



**UNIVERSIDADE FEDERAL DO CEARÁ
DEPARTAMENTO DE COMPUTAÇÃO
CURSO DE CIÊNCIA DA COMPUTAÇÃO**

DAVID BÔTO PAZ ARAGÃO

**PROPOSTAS PARA CONTROLE DE CONGESTIONAMENTO EM
REDE LTE OCASIONADO PELA COMUNICAÇÃO
MÁQUINA-A-MÁQUINA**

FORTALEZA, CEARÁ

2015

DAVID BÔTO PAZ ARAGÃO

**PROPOSTAS PARA CONTROLE DE CONGESTIONAMENTO EM
REDE LTE OCASIONADO PELA COMUNICAÇÃO
MÁQUINA-A-MÁQUINA**

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal do Ceará, como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

Área de concentração: Redes de Computadores

Orientador: Prof. Dr. Miguel Franklin de Castro

Co-Orientador: Prof. Dr. Dario Vieira

FORTALEZA, CEARÁ

2015

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

B767p Bôto Paz Aragão, David.
Propostas para Controle de Congestionamento em Rede LTE Ocasionado pela Comunicação Máquina-a-Máquina / David Bôto Paz Aragão. – 2015.
84 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Ciência da Computação, Fortaleza, 2015.

Orientação: Prof. Dr. Miguel Franklin de Castro.

Coorientação: Prof. Me. Dario Vieira.

1. Comunicação Máquina-a-máquina. 2. Redes LTE. 3. Controle de Congestionamento e Sobrecarga. I. Título.

CDD 005

DAVID BÔTO PAZ ARAGÃO

**PROPOSTAS PARA CONTROLE DE CONGESTIONAMENTO EM
REDE LTE OCASIONADO PELA COMUNICAÇÃO
MÁQUINA-A-MÁQUINA**

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação, da Universidade Federal do Ceará, como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação. Área de concentração: Redes de Computadores

Aprovada em: __/__/____

BANCA EXAMINADORA

Prof. Dr. Miguel Franklin de Castro
Universidade Federal do Ceará - UFC
Orientador

Prof. Dr. Dario Vieira
Ecole d'Ingénieur Généraliste en Informatique
et Technologies du Numérique - EFREI
Coorientador

Prof. Dr. Daniello Gonçalves Gomes
Universidade Federal do Ceará - UFC

Prof. Dr. Emanuel B. Rodrigues
Universidade Federal do Ceará - UFC

Aos meus pais e a minha irmã Fer-
nanda.

AGRADECIMENTOS

Em primeiro lugar agradeço a Deus, aos meus pais, pois sem eles eu não estaria aqui. Agradeço o amor, cumplicidade, compreensão e apoio que tenho recebido desde sempre dos meus pais e que não foram diferentes durante o período do mestrado. Agradeço a minha irmã, Fernanda, pelas infindáveis conversas sempre muito bem-humorada e pela revisão de alguns textos.

Agradeço ao meu orientador, Prof. Miguel Franklin, pelo direcionamento na pesquisa, pela sua dedicação na revisão dos artigos e pela busca por apoio financeiro sempre que precisei.

Agradeço ao meu coorientador, Prof. Dario Vieira, pela sua incansável atenção e orientação durante todo o processo de desenvolvimento, escrita e revisão dos trabalhos e artigos. Agradeço ao Prof. Dario Vieira pela sua ajuda e orientação no processo de imersão na vida e cultura francesa.

Agradeço a paciência e contribuição direta e indireta dos meus amigos.

Agradeço a todos os demais seres que de modo direto ou indireto contribuíram com esse trabalho.

Agradeço a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) pelo apoio financeiro.

“E um dia os homens descobrirão que esses discos voadores estavam apenas estudando a vidas dos insetos...”

(Mário Quintana)

RESUMO

O modelo de comunicação máquina-a-máquina (M2M) desempenhará um papel importante na visão da Internet das Coisas (IoT - *Internet of Things*). As redes *Long-term Evolution* (LTE) apresentam-se como potenciais redes de acesso para o M2M. Um dos problemas que emergem dessa integração é o congestionamento gerado pelo excesso de dispositivos M2M solicitando acesso a essas redes. Muitas são as soluções já apresentadas para este problema. Contudo, aspectos como priorização inter e intra-classes entre os dispositivos e estratégias para alocação dinâmica de recursos ainda são pouco explorados. Neste trabalho são propostos cinco mecanismos para controlar o congestionamento ocasionado pelos dispositivos M2M no contexto das redes LTE. O primeiro mecanismo proposto utiliza uma abordagem centrada nos dispositivos para controlar o uso dos recursos do canal de acesso aleatório. Os outros quatro mecanismos modelam a escassez dos recursos de tempo e preâmbulos para atender a demanda dos dispositivos como um problema de falência. Para cada tipo de recurso, tempo e preâmbulo, duas propostas são apresentadas. O problema da falência é abordado a partir de uma solução axiomática e usando uma estratégia utilizando jogos cooperativos. O impacto sobre os dispositivos H2H, a eficiência energética e a priorização entre as classes de dispositivos M2M são alguns dos principais aspectos analisados pelos mecanismos propostos. Com base nos resultados obtidos através de simulação, os mecanismos propostos apresentam vantagens em relação a um ou mais dos aspectos analisados. Todos os mecanismos foram simulados no simulador de rede NS-3.

Palavras-chave: Comunicação Máquina-a-máquina. Redes LTE. Controle de Congestionamento e Sobrecarga.

ABSTRACT

Machine-to-machine (M2M) is a communication model that will play an important role in vision of the Internet of Things (IoT). The Long-term Evolution (LTE) is expected to become a potential access networks for the M2M communication. A problem that arises from this scenario is the congestion caused by the huge number of M2M devices trying to access the network. Many solutions can be found in the literature that addresses this problem. However, only few mechanisms define priority among the different type of devices or present a dynamic approach to dynamically allocate resources of the random access channel. In this paper, we propose five mechanisms that aim to reduce the congestion caused by LTE network in the context of M2M devices. The first mechanism is centered in the devices and aim to control the access device to the network based in the congestion level presented by the network. The other four mechanisms model the scarcity of the time and preamble resources as a bankruptcy problem. For each resource, two approaches are proposed. The solutions presented to the bankruptcy problem follow from two an axiomatic approach and a cooperative game strategy. The impact over the H2H devices, the energy efficiency, and the prioritization among the different types of M2M devices are some of the main aspects analyzed by the proposed mechanisms. The result obtained through simulation show that the proposed mechanisms present advantages in one or more of the aspects analyzed. All approaches were simulated in the network simulator NS-3.

Keywords: Machine-to-Machine Communication. LTE Networks. Overload and Congestion Control.

LISTA DE FIGURAS

Figura 2.1	Evolução dos sistemas GSM e UTMS para o LTE. Rede pública de telefonia comutada (PSTN - <i>Public Switched Telephone Network</i>), Pacote de dados de rede (PDN - <i>Packet Data Network</i> -), Comutação de Circuito (CS - <i>Circuit Switch</i>) e Comutação de Pacote (PS - <i>Packet Swicth</i>) Domain, GERAN - <i>GSM EDGE Radio Access Network</i> . Adaptada de (SESIA; TOUFIK; BAKER, 2011).	22
Figura 2.2	Elementos do Núcleo da Rede LTE e suas respectivas interfaces. Adaptada de (SESIA; TOUFIK; BAKER, 2011).	22
Figura 2.3	Estrutura do quadro do LTE.	24
Figura 2.4	Máquina de Estado do Controle dos Recursos de Rádio do LTE (Radio Resource Control - RCC). Adaptada de (GRIGORIK, 2013).	25
Figura 2.5	Troca de mensagens entre dispositivo e estação base durante o procedimento de acesso aleatório baseado em contenção.	26
Figura 2.6	PRACH multiplexado com PUSCH e PUCCH. Adaptada de (SESIA; TOUFIK; BAKER, 2011).	28
Figura 2.7	Sobrecarga na eNodeB com solicitações.	29
Figura 2.8	Sobrecarga de elemento(s) do núcleo da rede.	29
Figura 3.1	CDF quando $\alpha = 1.0$	44
Figura 3.2	CDF quando $\alpha = 1.5$	45
Figura 3.3	Etapas do mecanismo PClass.	46
Figura 3.4	Probabilidade de Bloqueio.	48
Figura 3.5	Tempo Médio de Acesso.	48

Figura 3.6	Probabilidade de Bloqueio.	49
Figura 3.7	Tempo Médio de Acesso.	50
Figura 3.8	Dispositivos H2H.	50
Figura 3.9	Dispositivos M2M de alta e baixa prioridade.	51
Figura 3.10	Preâmbulos enviados pelos dispositivos H2H.	51
Figura 3.11	Preâmbulos enviados pelos dispositivos M2M.	52
Figura 4.1	Preâmbulos escolhidos por um único dispositivo.	56
Figura 4.2	Etapas comuns aos mecanismos Gtoken e Ptoken.	59
Figura 4.3	Probabilidade de acesso dos dispositivos H2H e total de acessos efetuados com sucesso dos dispositivos H2H.	61
Figura 4.4	Tempo médio de acesso dos dispositivos H2H.	61
Figura 4.5	Solicitações de acesso enviadas por dispositivos H2H.	62
Figura 4.6	Acessos realizados com sucesso pelos dispositivos $M2M_p$ e $M2M_{wp}$	62
Figura 4.7	Acessos realizados com sucesso pelos dispositivos H2H, $M2M_p$ e $M2M_{wp}$	63
Figura 4.8	Tempo médio de acesso dos dispositivos $M2M_p$ e $M2M_{wp}$	63
Figura 4.9	Probabilidade de acesso dos dispositivos M2M.	64
Figura 4.10	Solicitações de acesso enviadas por dispositivos $M2M_p$ e $M2M_{wp}$	64
Figura 4.11	Alocação de recursos de cada mecanismo.	65

Figura 4.12 Nível de utilização de recursos alocados de cada mecanismo.	65
Figura 4.13 I_1 , I_2 e I_3 representam os intervalo de <i>backoff</i> atribuídos respectivamente as classes H2H, $M2M_p$ $M2M_{wp}$. T é o tempo máximo de <i>backoff</i> que pode ser atribuído a uma classe.	66
Figura 4.14 Etapas Comuns aos Mecanismos Gcronos e Pcronos.	69
Figura 4.15 Acessos realizados com sucesso pelos dispositivos H2H.	70
Figura 4.16 Tempo médio de acesso dos dispositivos H2H.	71
Figura 4.17 Solicitações de acesso enviadas por dispositivos H2H.	72
Figura 4.18 Acessos realizados com sucesso pelos dispositivos $M2M_p$ e $M2M_{wp}$	72
Figura 4.19 Tempo médio de acesso dos dispositivos $M2M_p$ e $M2M_{wp}$	73
Figura 4.20 Solicitações de acesso enviadas por dispositivos $M2M_p$ e $M2M_{wp}$	73

LISTA DE TABELAS

Tabela 2.1	Características do M2M e H2H	20
Tabela 4.1	Tabela comparativa entre os mecanismos Gtoken, Ptoken, Gcronos e Pcronos.	68
Tabela A.1	Parâmetros utilizados nas simulações.	82
Tabela A.2	Parâmetros utilizados nas simulações.	83

SUMÁRIO

LISTA DE SIGLAS	
1 INTRODUÇÃO	17
2 REFERÊNCIAL TEÓRICO	20
2.1 Comunicação Máquina-a-Máquina	20
2.2 Long-Term Evolution - LTE	20
2.3 Arquitetura do LTE – Visão Geral	21
2.4 Estrutura de Quadros do LTE	23
2.5 Acesso Aleatório ao Canal em LTE	24
2.5.1 Procedimento de Acesso Aleatório ao Canal Baseado em Contenção	26
2.6 Sobrecarga e Congestionamento em LTE	28
2.7 O Problema da Sobrecarga e Congestionamento na Rede de Acesso via Rádio	30
2.7.1 Notação Utilizada	30
2.7.2 Formulação do Problema	31
2.7.3 Propostas da 3GPP	33
2.8 Teoria dos Jogos	34
2.9 Problema de Falência	35
2.9.1 Problema de Falência - Usando Jogos Cooperativos	36
2.9.2 Problema de Falência - Usando Regra Proporcional	37
2.10 Trabalhos Relacionados	38
3 PROPOSTA DE CONTROLE DE ACESSO AOS RECURSOS DO CANAL ..	42
3.1 Mecanismo PClass	42
3.1.1 Análise dos Resultados	47
3.1.1.1 Controle do Impacto sobre Dispositivos H2H	47
3.1.1.2 Priorização entre Dispositivos M2M	48
3.1.1.3 Acessos Realizados com Sucesso	49
3.1.1.4 Eficiência Energética	51
4 PROPOSTAS DE ALOCAÇÃO DINÂMICA DE RECURSOS DO CANAL ..	53
4.1 Mecanismos de Alocação Dinâmica de Códigos de Preâmbulos	53

4.1.1	Alocação de Código de Preâmbulos do RACH	56
4.1.2	Alocação de Códigos de Preâmbulos Usando Jogos Cooperativos (Gtoken)	57
4.1.3	Alocação de Códigos de Preâmbulos Usando a Regra Proporcional (Ptoken)	58
4.1.4	Análise dos Resultados	60
4.1.4.1	Impacto Sobre os Dispositivos H2H	60
4.1.4.2	Impacto Sobre os Dispositivos M2M.....	60
4.2	Mecanismos de Alocação Dinâmica de Tempo de Backoff	66
4.2.1	Alocação de Intervalos de Backoff Usando Jogos Cooperativos (Gcronos)	67
4.2.2	Alocação de Intervalos de Backoff Usando a Regra Proporcional (Pcronos)	68
4.2.3	Análise dos Resultados	70
4.2.3.1	Impacto Sobre os Dispositivos H2H	70
4.2.3.2	Impacto Sobre os Dispositivos M2M.....	71
4.3	Recomendações	74
5	CONCLUSÕES	75
5.1	Trabalhos Futuros	76
	REFERÊNCIAS BIBLIOGRÁFICAS	77
	APÊNDICE A – CONFIGURAÇÃO DOS EXPERIMENTOS.....	80
A.1	Simulação do Mecanismo Pclass	81
A.2	Simulação dos Mecanismos Gtoken, Ptoken, Gcronos e Pcronos	82
	APÊNDICE B – CONTRIBUIÇÕES.....	84

LISTA DE SIGLAS

IoT	Internet of Things
M2M	Machine-to-machine
MTC	Machine-Type Communication
LTE	Long-Term Evolution
H2H	Human-to-human
H2M	Human-to-machine
3GPP	Third Generation Partnership
RAN	Radio Access Network
CN	Core Network
RACH	Random Access Channel
MIMO	Multiple Input Multiple Output
GSM	Global System for Mobile Communications
UTMS	Universal Mobile Telecommunications System
FDD	Frequency Division Duplex
TDD	Time Division Duplex
VoIP	Voice over Internet Protocol
SAE	System Architecture Evolution
EPC	Evolved Packet Core
EPS	Evolved Packet System
CS- <i>Domain</i>	Circuit Switch Domain
eNodeB	Evolved Node B
MME	Mobile Management Entity
UE	User Equipment
P-GW	Packet Data Network Gateway
HSS	Home Subscriber Server
S-GW	Serving Gateway
GPRS	General Packet Radio Services
HSPA	High Speed Packet Access
CP	Cyclic Prefix
RB	Resource Block
RRC	Radio Resource Control
Short DRX	Short Discontinuous Reception
Long DRX	Long Discontinuous Reception
SIB	System Information Block
MIB	Master information block
SI	System Information
RA-RNTI	Random Access Radio Network Temporary Identifier

RAR	Random Access Response
BI	Backoff Indicator
PUSCH	Physical Uplink Shared Channel
HARQ	Hybrid Automatic Repeat reQuest
PRACH	Physical Random Access Channel
SOOC	Self-Optimizing Overload Control
CDF	Cumulative Density Function
RRC	Radio Resource Control
PDCP	Packet Data Convergence Protocol
RLC	Radio Link Control
SRS	Sounding Reference Signal
CQI	Channel Quality Indicator
SFN	System Frame Number

1 INTRODUÇÃO

O termo Internet das Coisas (Internet of Things - IoT), empregado pela primeira vez por Kevin Ashton em 1998 (PERERA et al., 2014a), é utilizado para designar um paradigma no qual objetos inteligentes estão interconectados e são capazes de trocar informações entre si em prol da realização de determinadas atividades. O modelo de comunicação máquina-a-máquina (M2M), também conhecido como *Machine-type Communication* (MTC), figura dentre um dos conceitos-chave para viabilização da IoT. Define-se por M2M o modelo de comunicação onde máquinas e dispositivos podem trocar informações entre si de modo autônomo, ou seja, sem quase ou nenhuma intervenção humana (LO et al., 2011). A rede *Long-Term Evolution* (LTE), devido as suas características de cobertura, segurança e acessibilidade tem um papel importante na concepção e implantação dos serviços e aplicações M2M (HASAN; HOSSAIN; NIYATO, 2013). No entanto, por ser um padrão inicialmente pensado para as redes celulares, este foi projetado para lidar com aplicações e serviços que seguem o modelo de comunicação humano-a-humano (H2H) e humano-a-máquina (H2M), e que diferem das características apresentadas pelo M2M. Assim, a quantidade de dispositivos esperados e a diversificação das aplicações e serviços M2M em relação ao H2H/H2M tem motivado a Third Generation Partnership (3GPP) a adaptar o LTE ao M2M (3GPP, 2014).

A 3GPP, responsável pela especificação e padronização do LTE, tem trabalhado na identificação e análise dos principais desafios para adaptar o LTE ao M2M e conseqüentemente aumentar sua aplicabilidade na IoT (3GPP, 2012). Dentre os desafios encontrados para adaptar o LTE ao modelo M2M, destacam-se os problemas de sobrecarga e congestionamento ocasionado pelo excesso de dispositivos M2M e a repercussão desses problemas sobre as aplicações e serviços H2H/H2M. O excesso de dispositivos solicitando acesso à rede pode impactar tanto na rede de acesso via rádio (*Radio Access Network* - RAN) como no núcleo da rede (*Core Network* - CN) do LTE. Motivados por estas e outras questões, estudos e propostas têm sido apresentadas (3GPP, 2014, 2012; LAYA; ALONSO; ALONSO-ZARATE, 2014; ZHENG et al., 2014; HUSSAIN; ANPALAGAN; VANNITHAMBY, 2014), que variam desde adaptações de mecanismos já existentes a propostas de novas arquiteturas para controlar ou mitigar esses problemas.

Uma das questões do controle de sobrecarga e congestionamento em LTE é a busca por mecanismos de controle que possam neutralizar ou reduzir o impacto dos dispositivos M2M sobre os dispositivos H2H durante o procedimento de acesso aleatório ao canal (RACH - *Random Access Channel*) e também possam provê acesso aos dispositivos M2M. Outros aspectos que também devem ser analisados por tais mecanismos é a latência no acesso e a eficiência energética, pois as aplicações podem apresentar diferentes restrições de tempo ou estarem embarcadas em dispositivos cuja fonte energética primária é proveniente de baterias. A eficiência energética das propostas apresentadas neste trabalho é analisada a partir do número médio de préambulos transmitidos.

Várias soluções para o problema de sobrecarga e congestionamento em LTE já foram propostas, como sumarizado em (LAYA; ALONSO; ALONSO-ZARATE, 2014). Contudo,

essas estratégias propõem melhorias na rede LTE para o M2M, mas não consideram a demanda gerada pelos dispositivos H2H (ZHENG et al., 2014). Além da priorização entre os dispositivos H2H e M2M, há poucos trabalhos que estabelecem priorização intra e inter-classe entre os dispositivos M2M durante o controle do congestionamento na RAN.

O objetivo deste trabalho é propor um conjunto de mecanismos para mitigar o congestionamento na rede LTE ocasionado pelo excesso de dispositivos M2M. Os mecanismos propostos neste trabalho diferem dos demais (e.g., (JIAN et al., 2013; PANG et al., 2014; LEE; KIM; YI, 2011)) por apresentarem uma estratégia dinâmica para divisão dos recursos do RACH. Os recursos em escassez gerenciados pelas propostas deste trabalho são os códigos de preâmbulos, utilizados durante o procedimento de RACH, e o intervalo de tempo disponível para ser dividido entre diferentes classes de *backoff*. Os mecanismos que tratam da gerência dos recursos (códigos de preâmbulos e tempo) são aplicados na estação base. Este trabalho também propõe um mecanismo para coordenar o acesso aos recursos do canal de acesso aleatório que prioriza e mitiga os efeitos do congestionamento. Este último é empregado na camada de acesso ao meio dos dispositivos. Em todos os cenários de simulação considerou-se a existência de dispositivos H2H.

O controle da sobrecarga e congestionamento na RAN do LTE tem como objetivo controlar o excesso de solicitações realizadas pelos dispositivos M2M e consequentemente reduzir o impacto ocasionado nos dispositivos H2H. Durante o procedimento de RACH, o dispositivo escolhe aleatoriamente um código de preâmbulo e aguarda o próximo *slot* disponível para enviar uma solicitação de acesso. Quando dois dispositivos utilizam o mesmo código tem-se uma colisão. Por comportar-se como o *slotted-alo*ha este mecanismo apresenta problemas quando há muitos dispositivos competindo pelos mesmos códigos e *slots* disponíveis. Assim, tem-se o aumento de modo exponencial da probabilidade de bloqueio com o aumento do número de dispositivos (LAYA; ALONSO; ALONSO-ZARATE, 2014).

Neste trabalho, são apresentados cinco mecanismos para mitigar o congestionamento que ocorre rede de acesso via rádio durante o procedimento de acesso aleatório ao canal. O primeiro mecanismo (PClass) segue uma abordagem essencialmente probabilística para controlar o envio dos preâmbulos pelas diferentes classes de dispositivos. Os outros quatro mecanismos têm como objetivo provê estratégias de alocação dinâmica de recursos (tempo e códigos de preâmbulos) de modo justo para as diferentes classes de dispositivos. Os mecanismos Gtoken e Ptoken são responsáveis pelas estratégias de alocação de código de preâmbulos e os mecanismos Gcronos e Pcronos pelas estratégias de alocação de tempo. O problema da escassez dos recursos é modelado como um problema de falência e para cada tipo de recurso tem-se uma modelagem diferente. Os mecanismos propostos neste trabalho são analisados segundo os seguintes aspectos:

- Número de acessos realizados com sucesso.
- Tempo de atraso no acesso.
- Quantidade de preâmbulos transmitidos.
- Probabilidade de bloqueio.

Sendo estes os indicadores-chave de desempenho comuns a todos os mecanismos simulados neste trabalho e aplicados aos dispositivos M2M e H2H.

Os mecanismos propostos neste trabalho originaram duas publicações. O artigo que apresenta o mecanismo PClass foi aprovado no Simpósio Brasileiro de Redes de Computadores - SBRC - 2015 (ARAGAO; VIEIRA; CASTRO, 2014). Um segundo trabalho, que apresenta os mecanismos Gcronos e Pcronos foi aprovado no *Global Information Infrastructure and Networking Symposium* - GIIS – 2015 (ARAGAO et al., 2015). Um terceiro artigo, que apresenta os Gtoken e Ptoken foi aprovado no Simpósio Brasileiro de Redes de Computadores - SBRC - 2016.

Esta dissertação está organizado da seguinte forma: o Capítulo 2 apresenta o referencial teórico e os trabalhos relacionados. Os Capítulos 3 e 4 apresentam os mecanismos propostos, os cenários de simulação e os resultados obtidos. Para finalizar, o Capítulo 5 apresenta as conclusões e os trabalhos futuros.

2 REFERÊNCIAL TEÓRICO

2.1 Comunicação Máquina-a-Máquina

A comunicação M2M é um modelo de comunicação utilizado por máquinas e dispositivos para troca de informações entre si com pouca ou nenhuma intervenção humana (LO et al., 2011). O M2M, também conhecida por MTC, nomenclatura atribuída pela 3GPP, é considerada uma das tecnologias base para viabilização da Internet das Coisas. Em contrastes com o modelo de comunicação humano-a-humano (H2H) e humano-a-máquina (H2M), o M2M pode apresentar uma ou mais das seguintes características listadas na Tabela 2.1.

A evolução da Internet das Coisas em conjunto com o M2M propicia o desenvolvimento de aplicações e serviços para os mais variados domínios. Dentre esses domínios estão o industrial, ambiental, social (PERERA et al., 2014b). Como exemplo de aplicações industriais tem-se as indústrias ligadas a atividades de aviação, *supply chain*, transporte e logística, aeroespacial, dentre outras. As aplicações pertencentes ao domínio social são voltadas para atividades como: saúde, ambientes inteligentes, entretenimento. No domínio ambiental as aplicações estão voltadas para atividades como: agricultura, reciclagem, alerta de desastre, monitoramento do meio ambiente, dentre outras (PERERA et al., 2014b). A IoT e suas aplicações exercerão um forte impacto nas atividades do dia-a-dia das pessoas com consequências sobre toda a sociedade.

2.2 Long-Term Evolution - LTE

O Long-Term Evolution (LTE), popularmente conhecida como rede de quarta geração (4G) de telefonia celular, é um padrão especificado pela 3GPP para as redes celulares. A 3GPP é formada por empresas e laboratório de pesquisa que juntos são responsáveis não somente por especificar, mas também identificar, reportar e sugerir soluções para as redes celulares. Com o aumento do número de usuários e aplicações demandando por taxas de transmissão cada vez mais elevadas, as operadoras estão tendo a necessidade de migrar suas redes para pa-

Características	Máquina-a-Máquina	Humano-a-Humano
Número de Dispositivos por Célula	Mil - Milhares	Algumas Centenas
Acesso à Estação Base	Massivo e Concorrente	Baixo
Mobilidade	Fixo - Lento	Fixo - Elevado
Dados Úteis Trafegados (quantidade)	Tipicamente pequeno	Pequena - Grande
Total de Tráfego por Dispositivo	Baixo	Baixo - Elevado
Fluxo do Tráfego	Unidirecional (tipicamente originado em dispositivos M2M)	Bidirecional
Frequência de Transmissão de Dados	Infrequente	Infrequente à Frequente
Consumo de Energia	Extremamente Baixa	Baixa à Elevada
Controle de Tempo	Em algumas aplicações específicas	Não há
Conexão Segura	Requerido	Requerido

Tabela 2.1: Características do M2M e H2H

drão LTE. Diante deste cenário, espera-se que nos próximos anos o LTE/LTE-A seja o padrão predominante nas redes celulares.

O padrão LTE é um padrão popularmente conhecido pelas altas taxas de transmissão apresentadas no sentido ascendente (*uplink*) e descendente (*downlink*) em comparação aos padrões que o antecede. Contudo, outras características também presentes no LTE são:

- Emprego da tecnologia MIMO (*Multiple Input Multiple Output*), importante para viabilização das taxas de transmissão no sentido ascendente e descendente de respectivamente 100 Mbps e 50 Mbps.
- Compatibilidade com outros padrões de redes, tais como o GSM (*Global System for Mobile Communications*) e UTMS (*Universal Mobile Telecommunications System*).
- Possibilidades de configuração da largura de banda para { 1.4, 2.5, 5, 10, 15, 20 } MHz.
- Capacidade de operar nos modos *Frequency Division Duplex* (FDD) e *Time Division Duplex* (TDD).
- Trabalha somente com a comutação por pacote (PS - *Packet Switching*), onde o serviço de Voz para o usuário é fornecido através de VoIP (*Voice over Internet Protocol*) ou utilizando outras tecnologias legadas.

Em contraste com os padrões anteriores para redes celulares, nos quais ainda ocorriam a comutação baseado em circuitos, o método de comutação empregado no LTE é baseado somente na comutação de pacotes. A Figura 2.1 destaca algumas das mudanças no GSM e UTMS em relação ao LTE.

2.3 Arquitetura do LTE – Visão Geral

A evolução do LTE em relação aos padrões que o antecede não ocorre somente na parte da rede de acesso via rádio com as *Evolved-UTRAN* (E-UTRAN), mas também na evolução da arquitetura do sistema (*System Architecture Evolution* - SAE), que é composto pelo *Evolved Packet Core* (EPC) (COX, 2012). Juntos, o LTE e o SAE formam o *evolved packet system* (EPS), como apresentado na Figura 2.1. O termo LTE foi inicialmente empregado para identificar os avanços ocorridos na rede de acesso via rádio, mas atualmente também tem sido utilizado para designar os avanços no núcleo da rede. Neste trabalho, o termo LTE é empregado para designar os avanços na rede de acesso via rádio e no núcleo da rede.

A rede LTE é composta pela rede de acesso via rádio (RAN) e pelo núcleo da rede (*Core Network* - CN). Como ilustrado na Figura 2.1, os avanços no EPC repercutem em significativas alterações em relação aos padrões GSM / UTMS. Como supracitado, no LTE o sistema de comutação é baseado somente em pacotes, deste modo, não há no EPC os elementos de rede responsáveis pelo suporte à comutação baseada em circuitos (*CS-Domain*).

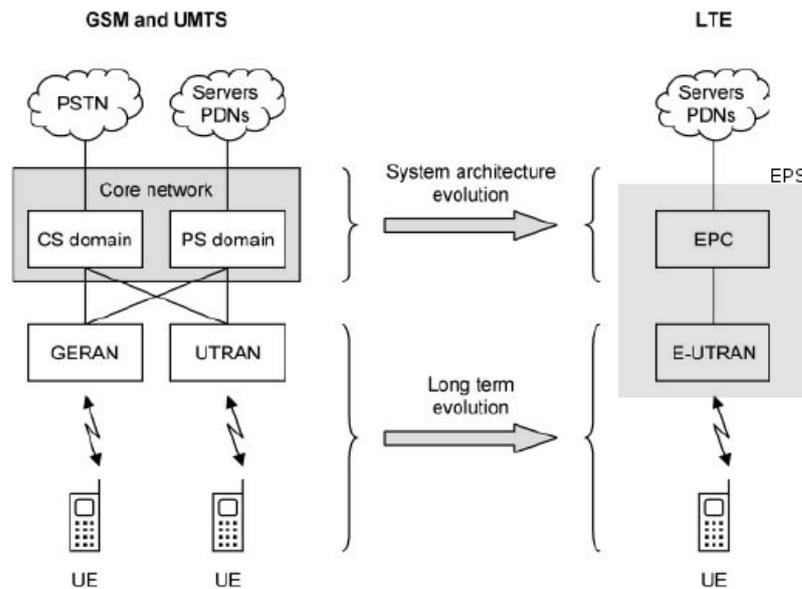


Figura 2.1: Evolução dos sistemas GSM e UTMS para o LTE. Rede pública de telefonia comutada (PSTN - *Public Switched Telephone Network*), Pacote de dados de rede (PDN - *Packet Data Network* -), Comutação de Circuito (CS - *Circuit Switch*) e Comutação de Pacote (PS - *Packet Swich*) Domain, GERAN - *GSM EDGE Radio Access Network*. Adaptada de (SESIA; TOUFIK; BAKER, 2011).

A rede de acesso via rádio (RAN) é formada pela estação base (*evolved Node B* - eNodeB) e é responsável pelas funções relacionadas ao rádio do LTE, tais como: alocação dinâmica de recursos para os dispositivos no *downlink* e *uplink* (escalonamento de pacotes), controle dos canais de rádio, controle de admissão, controle de mobilidade (*handover*) e suporte a vários sistemas de transmissão multi-antena (DAHLMAN; PARKVALL; SKOLD, 2013).

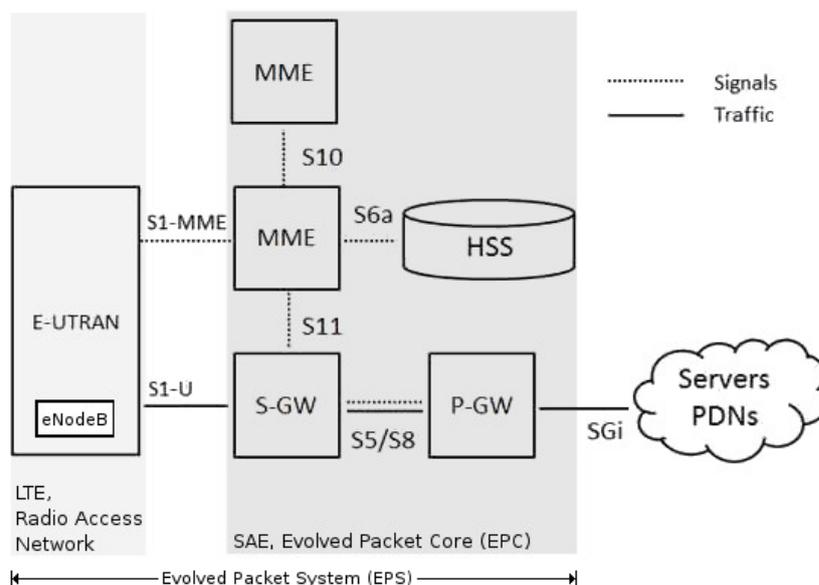


Figura 2.2: Elementos do Núcleo da Rede LTE e suas respectivas interfaces. Adaptada de (SESIA; TOUFIK; BAKER, 2011).

O núcleo da rede (CN) é responsável pelas atividades de autenticação, taxaço dos usuários e configuração de conexões fim-a-fim. O núcleo da rede é formado por diferentes nós, como apresentado na Figura 2.2 e descritos abaixo com mais detalhes.

- **Mobily Management Entity (MME).** O MME é um nó de controle que processa as informações trocadas entre os equipamentos de usuários (*User Equipament - UE*) / dispositivos e núcleo da rede (*Core Network - CN*). Consta dentre as suas funcionalidades: o estabelecimento de conexão e segurança entre os dispositivos e a rede; estabelecimento, manutenção e liberação de um canal de dados.
- **Packet Data Network Gateway (P-GW).** O P-GW é responsável pela alocação de endereço IP para os dispositivos, pela garantia de qualidade de serviço (QoS) e pela tarifação dos diferentes tipos de fluxo. O P-GW também age visando possibilitar a mobilidade dos dispositivos com outras redes que não seguem a padronização da 3GPP, tais como o CDMA2000 e as redes WiMax.
- **Home Subscriber Server (HSS).** O HSS atua como um banco de dados, no qual são armazenadas as informações sobre os usuários. Através do HSS, é possível, por exemplo, identificar a qual MME um determinado usuário está conectado, obter informações de profile relativas ao QoS do usuário, dentre outras.
- **Serving Gateway (S-GW).** O S-GW possibilita o movimento de terminais entre *eNodeBs*, sendo também responsável pela interação do LTE com outras tecnologias 3GPP, tais como GSM/GPRS e HSPA.

Vale ressaltar que os nós previamente descritos são apresentados a partir de um ponto de vista lógico. Em um ambiente real esses nós podem ser combinados, ou seja, um único nó físico pode atuar como MME, P-GW e S-GW.

2.4 Estrutura de Quadros do LTE

O quadro do LTE é mapeado no domínio do tempo e no domínio da frequência. No domínio do tempo, o quadro do LTE possui 10 ms, o qual é subdividido em 10 subquadros de 1 ms. Cada subquadro é novamente dividido em *slots* com duração de 0,5 ms e cada *slot* é formado por seis ou sete símbolos. A quantidade de símbolos por *slot* varia segundo as configurações do *Cyclic Prefix* (CP), que pode ser configurado como normal ou estendido. O tempo de duração do CP varia segundo as configurações da rede e é utilizado para reduzir a interferência entre símbolos. Consideramos, neste trabalho, a configuração mais usual, ou seja, a configuração na qual o CP tem tamanho normal, sete símbolos por *slot*. Um bloco de recurso (*Resource Block - RB*) ocupa 180 kHz, ou seja, o equivalente a 12 subportadoras de 15 kHz e 0.5 milissegundos no domínio do tempo, como apresentado na Figura 2.3 (SESIA; TOUFIK; BAKER, 2011).

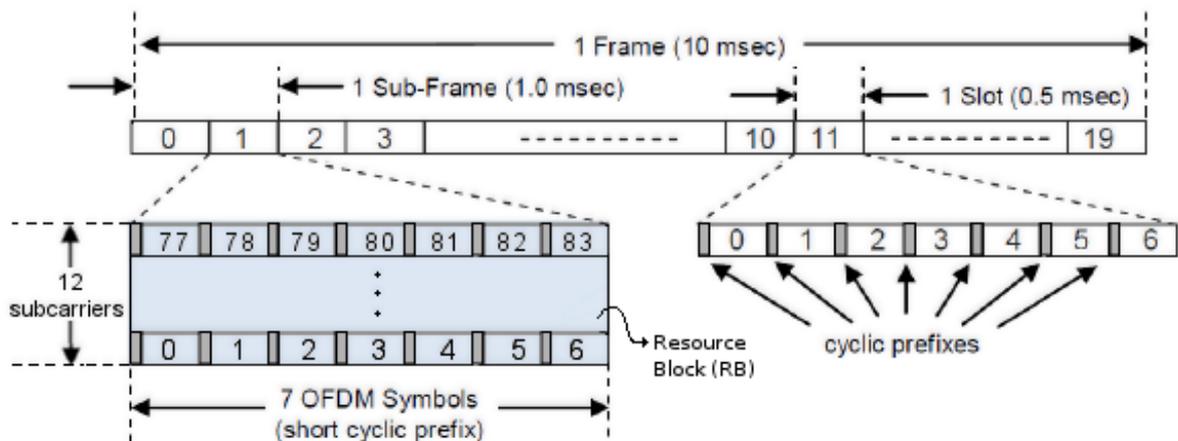


Figura 2.3: Estrutura do quadro do LTE.

2.5 Acesso Aleatório ao Canal em LTE

Na rede LTE, o procedimento de acesso aleatório ao canal (RACH) é iniciado para o estabelecimento de conexão de controle de rádio (*Radio Resource Control - RRC*) entre o dispositivo e a estação base. O procedimento de RACH tem como um de suas finalidades a sincronização do relógio do dispositivo com o relógio da estação base para e consequentemente viabilizar a troca de mensagens entre dispositivo e estação base. Os *slots* disponibilizados para o envio das solicitações são semialocados no tempo e frequência. Os possíveis estados do rádio dos dispositivos são definidos pela 3GPP através da máquina de estado apresentada na Figura 2.4. Apesar de haver possíveis alterações por parte das operadoras nos parâmetros que desencadeiam as transições, a máquina de estado em si é a mesma em toda a rede LTE.

A máquina de estado que representa o rádio dos dispositivos tem um total de quatro estados. No estado *RRC Idle*, o rádio do dispositivo consome menos de 15 mW e fica na escuta de tráfego de controle. No estado *RRC Connected*, o consumo de energia é bastante elevado (1000-3500 mW) devido ao envio ou recebimento de dados. Já os sub-estados *Short Discontinuous Reception (Short DRX)* e *Long Discontinuous Reception (Long DRX)* foram adicionados ao LTE para melhorar o desempenho energético do padrão. Nos estados *Short DRX* e *Long DRX* não ocorre a alocação de recursos da rede para os dispositivos. As transições que levam ao estado *RRC Connected* estão associadas às atividades de transmissão e recepção de dados. As transições para o estado *RRC Idle* são desencadeadas por (*timeouts*), como ilustrado na Figura 2.4 (GRIGORIK, 2013).

O procedimento de acesso aleatório no LTE pode ser categorizado em livre ou baseado em contenção. No primeiro, os recursos do canal de *uplink* são previamente alocados para os dispositivos pela eNodeB para o envio do pré-âmbulo de PRACH, primeira mensagem do procedimento de RACH. Assim, no procedimento de RACH livre de contenção não ocorre a competição entre os dispositivos pelos recursos do canal. Outra característica do método livre de contenção é a ausência de colisões, pois a comunicação com os dispositivos é iniciada por elementos que estão no núcleo da rede e propagada para os dispositivos pelas estações base.

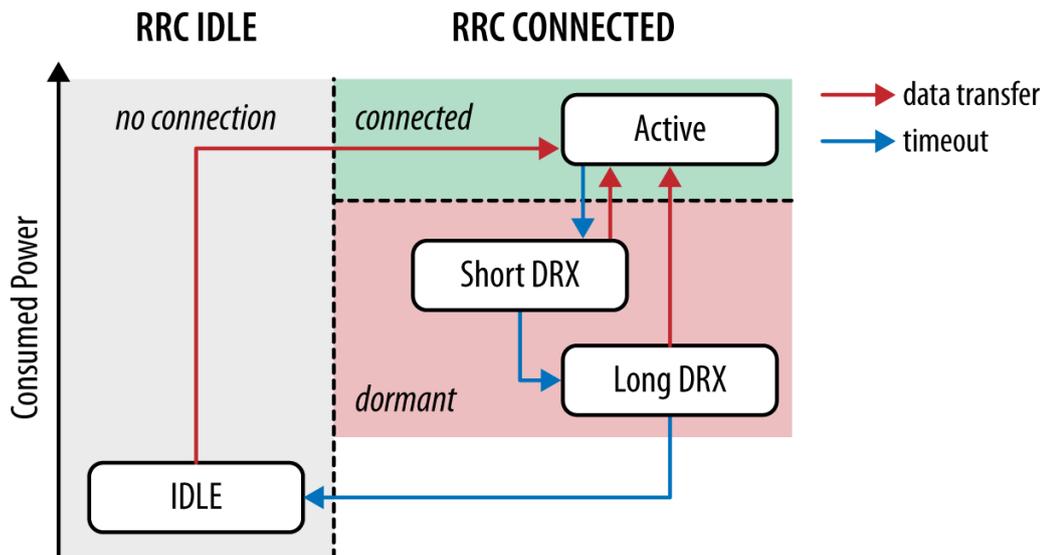


Figura 2.4: Máquina de Estado do Controle dos Recursos de Rádio do LTE (Radio Resource Control - RCC). Adaptada de (GRIGORIK, 2013).

O método livre de contenção é uma abordagem centralizada que normalmente está associada à atividade de controle que possuem fortes restrições de tempo, por exemplo, as atividades de *handover*. Em contrapartida, no método baseado em contenção não há alocação prévia de préambulos e os dispositivos competem entre si pelos recursos disponíveis quando precisam acessar a rede. Assim, quanto maior o número de dispositivos, mais colisões tendem a ocorrer. O procedimento de RACH é normalmente iniciando quando (DAHLMAN; PARKVALL; SKOLD, 2013):

1. O dispositivo possui dados que precisam ser enviadas, mas o canal de *uplink* não está sincronizado.
2. O dispositivo está no estado *RRC Connected*, ou seja, há um canal de dados estabelecido, mas o canal de *uplink* não está sincronizado para receber novos dados no canal de *downlink*, e assim transmite o *ACKnowledgment*, e *ACKnowledgment/Negative* (ACK, NACK) no *uplink*.
3. O dispositivo está no estado *RRC Connected* e está no processo de transição entre células.
4. Obtenção de informações de sincronização de relógio necessárias para determinar o posicionamento do dispositivo.
5. Durante a transição do estado *RRC Idle* para o *RRC Connected*, por exemplo, durante o estado inicial de acesso e atualização de localização de dispositivos na rede. No estado *RRC Idle* não há a presença de um canal de dados.
6. Ocorre falha no link de rádio e a recuperação do link faz-se necessária.

2.5.1 Procedimento de Acesso Aleatório ao Canal Baseado em Contenção

No procedimento RACH baseado em contenção, os dispositivos identificam os recursos disponíveis através das informações difundidas pelas mensagens de SIB (System Information Block) do tipo dois. No LTE, há treze tipos de mensagens de SIB. As mensagens de SIB, juntamente com a MIB, compõem o conjunto de mensagens de informação do sistema (*System Information - SI*). Algumas das informações contidas na SIB do tipo dois (SIB-2) são: i) mapeamento dos *slots* que estão disponíveis no canal para transmissão de pré-âmbulo do PRACH; ii) informações de sincronização; iii) parâmetros de controle de energia no *uplink*; iv) configurações de controle do canal de *uplink*; v) códigos de pré-âmbulos disponíveis. Dentre as informações repassadas aos dispositivos pelas MIB tem-se as configurações gerais da rede, como, por exemplo, a largura de banda utilizada na rede.

O acesso aleatório ao canal (RACH) baseado em contenção é um procedimento composto de quatro mensagens (msg1, msg2, msg3 e msg4) de sinalização, trocadas entre o dispositivo e a estação base, como ilustrado na Figura 2.5. As atividades realizadas durante cada um dos estágios do RACH são:

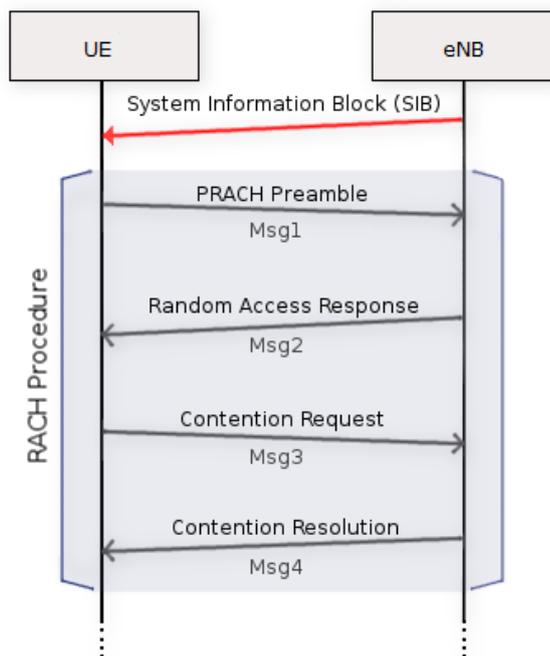


Figura 2.5: Troca de mensagens entre dispositivo e estação base durante o procedimento de acesso aleatório baseado em contenção.

- **Mensagem um (msg1), Envio de Pré-âmbulo.** Quando um dispositivo deseja obter acesso ao canal, este, primeiramente, deve aguardar o próximo *slot* (*RA-slot*) disponível. O LTE disponibiliza no máximo 64 códigos de pré-âmbulos para o procedimento de RACH, sendo parte desses códigos utilizados para o acesso aleatório livre de contenção. O dispositivo que deseja obter acesso à rede aguarda a mensagens de SIB difundidas pela eNodeB que contém as informações relacionadas aos *slots* e pré-âmbulos disponíveis. A

partir dessas informações, o dispositivo escolhe aleatoriamente um código de preâmbulo (PRACH) para ser enviado no próximo *slot* disponível. Se dois ou mais dispositivos enviarem o mesmo preâmbulo tem-se uma colisão. Caso contrário, a estação base é capaz de receber e processar os preâmbulos sem problemas. Contudo, há cenários que dificultam a detecção de uma colisão por parte da eNodeB, por exemplo, quando dois ou mais dispositivos que estão a uma mesma distância da estação base escolhem um mesmo código de preâmbulo. No entanto, apesar de não serem detectadas durante a primeira etapa, msg1, outras colisões tornam a ocorrer durante a terceira etapa, msg3, do RACH. Após a terceira etapa somente um dos dispositivos que colidiram será capaz de obter acesso à rede com sucesso. A janela de espera dos dispositivos por uma resposta da eNodeB pode variar de 2 a 10 milissegundos (LAYA; ALONSO; ALONSO-ZARATE, 2014). A contagem do tempo da janela de espera é iniciada a partir do terceiro subquadro, contando a partir do envio da solicitação (LAYA; ALONSO; ALONSO-ZARATE, 2014). O tempo da janela é flexível e informado pela eNodeB.

- **Mensagem dois (msg2), *Random Access Response*.** Após o recebimento e a decodificação das mensagens enviadas pelos dispositivos, a eNodeB gera o identificador *Random Access Radio Network Temporary Identifier* (RA-RNTI) com base no *RA-slot* em que o preâmbulo foi enviado. A mensagem de resposta (*Random Access Response* - RAR) é transmitida da eNodeB para os dispositivos através do canal físico compartilhado de *downlink*. Na mensagem de RAR, além do RA-RNTI, também são transmitidas as seguintes informações: i) instruções de tempo para sincronização do *uplink*; ii) alocação de recursos que serão utilizados pelos dispositivos para prosseguirem com o procedimento de RACH, isto é, para os dispositivos enviarem a terceira mensagem; iii) identificação de preâmbulos. Em caso de falha ou colisão, os dispositivos aguardam por um determinado intervalo de tempo determinado pelo indicador de *backoff* (*Backoff Indicator* - BI), propagado pela estação base. Os dispositivos só poderão retransmitir uma solicitação de acesso depois de transcorrido o tempo de espera.
- **Mensagem três (msg3), *Connection Request*.** Esta é a primeira mensagem do procedimento de RACH enviada no canal de *uplink* compartilhado (*Physical Uplink Shared Channel* PUSCH) que conta com o Hybrid Automatic Repeat reQuest (HARQ). Nesta mensagem, são transmitidas as requisições de conexão do controlador de recursos de rádio (*Radio Resource Control* - RRC), *tracking area updates* ou *scheduling request*. No caso de colisão de preâmbulo não detectada durante a primeira mensagem do RACH, um mesmo C-RNTI é atribuído aos dispositivos envolvidos na colisão e estes tornaram a colidir nesta etapa do RACH. Após um número máximo de retransmissões sem sucesso os dispositivos reiniciam o procedimento de RACH.
- **Mensagem quatro (msg4), *Connection Resolution*.** Esta mensagem é enviada pela eNodeB em resposta à mensagem *Connection Request*. Sendo a última mensagem do procedimento de acesso aleatório, o dispositivo que não a recebe pode inferir que houve uma falha durante o procedimento de RACH e conseqüentemente reprogramar uma nova tentativa de acesso, isto é, retorna para a primeira etapa do procedimento de RACH. As

tentativas de acesso são contabilizadas pelo dispositivo e após um determinado número de tentativas de acesso sem sucesso o dispositivo declarada a rede como não acessível.

As informações enviadas na primeira mensagem do procedimento de acesso aleatório ao canal são mapeadas na camada física, mas precisamente no canal PRACH. O PRACH é multiplexado no domínio do tempo e frequência com os canais PUSCH e PUCCH. Os recursos utilizados para o PRACH são alocados de modo semiestático na região do PUSCH, como mostrado na Figura 2.6. No domínio da frequência, o PRACH ocupa o equivalente a seis blocos de recursos (RBs - *Resource Blocks*), totalizando 1.08 MHz (6×180 kHz). No domínio do tempo, o PRACH ocupa o tempo equivalente a dois blocos de recursos, ou seja, 1 ms. O escalonamento do PRACH no PUSCH e a periodicidade com que os recursos no PUSCH são semialocados para o PRACH são determinados pelo índice de configuração do PRACH, repassado para os dispositivos na mensagem de SIB do tipo dois juntamente com outras configurações da rede relacionadas com o procedimento de RACH. Outras informações de configurações do RACH também repassadas na SIB do tipo dois são: o número de códigos de préambulos disponíveis para o procedimento de RACH (*numberOfRA-Preambles*), o número máximo de retransmissões de préambulos (*preambleTransMax*) que podem ser realizadas pelos dispositivos em caso de colisão, a duração da janela de espera da mensagem de resposta de solicitação do acesso aleatório (*ra-ResponseWindowSize*).

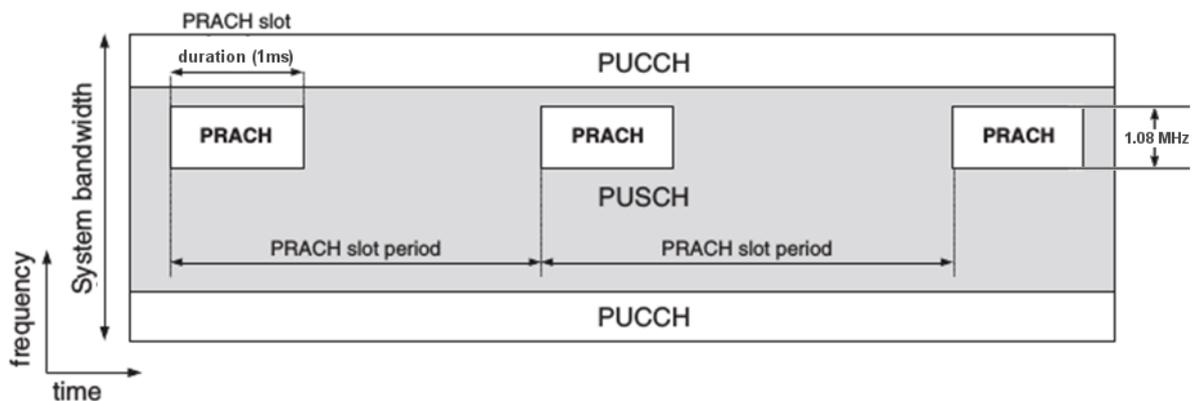


Figura 2.6: PRACH multiplexado com PUSCH e PUCCH. Adaptada de (SESIA; TOUFIK; BAKER, 2011).

2.6 Sobrecarga e Congestionamento em LTE

O número excessivo de dispositivos M2M solicitando acesso à rede pode acarretar congestionamento tanto no núcleo da rede como na rede de acesso via rádio (AMOKRANE et al., 2012). Para provê suporte as aplicações e serviços M2M, o LTE precisa se adaptar para lidar com as diferenças apresentadas por este modelo de comunicação (cf. Tabela 2.1). Na rede de acesso via rádio, um dos cenários de congestionamento ocorre quando há muitos dispositivos enviando solicitações de acesso para uma mesma eNodeB, como ilustrado na Figura 2.7. Neste

caso, devido ao número limitado de recursos no canal, tem-se o aumento significativo do número de colisões, impossibilitando a eNodeB de tratar as solicitações de acesso. Um exemplo de cenário que pode congestionar o núcleo da rede está ilustrado na Figura 2.8. Neste, o número de dispositivos associado às eNodeB é moderado, contudo, como há muitas eNodeB associadas a um mesmo nó do núcleo da rede, tem-se o congestionamento e eventual sobrecarga em um ou mais elementos do núcleo da rede.

Segundo (WEYRICH; SCHMIDT; EBERT, 2014), o número estimado de dispositivos até 2020 é de aproximadamente 50 bilhões. Com base nesses números, cenários como estes serão recorrentes. Tais problemas são considerados de alta prioridade e precisam ser controlados ou neutralizados para viabilizar o M2M em LTE (3GPP, 2014).

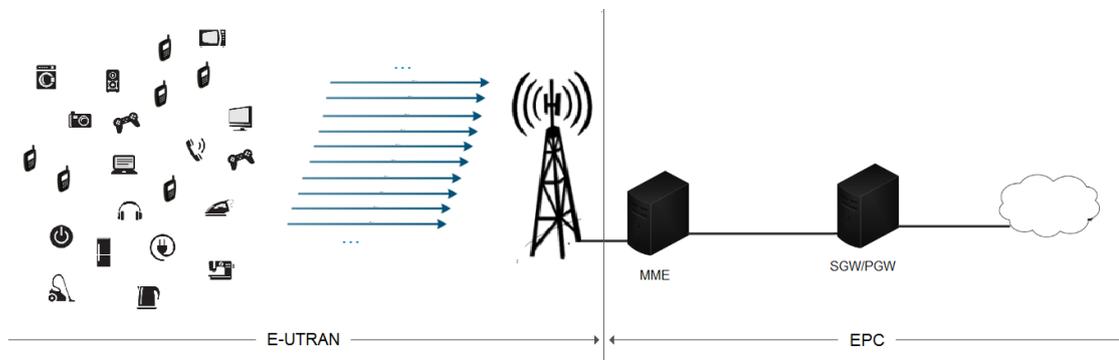


Figura 2.7: Sobrecarga na eNodeB com solicitações.

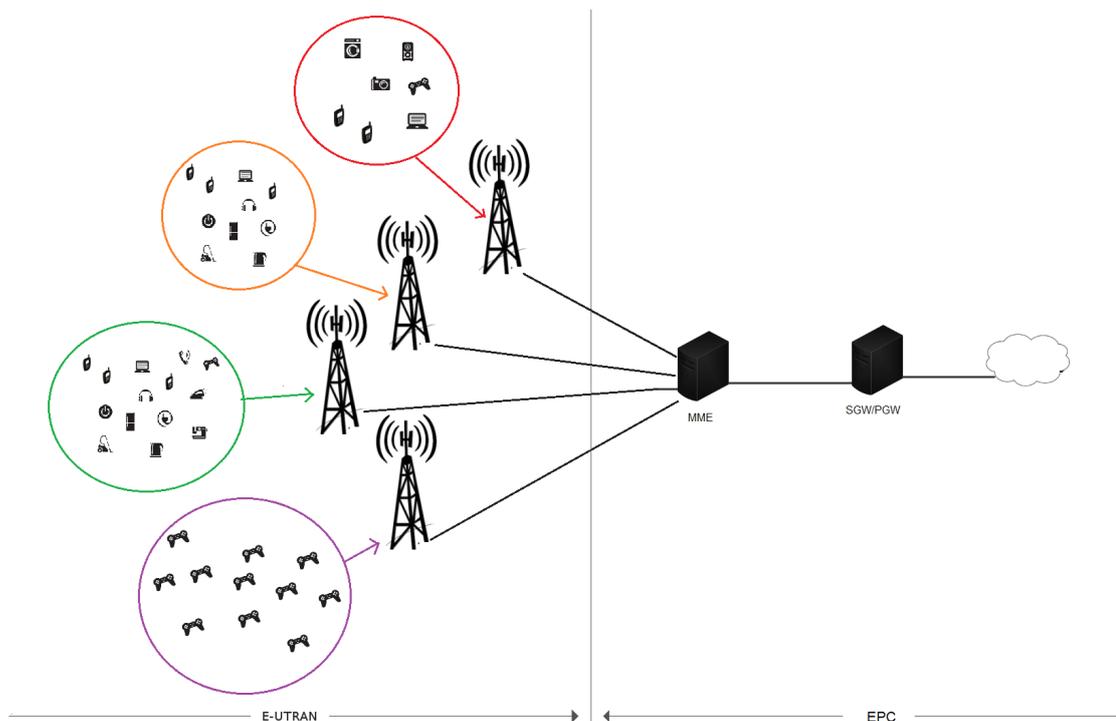


Figura 2.8: Sobrecarga de elemento(s) do núcleo da rede.

As estratégias apresentadas neste trabalho estão orientadas para mitigar o problema do congestionamento na RAN que ocorre durante o procedimento de acesso aleatório ao canal. Os cinco mecanismos propostos neste trabalho utilizam estratégias diferentes para mitigar este problema. Um mecanismo (PClass) está orientado para coordenar o envio das solicitações de acesso pelos dispositivos. Dois mecanismos (Gtoken e Ptoken) se baseiam na estratégia de alocação de recursos de RACH apresentada pela 3GPP, propondo que a reserva desses recursos seja efetuada dinamicamente. Para finalizar, dois mecanismos (Gcronos e Pcronos) se baseiam na estratégia de *backoff*, identificando e atribuindo diferentes valores de *backoff* para as diferentes classes de dispositivos simuladas.

2.7 O Problema da Sobrecarga e Congestionamento na Rede de Acesso via Rádio

O congestionamento na RAN ocorre devido à quantidade limitada de recursos de rádio para atender a demanda de solicitações de acesso durante o procedimento de RACH. Por comporta-se como uma variação do *slotted-alo*ha, o procedimento de RACH do LTE apresenta problemas de congestionamento com o aumento do número de dispositivos, com destaque para o aumento do número de colisões que cresce exponencialmente (ZHENG et al., 2014).

O modelo de acesso aleatório ao canal apresenta-se como um gargalo para o LTE. Um dos gargalos do procedimento de RACH ocorre devido ao número limitado de códigos de preâmbulos. O LTE conta com um total de 64 códigos de preâmbulos que devem ser divididos entre o procedimento de RACH baseado em contenção e o livre de contenção. A divisão padrão adotada pela 3GPP considera que 10, dos 64 preâmbulos, devem ser alocados para o procedimento de RACH livre de contenção. Deste modo, os 54 códigos restantes são objeto de disputa pelos dispositivos. Vale ressaltar que o procedimento de RACH padrão não reserva códigos de preâmbulos para classes de dispositivos. Assim, em um cenário composto por dispositivos das classes H2H e M2M, os códigos de preâmbulos disponíveis são disputados igualmente pelos indivíduos de ambas as classes.

2.7.1 Notação Utilizada

Esta seção descreve os símbolos e apresenta o problema da sobrecarga durante o procedimento de acesso aleatório ao canal. A notação utilizada na formulação do problema é mostrada a seguir:

- X - variável aleatória;
- $Unif(a, b)$ - distribuição uniforme no intervalo (a, b) ;
- L - número máximo de tentativas de acesso que podem ser realizadas pelos dispositivos;
- t - número de tentativas de acesso realizadas por um determinado dispositivo, ou seja, número de preâmbulos transmitidos por um determinado dispositivo, $t \in \{0 \dots L\}$.
- P_{ac} - parâmetro probabilidade de acesso (p) difundido pela estação base;

- P_{cong} - parâmetro propagado pela estação base para informa o nível de congestionamento na rede de acesso via rádio aos dispositivos;
- P_i - parâmetro probabilidade de transmissão de preâmbulo de solicitação de acesso calculado individualmente por cada dispositivo.
- i - índice utilizado para identificar a classe do dispositivo e conseqüentemente o nível de prioridade do dispositivo. A prioridade diminui à medida que o índice i aumenta;
- α - parâmetro utilizado para ajustar a probabilidade de acesso das diferentes classes de dispositivos de acordo com o nível de congestionamento na rede de acesso;
- $T_{backoff}(i)$ - função que retorna o tempo de *backoff* em milissegundos, onde i é a classe do dispositivo;
- P_c - probabilidade de colisão de preâmbulo durante o procedimento de acesso aleatório ao canal;
- e - constante matemática, número de Euler;
- λ - número médio de solicitações de acesso realizada por segundo;

2.7.2 Formulação do Problema

O procedimento de RACH segue os princípios do *slotted-aloha* (LAYA; ALONSO; ALONSO-ZARATE, 2014). Devido à analogia existente entre o *slotted-aloha* e o procedimento de RACH do LTE, pode-se dizer que a probabilidade de colisão (P_c) durante o procedimento de RACH na rede LTE é aproximado pela equação (PHUYAL et al., 2012; 3GPP, 2014):

$$P_c = 1 - e^{\left(\frac{-\lambda}{N_{preambulos}}\right)} \quad (2.1)$$

Onde $N_{preambulos}$ é o total de *slots* de acesso (*RA-Slots*) por segundo e λ a quantidade média de solicitações realizadas por segundo para uma estação base. Para uma largura de banda (B) dada em MHz, tal que $B \in \{1.4, 2.5, 5, 10, 15, 20\}$ a quantidade aproximada de blocos de recursos disponíveis por quadro é dada por: $PRB_{TotalQuadro} = \frac{B}{F_{subquadro}} \times T_{quadro} \times 2$, sendo T_{quadro} o tempo do quadro LTE (dado em milissegundo), $F_{subquadro}$ a largura de banda (dada em kHz) de um bloco de recurso no LTE. Os *RA-Slots* podem ser configurados para ocorrer com periodicidade n em um T_{quadro} , onde $n \in \{0.5, 1, 2, 3, 5, 10\}$, onde $n = 0.5$ indica que o *slot* é disponibilizado a cada 20 ms, ou seja, a cada 2 quadros do LTE tem-se um *slot*. A partir do número de *slots* disponibilizados por quadro, a quantidade de blocos de recursos disponíveis para o procedimento de RACH por segundo é dado por:

$$PRB_{Total} = \frac{PRB_{TotalQuadro}}{2} \times \frac{1000}{T_{quadro}} \times n. \quad (2.2)$$

Assim, o número de solicitações de acesso suportados por segundo pela estação base é dado por:

$$T_{RAR} = \frac{PRB_{Total}}{PRB_{RACH}} \quad (2.3)$$

Onde PRB_{RACH} identifica a quantidade de blocos de recursos utilizados durante o procedimento de RACH para o envio do preâmbulo de PRACH. Um preâmbulo de PRACH ocupa o equivalente a seis 6 blocos de recursos do canal de *uplink*. Fixando-se uma determinada taxa de colisão (P_c), o número de *slots* de acessos necessários para que a rede consiga suportar uma determinada intensidade de acesso é dado em (3GPP, 2014):

$$\lambda = -L \times \ln(1 - P_c) \quad (2.4)$$

Assim, como esperado, o aumento na quantidade de dispositivos, tem-se mais colisões devido ao número limitado de *slots* de acesso. O controle da sobrecarga e congestionamento é uma questão de caráter emergencial para que a rede celular, mais especificamente o padrão LTE, possa suportar a demanda esperada para os próximos anos de dispositivos M2M. Abaixo estão listadas alguns exemplos de ações realizadas por aplicações e serviços M2M que podem propiciar o surgimento de cenários de sobrecarga e congestionamento na RAN (3GPP, 2012):

- Mau funcionamento de aplicação M2M. Diante deste cenário as operadoras têm interesse por medidas para identificar e bloquear a aplicação que está com problemas, sem que as demais aplicações ou serviços sejam afetados.
- Número massivo de dispositivos M2M solicitando acesso à rede de modo simultâneo. Tal comportamento pode ser observado em aplicações de monitoramento, onde um determinado evento pode sensibilizar muitos dispositivos a realizarem uma mesma atividade, por exemplo, solicitar acesso à rede de modo quase que simultâneo.
- Aplicações que realizam atividades recorrentes e sincronizadas. Este comportamento pode gerar picos de solicitações, e conseqüentemente sobrecarregar a rede. Diante deste cenário, algumas das medidas realizada pelas operadoras das redes celulares a dispersão das solicitações no tempo.

As técnicas empregadas no controle do congestionamento podem ser classificadas em rígidas (*rigid*) ou suaves (*soft*) (AMOKRANE et al., 2012). As técnicas, ou mecanismos, suaves são utilizados pelas operadoras quando estas têm interesse em reduzir a frequência que determinados processos são executados pelos dispositivos, mas sem ter que eliminá-los. Em contrapartida, os mecanismos rígidos são empregados quando o objetivo é impossibilitar a conexão dos dispositivos MTC à rede, ou a execução de determinados processos (AMOKRANE et al., 2012). Uma das desvantagens da estratégia suave é a sua dependência por recursos para comunicação entre os elementos da rede. Contudo, diante de um cenário de sobrecarga ou congestionamento,

a comunicação entre esses elementos pode ficar comprometida, uma vez que tais canais podem estar congestionados. Os mecanismos rígidos agem nas bordas da rede impossibilitando o estabelecimento de novas conexões, não demandando, assim, mesmo diante de cenários de sobrecarga e congestionamento, recursos para troca de informações com os demais elementos da rede.

Os mecanismos propostos neste trabalho são categorizados como rígidos. O PClass atua na camada de controle de acesso ao meio para restringir o envio de solicitações de acesso à rede pelos dispositivos. Os Gtoken e Ptoken atuam na estação base e controlam o acesso à rede a partir da distribuição dos recursos necessários para o início do procedimento de RACH. O Gcronos e Pcronos agem na estação base alterando o tempo de *backoff* dos dispositivos, ou seja, determinado como as solicitações de acesso devem ser dispersadas.

2.7.3 Propostas da 3GPP

A 3GPP, ciente do problema e da necessidade de mecanismos para controlar a sobrecarga e o congestionamento ocasionado pela comunicação M2M em LTE, tem apresentado algumas possíveis estratégias para o problema. Dentre os esquemas propostos pela 3GPP, seis são predominantemente os mais conhecidos, sendo eles: *backoff* específico, reserva de recursos do RACH, *Access Class Barring* (ACB), *slotted access*, alocação dinâmica de recursos do PRACH e a estratégia *pull*. As principais propostas apresentadas pela 3GPP são descritas abaixo:

- **Backoff específico.** A estratégia de *backoff* visa controlar o congestionamento durante o acesso dispersando as solicitações no tempo. A proposta de *backoff* específico apresentada pela 3GPP consiste em atribuir diferentes intervalos de *backoff* para os dispositivos de acordo com a sua classe/tipo. No cenário apresentado pela 3GPP o valor dos intervalos atribuídos aos dispositivos H2H e M2M são respectivamente 20ms e 960ms. Contudo, em cenários mais congestionados tal proposta não controla o impacto gerado pelo excesso de dispositivos sobre o H2H. Além disso, nesta proposta os dispositivos M2M são tratados de modo homogêneo, ou seja, sem distinção ou qualquer tipo de priorização.
- **Reserva de recursos no RACH.** Nesta estratégia os recursos do RACH são divididos entre os dispositivos M2M e H2H, evitando que a sobrecarga durante o procedimento de RACH afete os dispositivos H2H. A separação dos recursos destinados aos dispositivos M2M e H2H assegura que os dispositivos H2H não sejam afetados pela comunicação M2M e vice-versa. Contudo, não há indicações, por parte da 3GPP, de como esses recursos devem ser distribuídos.
- **Access Class Barring (ACB).** Neste mecanismo a sobrecarga na RAN é controlada através da atribuição de valores ao parâmetro p (probabilidade de acesso) dos dispositivos. Antes de tentar se conectar à rede o dispositivo gera um valor aleatório (q) entre 0 e 1 que será comparado com o valor do parâmetro probabilidade de acesso. Se $q < p$ o dispositivo pode solicitar acesso à rede. Deste modo, quanto menor o valor atribuído ao parâmetro

p menor será a probabilidade do dispositivo de enviar uma solicitação de acesso. Essa proposta não estabelece nenhuma distinção entre os dispositivos M2M, ou seja, todos os dispositivos M2M são tratados de modo homogêneo.

- **Slotted Access.** Neste mecanismo um determinado *slot* de acesso é atribuído a cada dispositivo M2M. Quando um dispositivo M2M deseja transmitir, este deve identificar o *slot* que lhe foi reservado com base no seu identificador (ID) e no parâmetro *RA-Cycle*¹. As informações de como os recursos estão mapeados são difundidas periodicamente pela estação base. Quanto maior o número de dispositivos M2M em uma mesma célula mais dispositivos compartilham um mesmo *slot*, elevando assim o número de colisões. Uma das estratégias utilizada por esse mecanismo para controlar as colisões é aumentar o *RA-Cycle*. Porém, o aumento do *RA-Cycle* acarreta aumento do tempo de espera dos dispositivos.
- **Alocação dinâmica de recursos do PRACH.** A solução proposta por este mecanismo é de semialocar recursos no PUSCH para o PRACH. Deste modo, mais *slots* são disponibilizados para os dispositivos enviarem suas solicitações de acesso durante o procedimento de RACH. Neste mecanismo os recursos são dinamicamente alocados para o PRACH e variam de acordo com o nível de congestionamento na rede. Assim, quanto maior o nível de congestionamento mais recursos serão alocados para o PRACH. Porém, à medida que mais recursos do PUSCH são semialocados para o PRACH, reduz-se a quantidade de recursos utilizados para transferência de dados. Deste cenário emerge um dos desafios associados com a adoção dessa estratégia que consiste na necessidade de se analisar qual o impacto da alocação desses recursos do PRACH sobre o tráfego de dados e vice-versa. Como já mencionado em (LAYA; ALONSO; ALONSO-ZARATE, 2014), identificar quando e como os recursos de PRACH devem ser alocados ainda é uma questão em aberto.
- **Pull.** Nesta abordagem a comunicação é iniciada pela eNodeB, responsável por enviar uma solicitação de *paging* para o dispositivo. Esta estratégia tem um caráter centralizador, ou seja, quem determina quando os dispositivos podem transmitir é a estação base (eNodeB). No contexto real das aplicações M2M, os dispositivos enviam dados de modo esporádico, dificultando a aplicabilidade desta abordagem (3GPP, 2014).

2.8 Teoria dos Jogos

Teoria dos jogos é um conjunto de ferramentas utilizadas para modelagem e compreensão da interação entre indivíduos, também chamado de jogadores, diante de cenários de tomada de decisão. São assertivas comuns à teoria dos jogos o comportamento racional dos participantes do jogo e a tomada de decisão realizada pelos seus jogadores baseadas no conhecimento e comportamento esperado dos outros participantes (OSBORNE; RUBINSTEIN, 1994). Os principais elementos presentes nos jogos são:

¹Número inteiro enviado pela eNodeB múltiplo do quadro no qual foi enviado

- **Jogadores / Indivíduos:** participantes do jogo.
- **Estratégias:** possíveis ações que um jogador pode escolher durante o jogo.
- **Pagamento / *payoff*:** ganhos e perdas, ou seja, retorno obtido pelo jogador durante o jogo.

O jogador é a entidade básica comum a todos os modelos de jogos. Dependendo do tipo de jogo, pode-se analisar a interação dos jogadores a partir de duas perspectivas. Na primeira, a interação de cada jogador durante o jogo é analisada individualmente. No segundo, a análise da interação deixa de ser individualizada e passa a ser sobre o grupo de jogadores. Dependendo da perspectiva adotada tem-se um modelo composto por estratégias individuais associadas a cada jogador ou um modelo composto pelo conjunto de estratégias dos grupos de jogadores, conhecidos, respectivamente, como jogos não cooperativos e jogos cooperativos. Ao contrário dos jogos não cooperativos ou estratégicos, os jogadores, nos jogos cooperativos, tendem a cooperar e a formarem coalizões para maximizar o *payoff* para o grupo. Os jogos cooperativos também diferem dos jogos não cooperativos em relação ao conjunto de estratégias, pois nos jogos cooperativos não há a noção de conjunto de estratégias individuais (OSBORNE; RUBINSTEIN, 1994).

Os jogos cooperativos podem ser classificados em jogos com transferência de utilidade e jogos sem transferência de utilidade. Nos jogos com transferência de utilidade as ações tomadas pelas coalizões determinam o *payoff* para cada coalizão. Nos jogos sem transferência de utilidade as ações tomadas pela coalizão determinam o *payoff* de cada indivíduo. Além dos jogadores, nos jogos cooperativos tem-se o indivíduo que atua como árbitro. Um árbitro, para ser considerado imparcial, deve basear as suas decisões em princípios objetivos. Deste modo, pode-se dizer que um árbitro imparcial é capaz de realizar as mesmas decisões em jogos equivalentes.

2.9 Problema de Falência

O problema de falência (do inglês: *Bankruptcy Problem*) representa um cenário de distribuição de uma determinada quantidade de recursos perfeitamente divisíveis entre um grupo de indivíduos onde o total demandado pelos indivíduos ultrapassa a quantidade disponível para ser distribuída (HERRERO; VILLAR, 2001). O problema de falência pode ser representado a partir da tupla $B(N, E, c)$, onde $N = \{1, 2, 3, \dots, n\}$ é o conjunto de indivíduos, $E \in \mathbb{R}_+$ identifica a quantidade de recursos disponíveis e $c \in \mathbb{R}_+^n$ o vetor de demanda indexado por i . Várias propostas têm sido apresentadas e estudadas ao longo dos anos para lidar com problemas desta categoria (HERRERO; VILLAR, 2001).

O problema de falência pode ser analisado a partir de duas perspectivas (HERRERO; VILLAR, 2001). A primeira utiliza um *framework* de jogos, onde o problema é formulado como um jogo cooperativo com transferência de utilidade ou como um jogo de barganha. A segunda utiliza uma abordagem axiomática, onde as soluções são caracterizadas a partir de

propriedades distintas, as quais são utilizadas para nortear o julgamento de como os recursos devem ser distribuídos.

2.9.1 Problema de Falência - Usando Jogos Cooperativos

A abordagem utilizando jogos com transferência de utilidade para o problema de falência apresentada por O'Neil em (O'NEILL, 1982) associa a cada problema de falência $B(N, E, c)$ um jogo cooperativo $G(N, v_c)$, sendo v_c uma função característica $v_c \rightarrow 2^n$ dada por:

$$v_c(S) = \max\{E - \sum_{i \in N \setminus S} c_i, 0\} \quad (2.5)$$

onde $v_c(N) = E$ e $S \subseteq N$ é o subconjunto com as possíveis coalizões e $v_c(0) = 0$. Deste modo, a função característica é responsável por atribuir a cada subgrupo de jogadores uma valoração. Assim, em um jogo com n jogadores existem 2^n possíveis coalizões.

A partir da Equação 2.5 é possível verificar que o valor de uma dada coalizão S é dado pelo resultado do total de recursos disponíveis menos o total demandado pelos jogadores que não fazem parte da coalizão S . Depois de identificar qual o valor/relevância de uma dada coalizão, tem-se o desafio de identificar como os recursos devem ser distribuídos entre os indivíduos que formam a coalizão.

O valor de Shapley, proposto por Lloyd Shapley em 1953, é um conceito associado aos jogos cooperativos que propõe a distribuição dos recursos entre os indivíduos de uma dada coalizão de modo justo. Segue abaixo a função utilizada para calcular o valor de Shapley (OSBORNE; RUBINSTEIN, 1997):

$$\phi_i(v_c) = \sum_{S \subseteq N \setminus i} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \underbrace{(v_c(S \cup \{i\}) - v_c(S))}_{\text{contribuição marginal}} \quad (2.6)$$

Onde v_c é uma função característica (c.f., Equação 2.5). Na teoria dos jogos, o Valor de Shapley determina como os ganhos obtidos por um grupo de indivíduos pode ser dividido de modo justo entre seus membros. O senso de justiça apresentado pelo Valor de Shapley é sustentado pelas propriedades que o identifica de forma única, ou seja, até então não se conhece outra estratégia de divisão de recursos capaz de provê as mesmas propriedades providas por ele. São propriedades presentes no Valor de Shapley: a eficiência, a simetria, a agregação e o jogador nulo. Os quatro axiomas que sustentam essas propriedades estão descritos abaixo.

- Axioma da eficiência: a soma total dos valores de Shapley atribuídos aos jogadores é igual ao valor da coalizão que engloba todos os jogadores.
- Axioma da simetria: dois jogadores $i, j \in N$ são simétricos se apresentarem a mesma contribuição para a coalizão.
- Axioma da agregação ou aditivo: dado dois jogos $G(N, v)$ e $G(N, w)$, pode-se definir um $G(N, v + w)$, onde $(v + w)(S) = v(S) + w(S), \forall S \subset N$.

- Axioma do jogador nulo: a contribuição do jogador a qualquer coalizão é zero, ou seja, se i é um jogador nulo então $\phi_i(v) = 0$.

2.9.2 Problema de Falência - Usando Regra Proporcional

O problema de falência é amplamente conhecido e tem as seguintes, mas não somente, soluções axiomáticas (HERRERO; VILLAR, 2001):

- *Proportional Rule* (P). A quantidade de recursos atribuídos a um indivíduo deve ser proporcional ao valor demandado.
- *Constrained Equal-awards Rule* (CEA). Os recursos devem ser igualmente distribuídos entre os indivíduos, mas com a restrição de que nenhum indivíduo irá receber além do valor da sua demanda.
- *Constrained Equal-losses Rule* (CEL). A diferença entre a quantidade de recursos demandado e os recursos disponíveis é dividida igualmente entre os indivíduos, porém, garantindo que nenhum indivíduo obtenha uma quantidade negativo de recursos.

A regra Proporcional pode ser representada através da tupla $P(E, c) = \lambda c$, onde E representa a quantidade de recursos disponíveis que podem serem distribuídos, c o vetor de demanda que contém a quantidade de recursos reivindicados por cada indivíduo e $\lambda = \frac{E}{C}$, sendo $C = \sum_{i \in N} c_i$, ou seja, o montante demandado pelos dispositivos. A regra Proporcional é uma das soluções utilizadas neste trabalho para tratar o problema de falência. São propriedades presentes na regra Proporcional (HERRERO; VILLAR, 2001):

- Tratamento igual de iguais (*equal treatments of equals*): A mesma quantidade de recursos deve ser concedida aos indivíduos que apresentam reivindicações iguais.
- Invariância de escala ou Homogenia de grau um (*scale invariance*): A alocação independe da unidade utilizada. Por exemplo: a mesma regra utilizada para dividir um euro é utilizada para dividir mil euros.
- Composição (*Composition*): O ganho de um determinado indivíduo permanece inalterado diante da divisão dos recursos. Por exemplo: dado um patrimônio/recursos disponíveis R que foi dividido em duas partes R_1 e R_2 , tal que $R_1 + R_2 = R$, então, $P(R, c) = P(R_1, c) + P(R_2, c - P(R_1, c))$, sendo c o vetor de demanda ou reivindicações.
- Aditividade (*additivity*): A divisão nos recursos disponíveis/patrimônio e no vetor de demanda não devem alterar o resultado final da alocação dos recursos.
- Monotonicidade (*monotonic*): Variações positivas na quantidade de recursos disponíveis (patrimônio), não pode acarretar redução nos recursos alocados para os indivíduos. O mesmo vale para variações negativas, ou seja, redução nos recursos disponíveis não deve beneficiar nenhum indivíduo. Assim, se $R \geq R'$, então $P(E, c) \geq P(R', c)$, sendo R e R' os recursos disponíveis.

Dentre as três soluções apresentadas, a regra Proporcional é a mais conhecida e aplicada nos cenários de falência (HERRERO; VILLAR, 2001).

2.10 Trabalhos Relacionados

Esta subseção descreve os trabalhos relacionados as abordagens PClass, Gtoken, Ptoken, Gcronos e Pcronos propostas neste trabalho.

Em (LO et al., 2011), os autores propõem o mecanismo *Self-Optimizing Overload Control - SOOC* para alocar dinamicamente os *slots* disponíveis para o procedimento de RACH de acordo com o nível de congestionamento na rede de acesso via rádio. O nível de congestionamento é calculado com base no parâmetro *raRetryCounter* informado pelos dispositivos para a estação base durante o procedimento de RACH, mais especificamente após o recebimento da mensagem de RAR, ou seja, na terceira mensagem do RACH. O *raRetryCounter* é uma variável que armazena o número de tentativas de acesso realizadas pelo dispositivo. De posse deste parâmetro, a estação base infere o número médio de preâmbulos de PRACH enviados pelos dispositivos e o nível de congestionamento na rede. A partir do nível de congestionamento a estação base define se mais recursos devem ser reservados para o PRACH. À medida que mais dispositivos enviam solicitações mais colisões são esperadas e assim mais recursos são alocados para o PRACH. Os autores também propõem a reserva de recursos do RACH entre os dispositivos H2H, M2M e a criação de duas classes de dispositivos M2M. Em suma, o SOOC combina as técnicas *Dynamic Allocation of PRACH Resource*, *Access Class Barring*, *RACH resource allocation* e *p-persistent*, sendo esta última utilizada para evitar que os dispositivos fiquem ininterruptamente retransmitindo solicitações de acesso após uma colisão. Apesar de ser baseado nas estratégias propostas pela 3GPP, o SOOC apresenta alta complexidade de implementação. Como já mencionado, os mecanismos para controle de sobrecarga que alocam dinamicamente recursos para o PRACH também devem considerar o impacto sobre o tráfego de dados e vice-versa. Infelizmente os autores não realizam nenhuma análise da proposta, seja por simulação ou analiticamente. Deste modo, torna-se difícil analisar sua performance.

Em (JIAN et al., 2013), os autores propõem a criação de diferentes classes para os diferentes tipos de dispositivos, dispersando assim as (re-)tentativas de acordo com a classe do dispositivo. Os autores definem os dispositivos M2M em prioritários e não prioritários. A prioridade entre as classes é estabelecida segundo o tempo de *backoff*, sendo este calculado individualmente por cada dispositivo. Os autores propõem três estratégias para o cálculo do tempo de *backoff*. Na primeira, o tempo de *backoff* é calculado utilizando a equação:

$$T_{bf}^{MTC}(i) = T + Unif\left(i\frac{kT}{C}, (i+1)\frac{kT}{C}\right) \quad (2.7)$$

onde: i é a classe do dispositivo, T o intervalo máximo definido para os dispositivos H2H, C o número total de classes de dispositivos e k o parâmetro definido pelos autores para estender o intervalo total de *backoff* disponível para os dispositivos M2M. Deste modo, cada classe tem seu próprio intervalo de *backoff*. A segunda abordagem apresentada pelos autores para o cálculo

do tempo de *backoff* é dada por:

$$T_{bf}^{MTC}(i) = Unif(0, k_i T) \quad (2.8)$$

Em comparação com a primeira estratégia, nesta ocorre sobreposição dos intervalos de *backoff* dos dispositivos M2M. A abordagem utilizada pelos autores assemelha-se a apresentada pelo *backoff* específico, mas aplicada aos dispositivos M2M. A terceira estratégia apresentada altera a probabilidade de acesso dos dispositivos que passa a ser calculada segundo a equação:

$$P_{Ac}(i) = (C - i)\delta, \delta = \frac{(1 - p_0)P_{Ac}}{C} \quad (2.9)$$

onde $P_{Ac}, p_0 \in [0, 1]$, sendo P_{Ac} a probabilidade de acesso difundida pela eNodeB e p_0 o parâmetro utilizado para determinar o nível de congestionamento na rede. Deste modo, a probabilidade de acesso dos dispositivos H2H é de $100 \times P_{Ac} \%$ e a probabilidade de acesso de uma determinada classe i é dada por $100 \times P_{Ac}(i) \%$. Na primeira abordagem o tempo total de *backoff* é dividido de acordo com o número de classes, mas desconsidera a quantidade de dispositivos existente em cada classe. Assim, intervalos de *backoff* de mesmo tamanho são atribuídos às classes independentemente da quantidade de dispositivos. A segunda estratégia é uma variação do *backoff* específico e como tal apresenta intervalos que se sobrepõem. Na terceira abordagem os autores adaptam o ACB variando a probabilidade de acesso dos dispositivos. No entanto, quanto maior o número de dispositivos M2M maior será o número de solicitações e, conseqüentemente, menor será a probabilidade de acesso dos dispositivos, inclusive dos dispositivos H2H. Além disso, nenhuma dessas abordagens estabelece priorização intra-classe para os dispositivos M2M.

Em (PANG et al., 2014), os autores se baseiam na proposta de reserva de recursos do RACH para os diferentes tipos de dispositivos. Os autores propõem a criação de três *pools* de preâmbulos, sendo um para os dispositivos H2H (R_H), um para os dispositivos M2M (R_M) e um *pool* híbrido (R_B) cujos recursos podem ser utilizados por dispositivos H2H ou M2M. Para identificar como os recursos do *pool* híbrido devem ser utilizados, os autores modelam a disputa pelos recursos como um de jogo de estratégia mista. Os autores consideram um tráfego fixo de dispositivos H2H transmitindo no *pool* híbrido e definem três ações possíveis para os dispositivos M2M, sendo elas: utilizar o *pool* M2M ($a_i = M$), utilizar o *pool* híbrido ($a_i = B$) ou ficar em silêncio ($a_i = S$), ou seja, não transmitir o preâmbulo. Os jogadores/dispositivos M2M têm a probabilidade de suas ações determinadas pela Equação 2.10.

$$\sigma_i(a_i) = \begin{cases} P_{i,M} & \text{if } a_i = M, \\ P_{i,B} & \text{if } a_i = B, \\ 1 - P_{i,M} - P_{i,B} & \text{if } a_i = S \end{cases} \quad (2.10)$$

onde $P_s(a_i)$ é a probabilidade de obtenção de sucesso do procedimento de RACH quando o dispositivo i realiza a ação a_i . O pagamento ou custo para transmitir um preâmbulo é dado por $\lambda \in [0, 1]$, que descreve a complacência de um dispositivo de economizar energia. O *payoff* de cada jogador/dispositivo é dado por:

$$g_i(t) = \begin{cases} 1 - \lambda & \text{if transmission succeeds,} \\ -\lambda & \text{if transmission fails,} \\ 0 & \text{no transmission.} \end{cases} \quad (2.11)$$

As abordagens propostas pelos autores apresentam vantagens em relação à alocação fixa de recursos no RACH apresentada pela 3GPP. No entanto, os autores propõem a alocação de uma quantidade fixa de preâmbulos para os dispositivos H2H e M2M e não definem nenhuma priorização entre os dispositivos M2M. Na modelagem do jogo, os autores não consideram os dispositivos H2H como potenciais jogadores, ou seja, a utilização dos recursos do *pool* híbrido é analisada somente a partir da perspectiva dos dispositivos M2M.

Em (LEE; KIM; YI, 2011), os códigos de preâmbulos são divididos entre os dispositivos H2H e M2M. Os autores apresentam duas estratégias para divisão desses preâmbulos. Na primeira, os preâmbulos são divididos em duas partes disjuntas para os dispositivos das classes H2H e M2M. Na segunda, os preâmbulos também são divididos entre os dispositivos H2H e M2M. No entanto, ao contrário do que ocorre no primeiro método, os preâmbulos antes alocados somente para o M2M também podem ser utilizados pelos dispositivos H2H. Combinada a estratégia de divisão de recursos do RACH, os autores também utilizam o *Access Class Barring* (ACB) para restringir, dentre os dispositivos H2H e M2M, quais dispositivos terão acesso aos recursos do RACH. Os autores não estabelecem nenhum tipo de priorização entre os dispositivos M2M nesta estratégia.

Em (LIEN et al., 2012), os autores propõem uma estratégia para o cálculo do parâmetro que define a probabilidade de acesso (p) do *Access Class Barring* (ACB). No ACB cada eNodeB calcula individualmente o parâmetro probabilidade de acesso. Contudo, os autores argumentam que diante de cenários mais severos de congestionamento a eNodeB atribui valores cada vez menores para a probabilidade de acesso (p), resultando, assim, no aumento do tempo de acesso dos dispositivos M2M. Devido a sua arquitetura de rede *multi-tier*, tem-se, no LTE, as *picocells*², que são cobertas pela *macrocell*³, as quais além do ganho na potência do sinal também ajudam a reduzir a carga sobre a *macrocell*. Com base nesta arquitetura, os autores propõem que o cálculo da probabilidade de acesso, difundida pela estação base, seja realizando levando em consideração o nível de congestionamento das estações base próximas. O algoritmo pode apresentar problemas em cenários onde não houver *picocells*. Durante a avaliação de desempenho da proposta os autores restringem a análise aos dispositivos M2M, ou seja, não analisam o impacto sobre os dispositivos H2H. Os autores também não definem nenhum tipo de priorização entre os dispositivos M2M, ou seja, todos os dispositivos M2M são tratados de modo homogêneo.

Em (HASAN; HOSSAIN; NIYATO, 2013), os autores propõem um mecanismo para seleção de estação base para os dispositivos MTC. Baseado no procedimento de acesso aleatório do LTE, os dispositivos identificam o tempo de atraso das estações base para transmissão de pacotes. A partir dessas informações, o dispositivo pode identificar e escolher qual

²Estação base celular com baixa área de cobertura.

³Estação base celular com grande área de cobertura.

estação base melhor atende aos seus requisitos de qualidade de serviço (*Quality of Service - QoS*). Para auxiliar os dispositivos no processo de escolha da estação base, os autores propõem um algoritmo de aprendizagem por reforço (i.e., *Q-learning*). Uma das desvantagens da proposta é que esta considera que os dispositivos M2M estão na área de cobertura de mais de uma eNodeB. Os autores não fazem distinção entre os dispositivos M2M. Além disso, não constam dispositivos H2H nos cenários de simulação utilizados.

Em (JIANG et al., 2014), os autores apresentam um mecanismo para aprimorar o cálculo do parâmetro de probabilidade de acesso (p) do *Access Class Barring* (ACB). O cenário e o problema descritos pelos autores assemelham-se ao apresentado em (LIEN et al., 2012), ou seja, a necessidade de adequar o parâmetro (p) para que diante de cenários mais severos de congestionamento a carga gerada pelas solicitações de acesso fique distribuída entre as estações base. Os autores propõem uma solução utilizando jogos evolucionários. Primeiramente os autores estimam o número de dispositivos que estão aguardando para iniciar uma solicitação de acesso à rede. Em um segundo momento, a demanda dos dispositivos é distribuída entre as diferentes estações base que calculam individualmente a probabilidade de acesso. O algoritmo pode apresentar problemas em cenários onde não houver *picocells*. Os autores restringem a análise aos dispositivos M2M, ou seja, não analisam o impacto sobre os dispositivos H2H durante a avaliação de desempenho da proposta. Os autores também não definem nenhum tipo de priorização entre os dispositivos M2M, ou seja, todos os dispositivos M2M são tratados de modo homogêneo.

3 PROPOSTA DE CONTROLE DE ACESSO AOS RECURSOS DO CANAL

3.1 Mecanismo PClass

O mecanismo PClass é implementado em dois pontos da rede LTE, na estação base e nos dispositivos. O algoritmo executado na estação base tem como objetivo identificar o nível de congestionamento na RAN, ver Algoritmo 1. O segundo algoritmo é executado individualmente por todos os dispositivos e determina a probabilidade de transmissão do preâmbulo pelo dispositivo.

Algoritmo 1: ALGORITMO PARA IDENTIFICAR O NÍVEL DE CONGESTIONAMENTO NA RAN DO LTE.

Entrada: número de solicitações de acesso recebidas e os parâmetros de referência

$$a = 25 \text{ e } b = 50$$

Saída: retorna o nível de congestionamento (baixo, médio, alto)

```

1 enquanto ler todos os códigos de preâmbulos recebidos faça
2   | conta o número de solicitações recebidas;
3 fim
4 se número de solicitações  $\leq a$  então
5   | retorna nível de congestionamento baixo;
6 fim
7 senão se  $a > \text{número de solicitações} \leq b$  então
8   | nível de congestionamento médio;
9 fim
10 senão
11   | nível de congestionamento alto;
12 fim

```

O algoritmo implementado na estação base é utilizado para identificar o nível de congestionamento na RAN e se baseia nos resultados apresentados em (CHENG et al., 2012). Nesse artigo, os autores estabelecem uma relação entre o nível de utilização dos preâmbulos, a probabilidade de colisão e a probabilidade de envio com sucesso dos preâmbulos de PRACH. A partir dos resultados apresentados, percebe-se que para manter o nível de utilização máximo de 50% e a probabilidade de colisão em aproximadamente 22%, o número de solicitações por *RA-slot* é aproximadamente igual ao número máximo de códigos de preâmbulos disponíveis. Ainda com base nos resultados supracitados e número de preâmbulos disponíveis para o procedimento de RACH, considera-se o nível de congestionamento médio quando o número de solicitações estiver entre 25 e 50. O nível de congestionamento é considerado baixo quando o número de solicitações recebidas pela estação base for menor do que 25 e alto quando for maior do que 50 solicitações.

A priorização estabelecida entre os dispositivos pelo mecanismo PClass decresce com o aumento do índice i , ou seja, os dispositivos pertencentes a classe de índice $i = 0$ têm prioridade maior do que os dispositivos da classe $i = 1$, e assim sucessivamente. Nos dispo-

sitivos, o algoritmo responsável por identificar e calcular a probabilidade de transmissão de um preâmbulo de solicitação, mensagem 1 do procedimento de RACH, baseia-se na classe do dispositivo (i), na quantidade de tentativas de acesso já realizadas (t), no número máximo de tentativas permitidas (L), no nível de congestionamento na RAN e na probabilidade de acesso difundida pela estação base (p_{ac}), como ilustrado na Equação 3.1.

$$P_s(i, t, L) = \begin{cases} p_{ac} & \text{para } P_{cong} = 0.0; \\ p_{ac} \times (((\frac{L}{t}) \times i) \times \alpha) & \text{para } P_{cong} = 0.5; \\ p_{ac} \times (((\frac{L}{t}) \times i) \times \alpha) & \text{para } P_{cong} = 1.0; \end{cases} \quad (3.1)$$

O procedimento de RACH é iniciado pelo dispositivo quando este deseja se conectar à rede (cf. Seção 2.5.1). O algoritmo é executado pelo dispositivo antes do envio do preâmbulo. Diferentemente do procedimento de RACH padrão, onde o dispositivo envia uma solicitação tão logo deseje acessar a rede, no mecanismo proposto neste trabalho uma nova etapa antecede o processo de envio do preâmbulo. Antes de enviar uma solicitação cada dispositivo calcula a probabilidade de envio do preâmbulo com base na Equação 3.1. Dada a classe do dispositivo (i), o número de tentativas de acesso já realizadas pelo dispositivo (t) e o nível de congestionamento na RAN (P_{cong}) cada dispositivo apresenta diferentes valores para a probabilidade de envio de preâmbulo.

O algoritmo proposto altera a probabilidade de transmissão do preâmbulo com base no tipo de dispositivo, garantindo que determinadas classes de dispositivos tenham maior ou menor probabilidade de transmissão de preâmbulos de solicitação de acesso. O número de tentativas de acesso possibilita a priorização intraclasse, ou seja, entre os dispositivos que pertencem a uma mesma classe. Os dispositivos que estão próximos de atingir a quantidade máxima de transmissões permitidas pela rede têm maior prioridade de transmissão. O nível de congestionamento informado pela RAN afeta proporcionalmente todas as classes de dispositivos através da alteração dos valores atribuídos ao α . Pelos resultados obtidos através de simulações, os melhores valores obtidos para α , quando $P_{cong} = 0.5$ e $P_{cong} = 1.0$ são respectivamente $\alpha = 1.0$ e $\alpha = 1.5$. O congestionamento pode ser classificado em baixo, médio e alto e o algoritmo proposto para os dispositivos varia de acordo com esses três possíveis níveis de congestionamento, como apresentado no Algoritmo 3.1.

Antes de transmitir uma mensagem de solicitação de acesso o dispositivo gera um valor aleatório $X \sim Unif(0, P_s)$, sendo P_s calculado através da Equação 3.1 e $Unif(0, P_s)$ uma distribuição uniforme no intervalo $(0, P_s)$. De posse da variável aleatória X e do parâmetro p_{ac} , propagado pela estação base, o dispositivo verifica se X é menor do que p_{ac} . Se sim, o dispositivo continua o procedimento de acesso ao canal enviando uma solicitação de acesso. Caso contrário, o dispositivo entra no estado de espera e só poderá enviar uma nova solicitação depois de passado o intervalo de espera. O valor do intervalo de espera ou valor de *backoff*, $T_{backoff}$, varia de acordo com a classe do dispositivo e é calculado individualmente para cada dispositivo através da Equação 3.2. O funcionamento do mecanismo PClass está representado na Figura 3.3.

$$T_{backoff}(i) = \begin{cases} 20ms & \text{para } i = 0; \\ 50ms \times i & \text{para } i \geq 1; \end{cases} \quad (3.2)$$

A função (F_X) de densidade acumulada (*Cumulative Density Function* - CDF) do dispositivo k da classe i após t tentativas de acesso é dada pela Equação 3.3. Assim, fixando-se p_{ac} e variando o X , tem-se que quando maior o valor de P_s menor será a probabilidade de transmissão do preâmbulo para uma dada classe de dispositivo.

$$F_X(p_{ac}) = P(X \leq p_{ac}) = \frac{p_{ac}}{P_s} \quad (3.3)$$

Sabendo que o P_s também está relacionado com o número de tentativas de acesso realizadas pelo dispositivo, tem-se que a probabilidade de envio de solicitação do dispositivo k da classe i , que está solicitando acesso pela t -ésima vez, aumenta proporcionalmente com t e k e de modo inversamente proporcional em relação a i . Contudo, tem-se o aumento do número de colisões quando há muitos dispositivos solicitando acesso. Em caso de colisão, o intervalo de *backoff* é calculado seguindo a estratégia convencional do LTE, ou seja, $T_{backoff} \sim Uniforme(0, BI)$, onde BI (*Backoff Indicator*) é informado pela estação base, mas precisamente na segunda mensagem do procedimento de RACH.

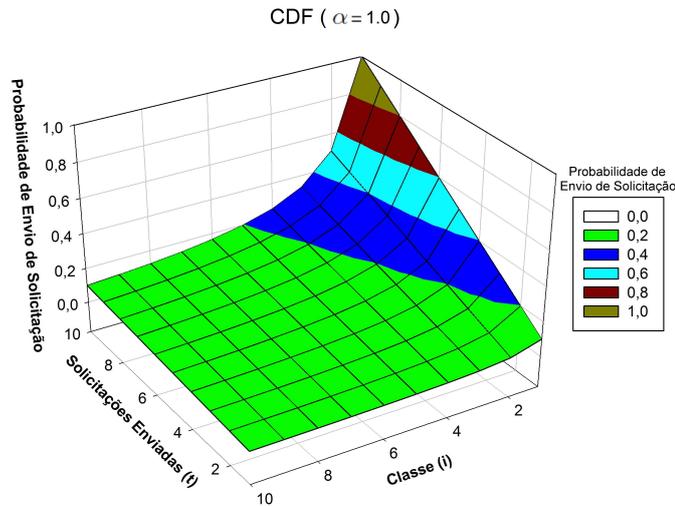


Figura 3.1: CDF quando $\alpha = 1.0$.

As Figuras 3.1 e 3.2 ilustram como a probabilidade de envio de preâmbulo varia segundo os parâmetros: classe de dispositivo, número de transmissões e nível de congestionamento na RAN.

Vale ressaltar que as condições para o envio de preâmbulo são analisadas somente pelos dispositivos M2M, ou seja, a Equação 3.1 não se aplica aos dispositivos da classe H2H. Assim, independentemente do número de dispositivos M2M e nível de congestionamento na rede, não há alterações nas etapas do procedimento de RACH para os dispositivos H2H.

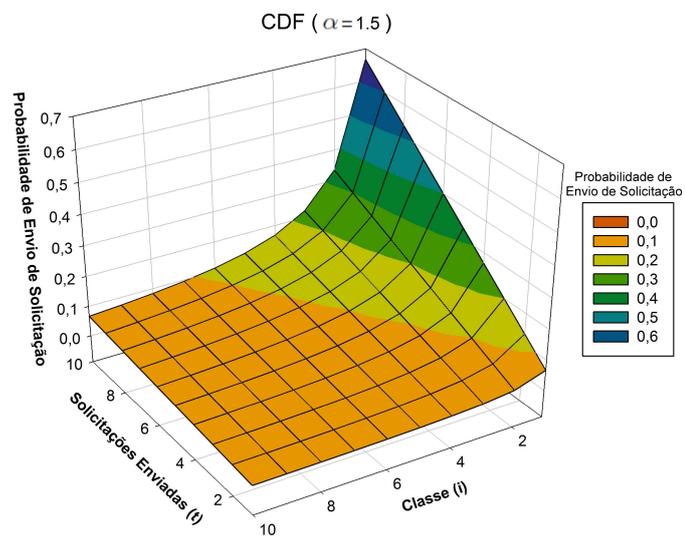


Figura 3.2: CDF quando $\alpha = 1.5$.

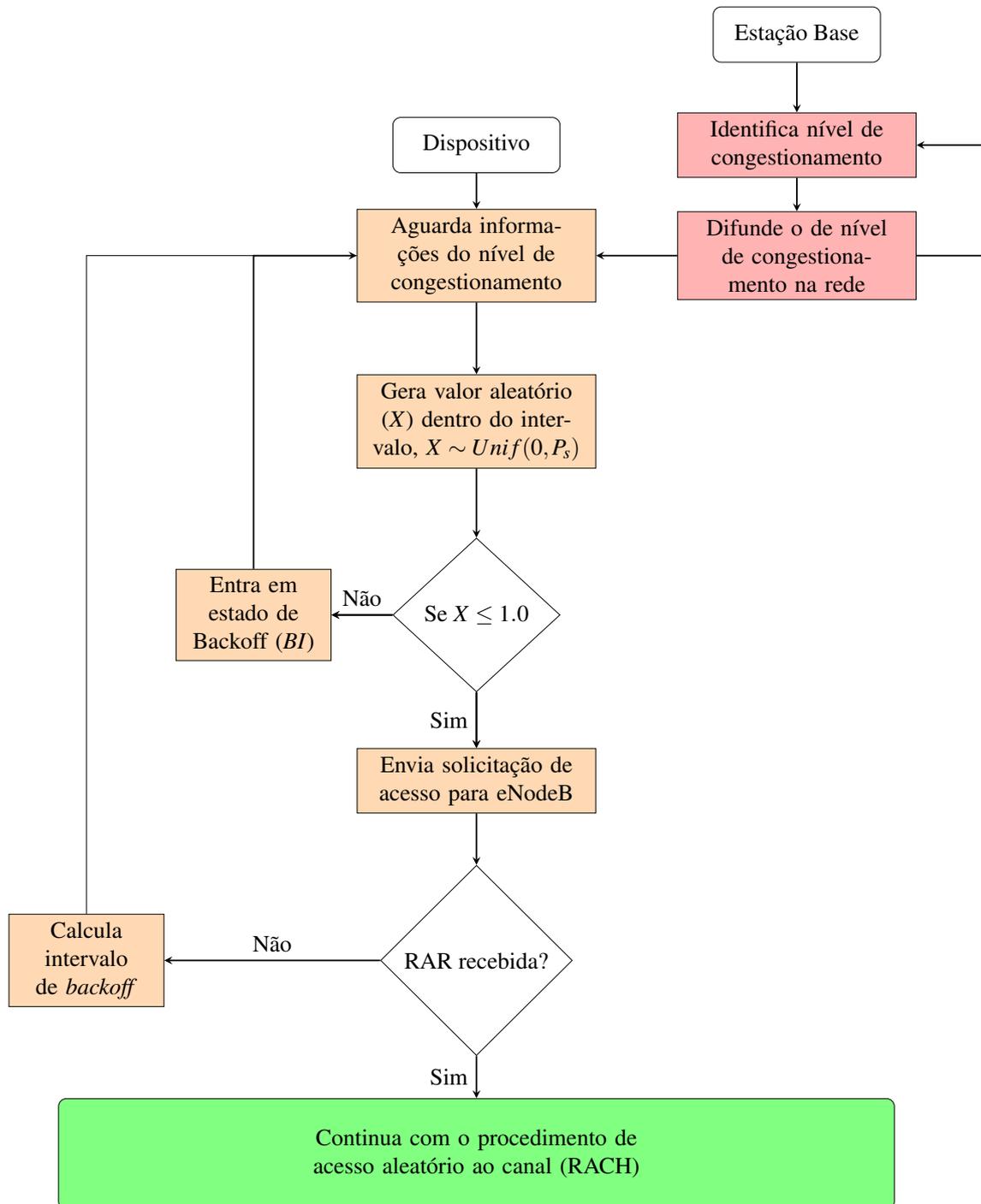


Figura 3.3: Etapas do mecanismo PClass.

3.1.1 Análise dos Resultados

Os resultados apresentados nesta seção comparam o desempenho do mecanismo PClass com outros três mecanismos. Dentre os mecanismos comparados tem-se o Slotted Aloha, o Backoff Específico (ZTE, 2011) e a proposta de Classes de Backoff apresentada em (JIAN et al., 2013) e serão referenciados respectivamente por “SlotAloha”, “BE” e “Jian”. Os níveis de congestionamento dos cenários considerados variam de baixo, médio e alto.

Os Indicadores-Chaves de Performance (*Key Performance Indicators - KPI's*) são: a latência no acesso, a probabilidade de bloqueio, o total de solicitações de acesso transmitidas e o total de acessos realizados com sucesso. Sendo a latência no acesso o intervalo de tempo transcorrido entre o instante em que o dispositivo envia a primeira solicitação de acesso até o momento de recebimento da mensagem de contenção, ou seja, a quarta mensagem do procedimento de RACH. O total de solicitações de acesso contabiliza as solicitações realizadas pelo dispositivo. Quanto maior o número de solicitações transmitidas, maior o consumo energético, pois as atividades de transmissão e recepção por rádio normalmente demandam uma quantidade significativa de recursos energéticos.

3.1.1.1 Controle do Impacto sobre Dispositivos H2H

A variação da probabilidade de bloqueio em relação a quantidade de dispositivos mostrada na Figura 3.4 indica que para 5100 dispositivos o algoritmo PClass apresenta performance melhor do que Jian e SloAloha de, respectivamente, 19% e 40% aproximadamente. Para 1350 dispositivos, percebe-se que o algoritmo de Jian apresenta desempenho máximo de aproximadamente 10% em relação ao PClass. Contudo, com o aumento do número de dispositivos, tem-se uma redução na performance do algoritmo Jian. Entre o intervalo de 1800 a 2000 dispositivos percebe-se que há uma inversão da probabilidade de bloqueio, onde o PClass passa a apresentar probabilidade de bloqueio menor do que o algoritmo de Jian. Nos cenários com mais dispositivos a probabilidade de bloqueio do mecanismo PClass continua menor do que a do mecanismo Jian.

A inversão, apresentada na Figura 3.4, que ocorre próxima dos 1975 dispositivos está relacionada ao nível de congestionamento (P_{cong}), detalhado na Seção 3.1. Nos cenários onde o nível de congestionamento é moderado mais dispositivo M2M enviam requisições de acesso à rede. Consequentemente, tem-se o aumento do número de colisões e impacto no acesso do H2H.

Como ilustrado na Figura 3.5, o tempo médio de acesso permanece praticamente constante em todos os algoritmos simulados. Contudo, vale ressaltar que o cálculo do tempo médio de acesso considera somente os dispositivos que conseguiram acesso à rede com sucesso. Assim, o tempo médio de acesso dos dispositivos M2M permanece constante, mas, em contrapartida, tem-se a diminuição do número de dispositivos que conseguem acessar a rede, como ilustrado na Figura 3.4. Devido ao modo como o algoritmo Jian cria as classes de *backoff*, tem-se a elevação do tempo de acesso para os dispositivos H2H.

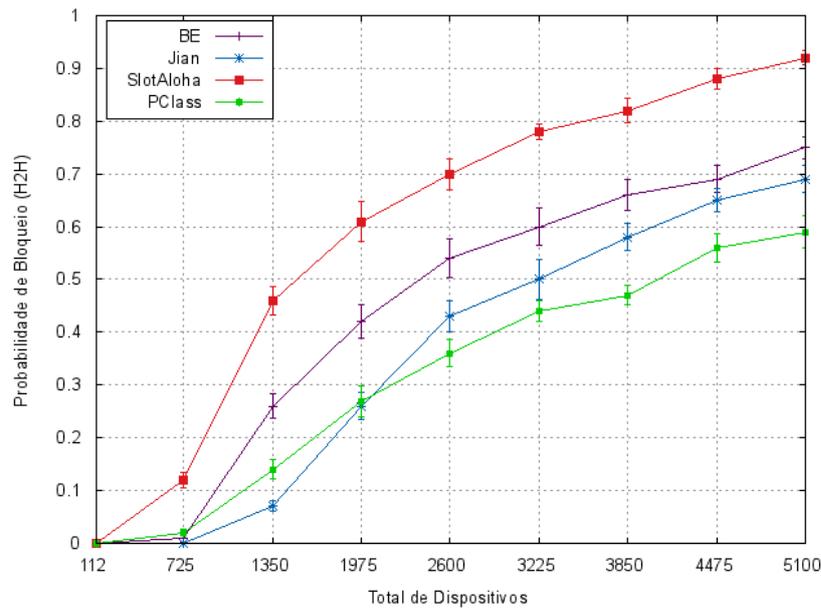


Figura 3.4: Probabilidade de Bloqueio.

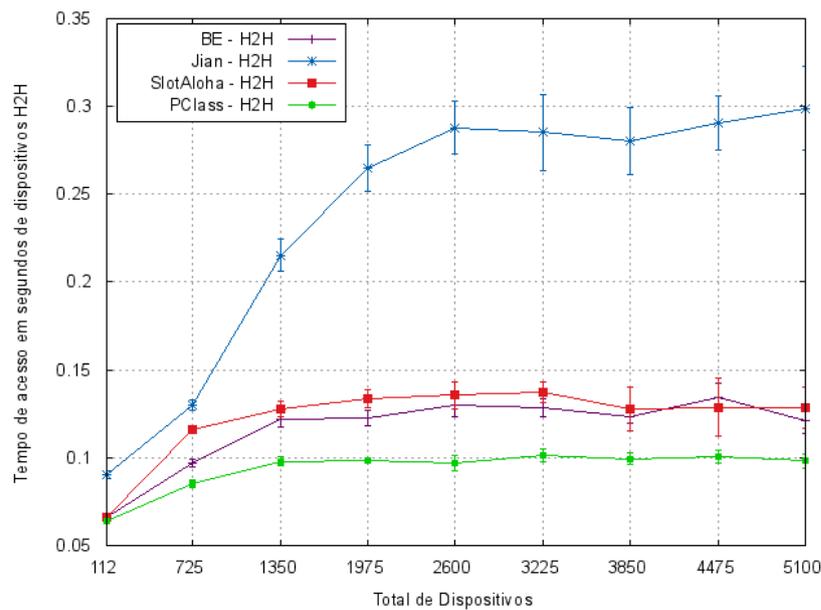


Figura 3.5: Tempo Médio de Acesso.

3.1.1.2 Priorização entre Dispositivos M2M

No caso dos dispositivos M2M, a probabilidade de bloqueio aumenta com o número de dispositivos em todas as propostas analisadas, como apresentada na Figura 3.6. Os algoritmos SlotAloha e BE não estabelecem priorização entre os dispositivos M2M, penalizando igualmente todos os dispositivos M2M. Comparados entre si, a diferença existente entre o SlotAloha e o BE ocorre devido à diferença nos valores de *backoff*.

No SlotAloha, os dispositivos M2M são dispersados em intervalos de tempo de 20ms e no BE em intervalos de 100ms. Um resultado imediato da dispersão é o aumento no

tempo de acesso. Contudo, tem-se a redução da probabilidade de bloqueio. Os resultados apresentados por Jian são melhores do que o SlotAloha e BE, pois estes não apresentam priorização entre os dispositivos M2M. Em relação ao PClass, o Jian apresenta melhora na performance de aproximadamente 12% na probabilidade de bloqueio entre dispositivos M2M de alta prioridade para cenários com aproximadamente 725 a 3530 dispositivos. Porém, esse quadro se inverte quando o número de dispositivos M2M é superior a 3530. Em relação ao M2M de baixa prioridade a estratégia adotada em Jian comporta-se como o SlotAloha e o BE, ou seja, o tempo de acesso aumenta gradativamente com o aumento do número de dispositivos.

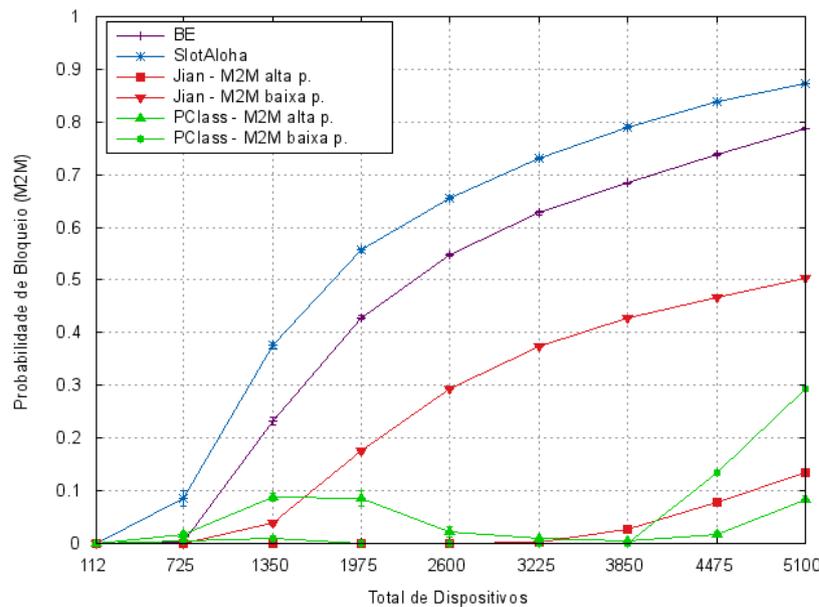


Figura 3.6: Probabilidade de Bloqueio.

O mecanismo PClass apresenta a mesma performance do mecanismo Jian quando comparado com o SlotAloha e BE em relação a priorização entre os dispositivos M2M. Percebe-se que o mecanismo PClass excede o número de tentativas de acesso nos cenários menos congestionados, ou seja, aqueles em que há poucos dispositivos (e.g., Fig. 3.6 com 1350 dispositivos M2M). Porém, à medida que o nível de congestionamento aumenta o PClass melhora. O comportamento apresentado pelo mecanismo PClass para cenários com 1350 a 3850 dispositivos deve-se ao parâmetro (P_{cong}), pois este repercute na probabilidade de transmissão de préambulo dos dispositivos, como apresentado nas Equações 3.1 e 3.3.

3.1.1.3 Acessos Realizados com Sucesso

Através das Figuras 3.8 e 3.9 é possível verificar o impacto gerado nos dispositivos H2H com o aumento do número de dispositivos M2M. Os mecanismos Jian e PClass apresentam comportamento parecido, contudo, mais dispositivos H2H conseguem acessar a rede com o PClass do que com o Jian, como apresentado na Figura 3.8. Nos cenários em que o número total de dispositivos simulados é maior do que 1975, o total e o tempo de acesso dos dispositivos H2H do mecanismo PClass apresentam resultados melhores do que todas as outras estratégias simuladas (cf. Figura 3.5).

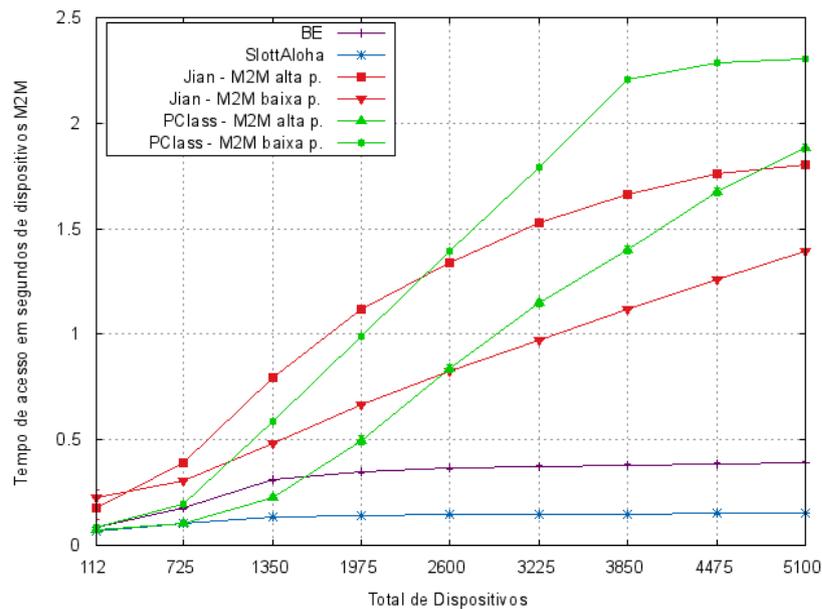


Figura 3.7: Tempo Médio de Acesso.

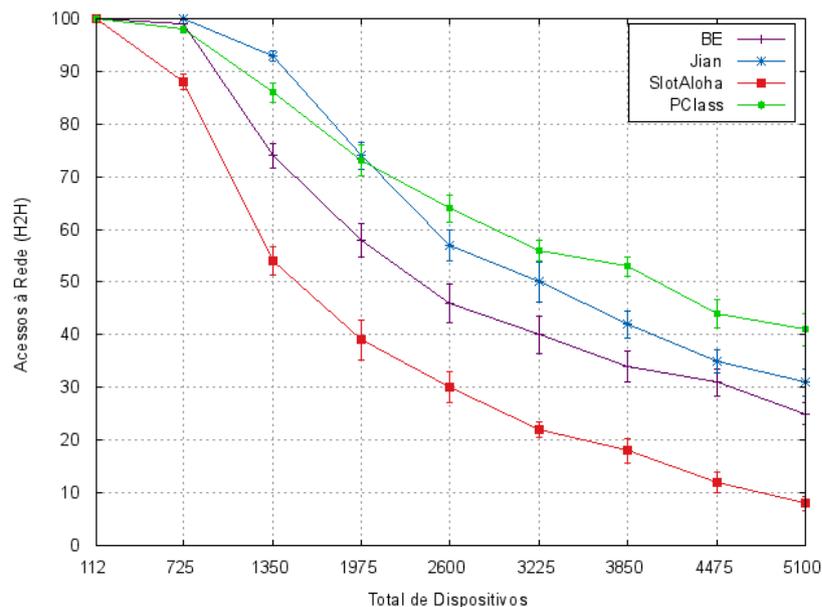


Figura 3.8: Dispositivos H2H.

Como apresentado na Figura 3.9, para os dispositivos M2M de alta prioridade o PClass e o Jian apresentam poucas diferenças em relação ao total de acessos e ambos conseguem garantir que quase todos os dispositivos M2M de baixa prioridade efetuem acesso à rede com sucesso. Porém, quando o total de acessos é analisado em conjunto com o tempo de acesso dos dispositivos M2M de alta prioridade (cf. Figura 3.7) percebe-se que PClass apresenta tempo de acesso predominantemente menor do que Jian. Ou seja, o mecanismo PClass mantém o número de acesso igual ao de Jian sem afetar no tempo de acesso. Para os dispositivos M2M de baixa prioridade, o PClass garante mais acessos do que Jian, mas apresenta tempo de acesso maior do que Jian.

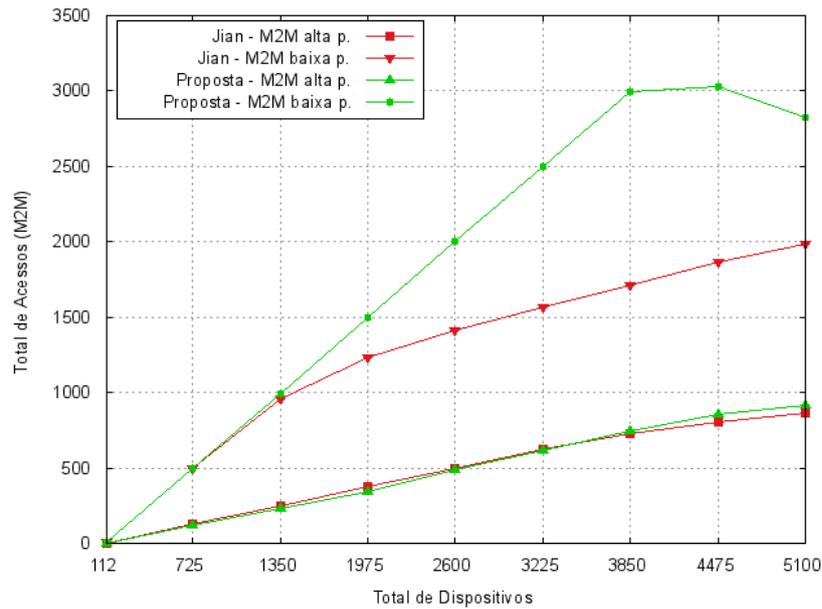


Figura 3.9: Dispositivos M2M de alta e baixa prioridade.

3.1.1.4 Eficiência Energética

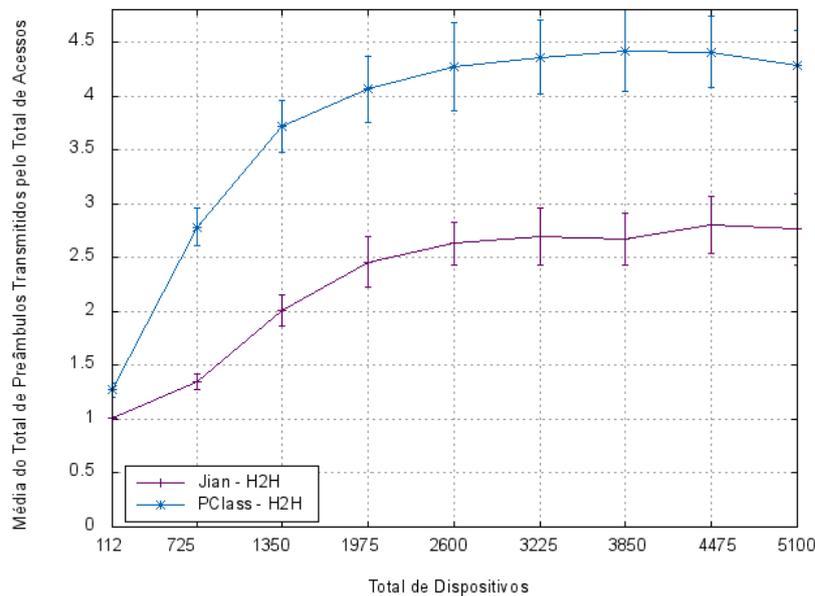


Figura 3.10: Preâmbulos enviados pelos dispositivos H2H.

As Figuras 3.10 e 3.11 apresentam a quantidade total de solicitações de acesso transmitidas pelos dispositivos H2H e M2M para os mecanismos PClass e Jian. A quantidade de preâmbulos transmitidos pelos dispositivos H2H no mecanismo PClass é maior do que a quantidade transmitida no Jian, como ilustrado na Figura 3.10. A redução na quantidade de preâmbulos transmitidos pelos dispositivos M2M, que ocorre a partir dos 2600 dispositivos, está relacionado com o nível de congestionamento na rede, pois quando o nível de congestionamento está elevado o mecanismo PClass diminui a probabilidade de envio de solicitações

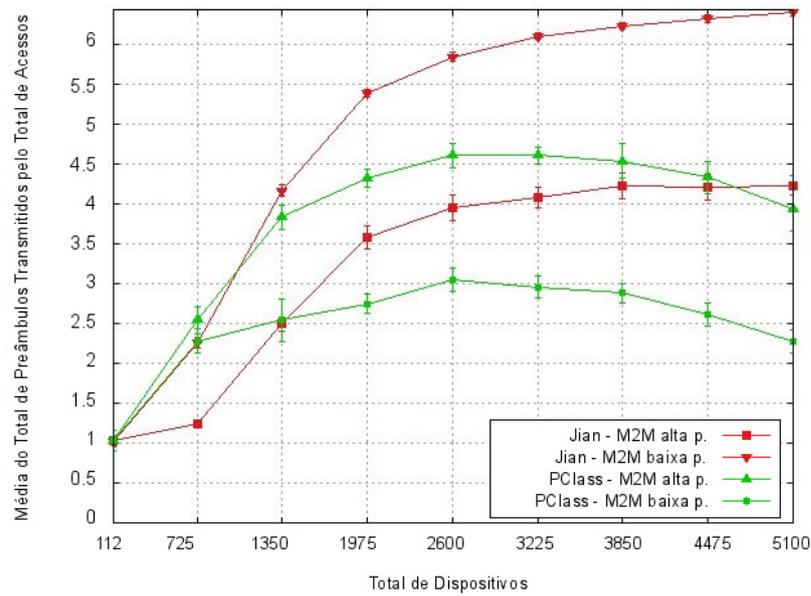


Figura 3.11: Preâmbulos enviados pelos dispositivos M2M.

dos dispositivos. Como o número de dispositivos M2M é maior do que a dos dispositivos H2H, pode-se dizer que de modo geral o PClass economiza mais energia do que o Jian. Além disso, o caráter autônomo apresentado pelos dispositivos M2M torna o gerenciamento dos recursos energéticos nesses dispositivos um fator crítico para a viabilização de determinadas aplicações, pois dependendo do tipo da aplicação (e.g., monitoramento ambiental, cidades inteligentes, etc.), determinados dispositivos podem ter baterias como fonte de energia primária.

4 PROPOSTAS DE ALOCAÇÃO DINÂMICA DE RECURSOS DO CANAL

Este capítulo apresenta os mecanismos Gtoken, Ptoken, Gcronos e Pcronos propostos neste trabalho para o controle do congestionamento ocasionado pelo excesso de dispositivos na RAN do LTE durante o procedimento de RACH. Os mecanismos Gtoken e Ptoken propõem uma estratégia para alocação de recursos de RACH de modo dinâmico, ou seja, os recursos são alocados de acordo com a necessidade e prioridade de cada classe de dispositivo. Os mecanismos Gcronos e Pcronos se baseiam em uma estratégia de *backoff*, onde intervalos de *backoff* disjuntos entre si são atribuídos as diferentes classes de dispositivos.

4.1 Mecanismos de Alocação Dinâmica de Códigos de Preâmbulos

As propostas apresentadas nesta seção se baseiam na estratégia da 3GPP de separação dos recursos do RACH para os dispositivos M2M e H2H. No entanto, não é abordado pela 3GPP como esses recursos devem ser divididos entre as diferentes classes de dispositivos. Além disso, as estratégias originalmente apresentadas pela 3GPP não estabelece nenhum tipo de priorização entre os dispositivos M2M, ou seja, todos os dispositivos são tratados de modo homogêneo.

A colisão de preâmbulos durante o procedimento de acesso aleatório ao canal aumenta à medida que mais dispositivos solicitam acesso à rede, como apresentado na Seção 2.7. O canal não possui recursos suficientes para atender ao elevado número de solicitações e conseqüentemente tem-se a sobrecarga e o congestionamento na rede de acesso via rádio. Os mecanismos Gtoken e Ptoken apresentados nesta seção controlam o impacto ocasionado pelo excesso de dispositivos M2M sobre os dispositivos H2H através da separação dos códigos de preâmbulos utilizados durante o procedimento de acesso aleatório. Com a separação dos recursos têm-se grupos distintos de preâmbulos associados às diferentes classes. Assim, as eventuais colisões ficam restritas aos dispositivos de uma mesma classe. Com isso, a probabilidade de colisão deixa de estar associada somente ao número de dispositivos, mas também a quantidade de códigos de preâmbulos disponíveis para cada classe de dispositivo.

Baseado na quantidade de códigos de preâmbulos disponíveis para o procedimento de RACH e número de dispositivos, tem-se que o número esperado de preâmbulos escolhidos por somente um dispositivo pode ser calculado como descrito. Seja X uma variável aleatória que segue a distribuição de Bernoulli, ou seja, pode assumir os valores 0 ou 1.

Vamos indicar por $X = 1$ se um dado preâmbulo foi escolhido por somente um dispositivo; caso contrário $X = 0$. Tem-se que o número de preâmbulos escolhidos por somente um dispositivo pode ser dado por:

$$Y = X_1 + X_2 + X_3 + \dots + X_M \quad (4.1)$$

Onde M representa o número total de códigos de preâmbulos disponíveis. Assim:

$$E[Y] = E[X_1 + X_2 + X_3 + \dots + X_M] \stackrel{(a)}{=} E[X_1] + E[X_2] + E[X_3] + \dots + E[X_M] \quad (4.2)$$

(a) pela propriedade da linearidade da esperança. Por serem eventos independentes, tem-se que:

$$E[X_1] = E[X_2] = E[X_3] = \dots = E[X_M] \quad (4.3)$$

Logo:

$$E[Y] = M \times E[X_1] \quad (4.4)$$

Seja n a quantidade de dispositivos competindo pelos preâmbulos, a probabilidade de que apenas um preâmbulo seja escolhido por um dos n dispositivos é dada por:

$$\begin{aligned} P_{sucesso} &= \binom{n}{1} p(1-p)^{n-1} \\ &= np(1-p)^{n-1} \end{aligned} \quad (4.5)$$

Onde $\binom{n}{1}$ identifica os n dispositivos que competem pelo preâmbulo com probabilidade p e $(1-p)^{n-1}$ é a probabilidade dos outros dispositivos não escolherem o preâmbulo. Tem-se que a probabilidade máxima de sucesso na escolha do preâmbulo é máxima quando $\frac{d}{dp}P_{sucesso} = 0$, ou seja:

$$\begin{aligned} -np(1-n)(1-p)^{n-2} + n(1-p)^{n-1} &= 0 \\ np(n-1)(1-p)^{n-2} &= n(1-p)^{n-1} \\ p(n-1) &= 1-p \\ np-p &= 1-p \\ p &= \frac{1}{n} \end{aligned} \quad (4.6)$$

Substituindo p na Equação 4.5:

$$\begin{aligned} P_{sucesso} &= np(1-p)^{n-1} \\ &= \left(1 - \frac{1}{n}\right)^{n-1} \end{aligned} \quad (4.7)$$

Desenvolvendo a Equação 4.7 temos:

$$\begin{aligned}
P_{sucesso} &= \left(1 - \frac{1}{n}\right)^{n-1} \\
&= \frac{(n-1)^{n-1}}{n^{n-1}} \\
&= n^{-(n-1)}(n-1)^{n-1} \\
&= n^{-(n-1)} \sum_{k=0}^{n-1} \binom{n-1}{k} n^{n-1-k} (-1)^k \\
&= \sum_{k=0}^{n-1} \binom{n-1}{k} n^{-k} (-1)^k \\
&= \sum_{k=0}^{n-1} n^{-k} (-1)^k \frac{(n-1)!}{k!(n-1-k)!} \\
&= \sum_{k=0}^{n-1} \frac{(-1)^k}{k!} \frac{(n-1)!}{n^k(n-1-k)!}
\end{aligned} \tag{4.8}$$

Considerando uma análise assintótica, ou seja, $n \rightarrow \infty$ tem-se:

$$\begin{aligned}
\lim_{n \rightarrow \infty} P_{sucesso} &= \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} \frac{(-1)^k}{k!} \frac{(n-1)!}{n^k(n-1-k)!} \\
&= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \\
&\stackrel{(a)}{=} e^{-1} \approx 0.37
\end{aligned} \tag{4.9}$$

Sendo o passo (a) dado pela definição da série de Taylor para e^x . Assim, a eficiência máxima de utilização dos preâmbulos é de 37%. Em um cenário onde há M preâmbulos disponíveis, o número esperado de preâmbulos escolhidos por somente um dispositivo é dada por: $E[Y] = n(1 - \frac{1}{M})^{n-1}$, onde $p = \frac{1}{M}$. Sabendo que a performance máxima do sistema é obtida quando $M = n$, se há n dispositivos solicitando acesso à rede a quantidade média de preâmbulos necessária para satisfazer esta demanda pode ser dada por: $n \times \frac{1.0}{\delta}$, sendo $\delta = 0.37$.

A separação dos preâmbulos entre as diferentes classes de dispositivos implica divisão dos códigos (M) entre essas diferentes classes de dispositivos. Deste modo, é possível verificar que a separação dos preâmbulos influencia na probabilidade de colisão e consequentemente no número esperado de preâmbulos escolhidos por somente um dispositivo, ou seja, preâmbulos recebidos com sucesso. As faixas de preâmbulos são atribuídas as diferentes classes de dispositivos, restringindo as colisões entre os dispositivos de uma mesma classe. Através da Figura 4.1 é possível observar como o número esperado de preâmbulos escolhidos por somente um dispositivo varia com a quantidade de solicitações e número de preâmbulos disponíveis.

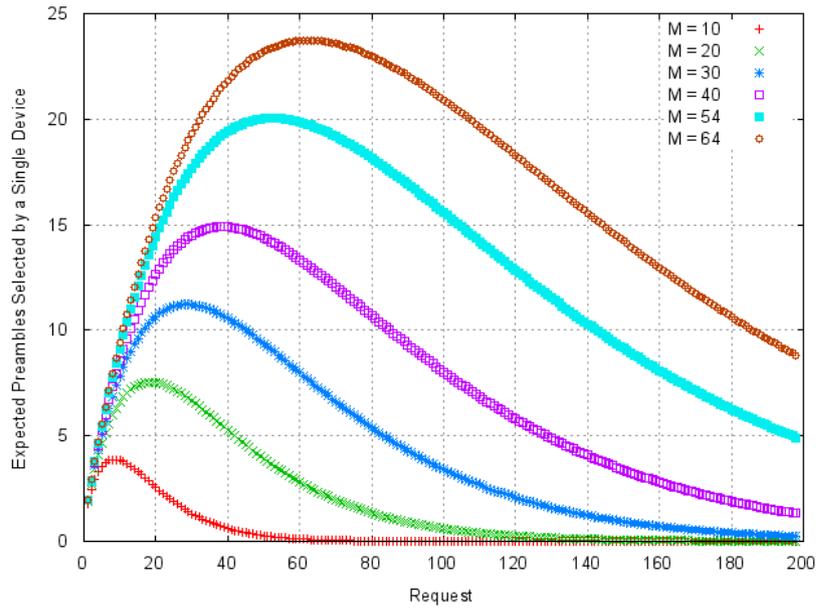


Figura 4.1: Preâmbulos escolhidos por um único dispositivo.

4.1.1 Alocação de Código de Preâmbulos do RACH

A divisão dos códigos de preâmbulos entre as diferentes classes pode ser caracterizada como um problema de fâlcia, onde a quantidade de códigos de preâmbulos disponíveis não é suficiente para atender a quantidade solicitada pelas classes. O cálculo da quantidade de preâmbulos demandado por cada classe é baseado no número de preâmbulos recebidos pela estação base. Após o procedimento de RACH, a estação base calcula quantos preâmbulos foram recebidos com sucesso, quantas colisões ocorreram e quantos preâmbulos não foram utilizados, representados, respectivamente, por $PreL_i$, $PreC_i$ e $PreE_i$. O vetor de demanda c_i é dado pela Equação 4.10.

$$c = \frac{1}{j} \sum_{l=1}^j (PreL_{i,k-l} + PreC_{i,k-l}) \times \delta \quad (4.10)$$

A Equação 4.10 é a média móvel simples dos preâmbulos recebidos com sucesso e dos preâmbulos que colidiram, onde k identifica qual é o número do RACH corrente.

Atualmente o procedimento de RACH não é capaz de identificar qual a classe de um dispositivo através da mensagem de solicitação de acesso ao canal, ou seja, a primeira mensagem do procedimento de RACH. No procedimento de RACH baseado em contenção, a eNodeB só é capaz de identificar a classe do dispositivo somente após a terceira mensagem do procedimento de RACH. Para viabilizar a contagem dos preâmbulos realizada pelos mecanismos propostos durante a primeira mensagem do RACH, neste trabalho atribuiu-se a cada dispositivo um código fixo e único. A estação base é capaz de identificar a qual classe o dispositivo pertence através do código do dispositivo a

Durante o momento inicial do sistema, ou seja, momento em que não há histórico

das solicitações de acesso, a estação base adota a estratégia de dividir os preâmbulos igualmente dentre as três classes de dispositivos. Para priorizar as diferentes classes de dispositivos consideradas neste trabalho (H2H, M2M_p, M2M_{wp}), os seguintes pesos (α , β , γ) são atribuídos a cada uma das classes durante o cálculo do vetor de demanda realizado pelas Equações 4.11, 4.12 e 4.13.

$$c_{H2H} = \alpha \left(\frac{1}{j} \sum_{l=1}^j (PreL_{H2H,k-l} + PreC_{H2H,k-l}) \times \delta \right) \quad (4.11)$$

$$c_{M2M_p} = \beta \left(\frac{1}{j} \sum_{l=1}^j (PreL_{M2M_p,k-l} + PreC_{M2M_p,k-l}) \times \delta \right) \quad (4.12)$$

$$c_{M2M_{wp}} = \gamma \left(\frac{1}{j} \sum_{l=1}^j (PreL_{M2M_{wp},k-l} + PreC_{M2M_{wp},k-l}) \times \delta \right) \quad (4.13)$$

Os pesos α , β , γ são utilizados pelos mecanismos Gtoken e Ptoken e foram obtidos através de simulações. Os pesos são atribuídos as classes de dispositivos visando garantir que a prioridade dos dispositivos M2M_p seja inferior à dos dispositivos H2H, mas superior à dos dispositivos M2M_{wp}. Os valores atribuídos ao α , β , γ para os mecanismos Gtoken e Ptoken são: $\alpha = 3$, $\beta = 2$, $\gamma = 0.265$ e $\alpha = 3$, $\beta = 2$, $\gamma = 1$, respectivamente.

4.1.2 Alocação de Códigos de Preâmbulos Usando Jogos Cooperativos (Gtoken)

O número máximo de preâmbulos disponíveis no LTE para o procedimento de RACH é limitado em 64 códigos e em média 10 desses códigos são alocados para o procedimento de RACH livre de contenção. A quantidade de códigos restantes deve ser alocados para as classes de dispositivos é $E = 54$ (64-10). Considerando que há três classes de dispositivos, $N = \{H2H, M2M_p, M2M_{wp}\}$ e i um índice de N , o vetor de demanda de cada classe é dado pelas Equações 4.11, 4.12, 4.13. Deste modo, tem-se que o vetor de demanda para os dispositivos das classes H2H, M2M_p e M2M_{wp} é dado por:

$$c_{H2H} = 3 \times \left(\frac{1}{4} \sum_{l=1}^4 (PreL_{H2H,k-l} + PreC_{H2H,k-l}) \times 0.37 \right) \quad (4.14)$$

$$c_{M2M_p} = 2 \times \left(\frac{1}{4} \sum_{l=1}^4 (PreL_{M2M_p,k-l} + PreC_{M2M_p,k-l}) \times 0.37 \right) \quad (4.15)$$

$$c_{M2M_{wp}} = 0.265 \times \left(\frac{1}{4} \sum_{l=1}^4 (PreL_{M2M_{wp},k-l} + PreC_{M2M_{wp},k-l}) \times 0.37 \right) \quad (4.16)$$

Onde nas Equações 4.11, 4.12, 4.13 o cálculo da média é realizada considerando

os últimos 4 preâmbulos recebidos ($j = 4$). O mecanismo é executado na eNodeB que age como árbitro do jogo, sendo responsável pela divisão dos recursos e divulgação, através da mensagem de SIB2, de como os códigos de preâmbulos estão alocados para os dispositivos. Assim, o problema de falência pode ser modelado utilizando o *framework* de jogos cooperativos $G(N, v_c)$, onde $v_c(S) = \max\{E - \sum_{i \in N \setminus S} c_i, 0\}$, e $v_c(N) = E$. Desenvolvendo $v_c(S)$:

$$\begin{aligned} v_c(H2H) &= \max\{54 - (c_{M2M_p} + c_{M2M_{wp}}), 0\} \\ v_c(M2M_p) &= \max\{54 - (c_{H2H} + c_{M2M_{wp}}), 0\} \\ v_c(M2M_{wp}) &= \max\{54 - (c_{H2H} + c_{M2M_p}), 0\} \\ v_c(H2H, M2M_p) &= \max\{54 - (c_{M2M_{wp}}), 0\} \\ v_c(M2M_p, M2M_{wp}) &= \max\{54 - (c_{H2H}), 0\} \\ v_c(H2H, M2M_p, M2M_{wp}) &= 54 \end{aligned}$$

Para finalizar, a quantidade de preâmbulos que deve ser atribuído a cada uma das classe através do ϕ_i , (cf., Equação 2.6), ou seja, ϕ_{H2H} , ϕ_{M2M_p} , $\phi_{M2M_{wp}}$.

4.1.3 Alocação de Códigos de Preâmbulos Usando a Regra Proporcional (Ptoken)

O problema de falência pode ser representado usando a regra proporcional através da tupla $P(E, c)$, como apresentado na Seção 2.9.2. Segundo a modelagem adotada neste trabalho, os códigos de preâmbulos de RACH são os recursos em escassez. A quantidade de recursos demandados pelos dispositivos das classes H2H, M2M_p e M2M_{wp} é dado respectivamente pelas Equações 4.11, 4.12, 4.13. Assim como o mecanismo Gtoken o Ptoken também é executado na eNodeB. A eNodeB age como árbitro do jogo, sendo responsável pela divisão dos recursos e pela divulgação de como os códigos de preâmbulos estão alocados para os dispositivos. Os dispositivos têm ciência de como os recursos estão alocados a partir das mensagens de SIB2 enviada pela eNodeB. Deste modo, tem-se: $E = 54$, $\lambda = E/C$, onde $C = \sum_{i \in N} c_i$ e $N = \{H2H, M2M_p, M2M_{wp}\}$. Deste modo, o C é dado por:

$$\begin{aligned} C = c_{H2H} &= 3 \times \left(\frac{1}{4} \sum_{l=1}^4 (PreL_{H2H, k-l} + PreC_{H2H, k-l}) \times 0.37 \right) + \\ c_{M2M_p} &= 2 \times \left(\frac{1}{4} \sum_{l=1}^4 (PreL_{M2M_p, k-l} + PreC_{M2M_p, i, k-l}) \times 0.37 \right) + \\ c_{M2M_{wp}} &= 1 \times \left(\frac{1}{4} \sum_{l=1}^4 (PreL_{M2M_{wp}, k-l} + PreC_{M2M_{wp}, k-l}) \times 0.37 \right) \end{aligned}$$

O funcionamento dos mecanismos Gtoken e Ptoken está representado na Figura 4.2.

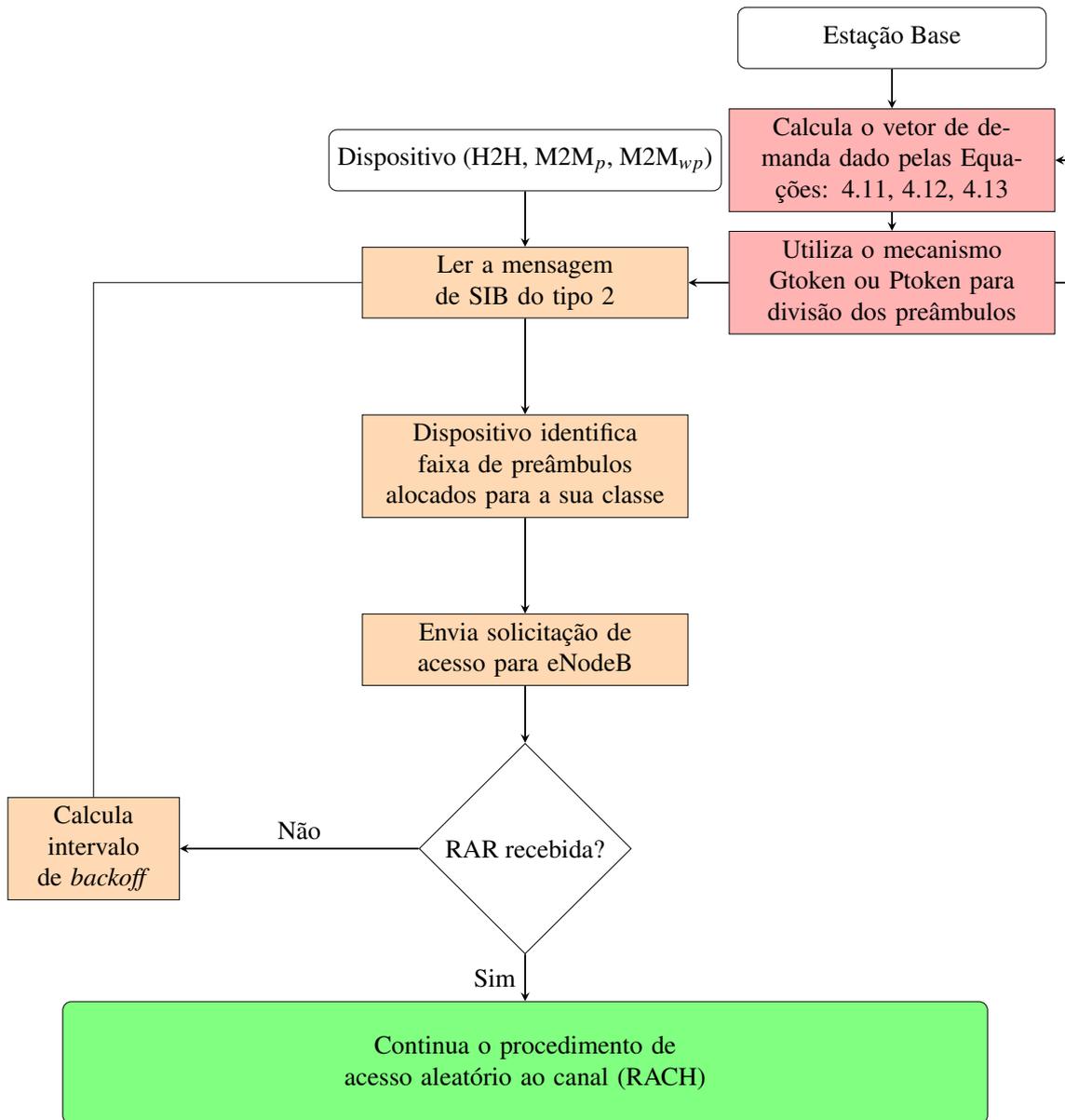


Figura 4.2: Etapas comuns aos mecanismos Gtoken e Ptoken.

4.1.4 Análise dos Resultados

Os resultados apresentados nesta seção comparam o desempenho dos mecanismos Gtoken, Ptoken e o mecanismo Fixo. O mecanismo Fixo representa indiretamente o comportamento das estratégias presentes em outros trabalhos (e.g., (PANG et al., 2014), (LEE; KIM; YI, 2011)). No mecanismo Fixo, como o nome já evidencia, os códigos de preâmbulos são igualmente distribuídos entre as três classes de dispositivos. Os Indicadores-Chaves de Performance (*Key Performance Indicators* - KPI's) são: a latência no acesso, o total de acessos realizados com sucesso e o total de solicitações de acesso transmitidas.

4.1.4.1 Impacto Sobre os Dispositivos H2H

O impacto sobre o número total de acessos de dispositivos H2H é baixo. Independentemente do número de dispositivos M2M, quase todos os dispositivos da classe H2H conseguem acessar a rede, como ilustrado na Figura 4.3. Os recursos solicitados pela classe H2H são priorizados em relação as outras classes nos mecanismos Gtoken e Ptoken. Diferentemente do mecanismo Fixo, nos mecanismos Gtoken e Ptoken a alocação dos recursos do RACH é realizada dinamicamente. O mecanismo Fixo também apresenta poucas variações, resultado já esperado, pois o número de preâmbulos fixado por este mecanismo é suficiente para atender a demanda de acesso dos dispositivos H2H. A alta prioridade estabelecida para os dispositivos H2H repercute positivamente na probabilidade de acesso dos três mecanismos, que se mantém aproximadamente entre 1.0% e 0.995%, como ilustrado na Figura 4.3. Em suma, os mecanismos apresentam bons resultados em relação ao controle do impacto sobre os dispositivos H2H.

A latência no acesso aumenta quando há mais dispositivos tentando acessar a rede, como já esperado e ilustrado na Figura 4.4. Dentre os três mecanismos simulados, os mecanismos Gtoken e o Fixo apresentam resultados muito próximos e melhores do que o Ptoken. Contudo, a diferença dos mecanismos Gtoken e Fixo para o Ptoken é de aproximadamente 0.02 segundos, que de modo geral tem baixas implicações sobre as aplicações H2H.

A Figura 4.5 apresenta o número total de preâmbulos transmitidos pelos dispositivos H2H de cada mecanismo. Para o cenário com 3050 dispositivos, o mecanismo Ptoken apresenta o pior resultado, onde são transmitidos aproximadamente 450 preâmbulos. Resultados melhores são apresentados pelos mecanismos Gtoken e o Fixo, onde são transmitidos aproximadamente 300 e 290 preâmbulos, respectivamente. Com exceção dos primeiros cenários em que há menos de 500 dispositivos sendo simulados, as vantagens apresentadas pelos mecanismos Gtoken e Fixo se mantêm sobre o mecanismo Ptoken.

4.1.4.2 Impacto Sobre os Dispositivos M2M

Para analisar o desempenho dos mecanismos simulados optou-se por verificar como o aumento do número de dispositivos $M2M_{wp}$ afeta no acesso dos dispositivos $M2M_p$. Os três mecanismos apresentam o número máximo de acessos limitados em 300. O mecanismo Ptoken

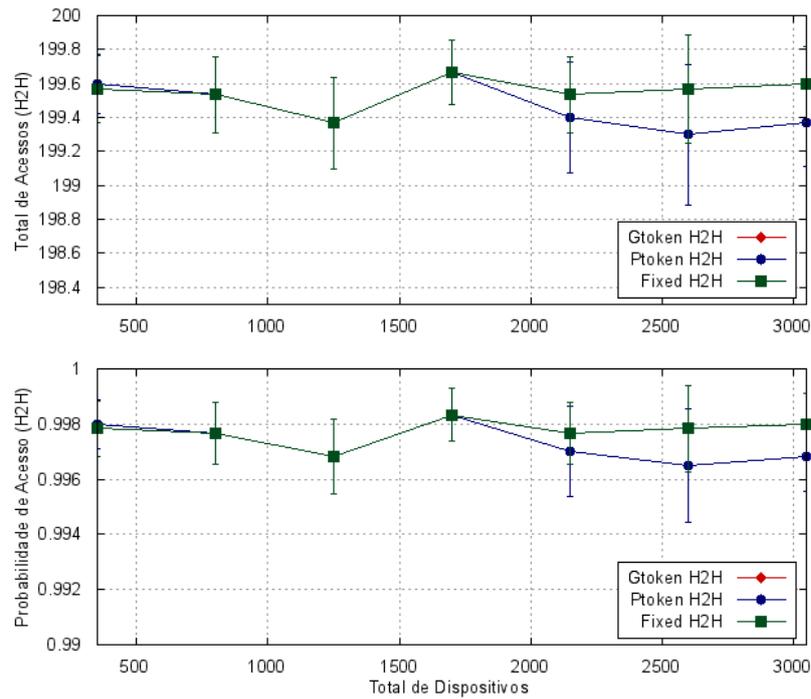


Figura 4.3: Probabilidade de acesso dos dispositivos H2H e total de acessos efetuados com sucesso dos dispositivos H2H.

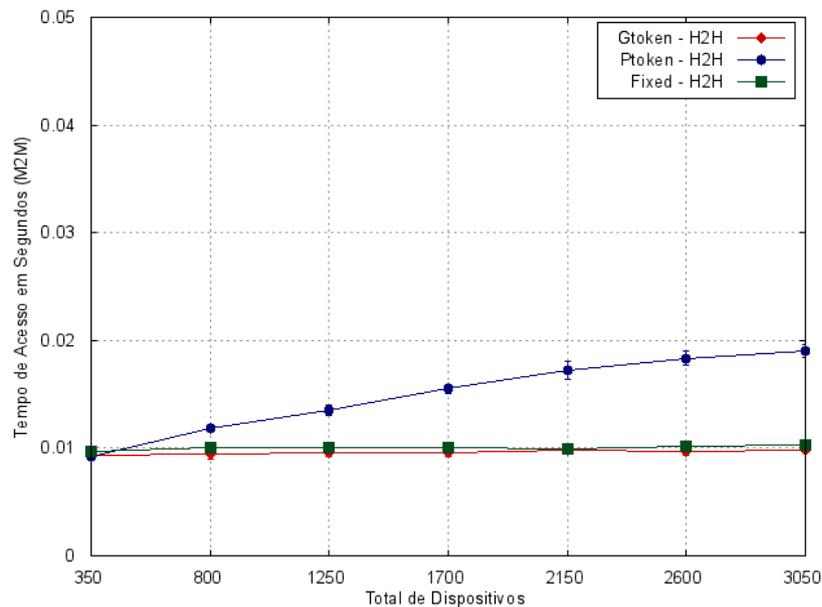


Figura 4.4: Tempo médio de acesso dos dispositivos H2H.

apresenta resultados melhores do que os mecanismos Gtoken e Fixo em termos de priorização das classes. Com o aumento do número de dispositivos $M2M_{wp}$ e $M2M_p$, o número de acessos efetuados com sucesso pelos dispositivos $M2M_{wp}$ começa a diminuir, possibilitando assim que mais dispositivos $M2M_p$ acessem a rede, como ilustrado na Figura 4.6. Isto ocorre porque os códigos de pré-âmbulo antes alocados para a classe $M2M_{wp}$ são alocados para os dispositivos da classe $M2M_p$. Assim, percebe-se que ocorre a realocação dos recursos do RACH nos mecanis-

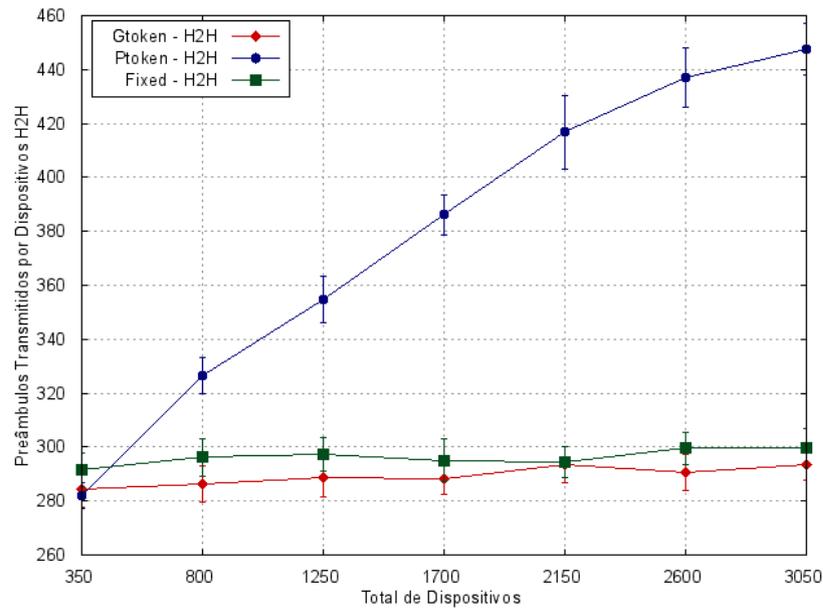


Figura 4.5: Solicitações de acesso enviadas por dispositivos H2H.

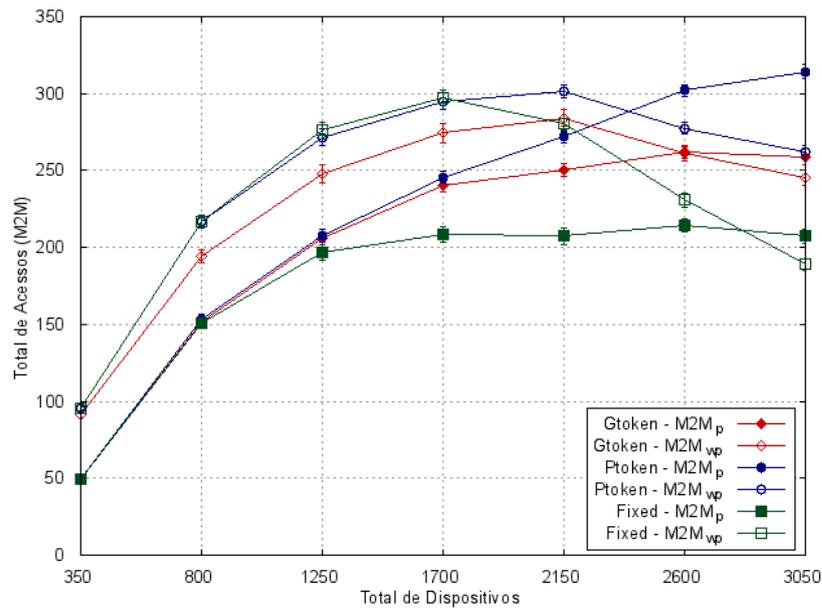


Figura 4.6: Acessos realizados com sucesso pelos dispositivos M2M_p e M2M_{vp}.

mos Gtoken e Ptoken. Em contrapartida, o decremento no número de acessos dos dispositivos M2M_{vp} experimentado pelo mecanismo Fixo não é acompanhado pelo incremento no número de acesso dos dispositivos M2M_p.

A Figura 4.7 ilustra o número total de acessos considerando as três classes de dispositivos simuladas. Devido ao número fixo de dispositivos H2H, o desempenho apresentado pelos mecanismos Gtoken e Ptoken deve-se aos dispositivos M2M. Com base nos resultados ilustrados na Figura 4.7 é possível observar que o número máximo de acessos alcançado por cada proposta é de aproximadamente 700, 740 e 780 para os mecanismos Fixo, Gtoken e Ptoken respectivamente. Com o aumento do número de dispositivos solicitando acesso à rede, ou

seja, aumento do nível de congestionamento tem-se o decréscimo do número de acessos efetuados com sucesso.

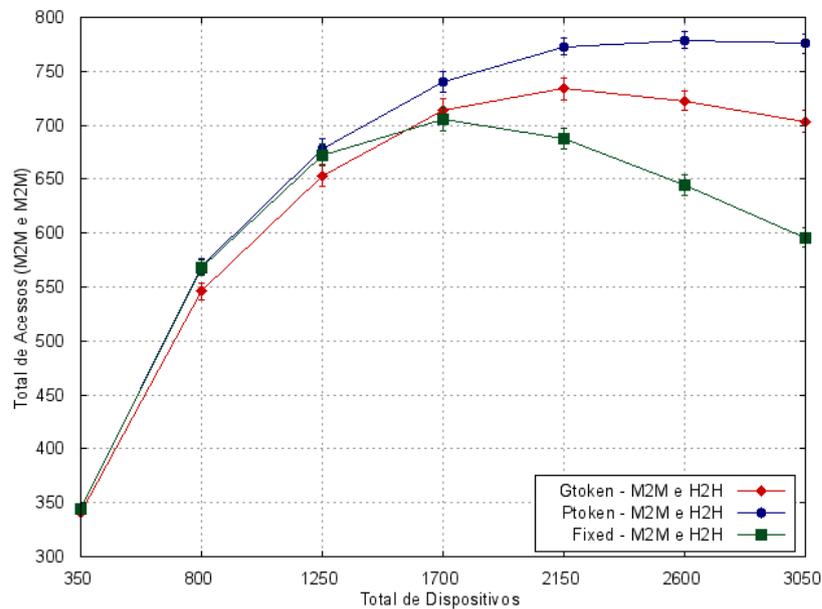


Figura 4.7: Acessos realizados com sucesso pelos dispositivos H2H, $M2M_p$ e $M2M_{wp}$.

O tempo médio de acesso para os dispositivos das classes $M2M_{wp}$ e $M2M_p$ são respectivamente 0.09 e 0.055 segundos, como ilustrado na Figura 4.8. Com base no número e tempo de acesso é possível verificar que os mecanismos Gtoken e Ptoken apresentam resultados melhores do que o mecanismo Fixo, garantindo mais acessos com baixo impacto no tempo de acesso.

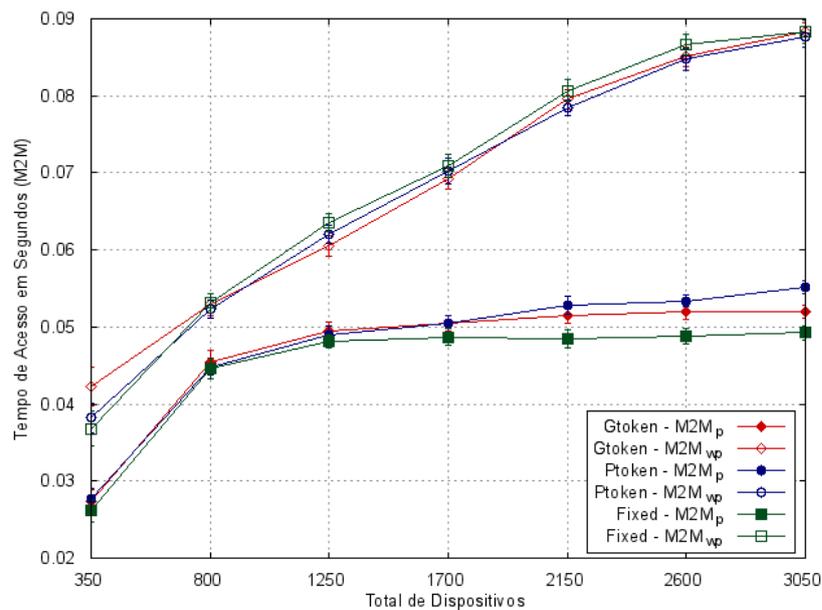


Figura 4.8: Tempo médio de acesso dos dispositivos $M2M_p$ e $M2M_{wp}$.

Os três mecanismos simulados respeitam a prioridade dos dispositivos M2M de alta prioridade sobre os de baixa prioridade, como ilustrado na Figura 4.9. Todas as classes garantem

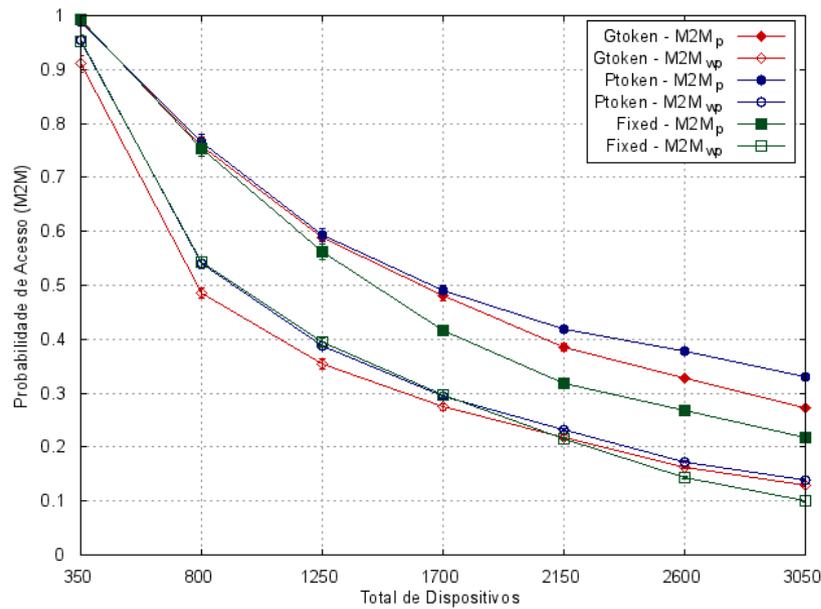


Figura 4.9: Probabilidade de acesso dos dispositivos M2M.

a probabilidade de acesso dos dispositivos M2M prioritário maior do que a dos dispositivos M2M de baixa prioridade. Como explicado na Seção 4.1.1, os mecanismos Gtoken e Ptoken ponderam com valores diferentes as classes de dispositivos. Deste modo, a diferença entre as classes M2M prioritária e M2M não prioritária é maior nos mecanismos Gtoken e Ptoken, como representado na Figura 4.9.

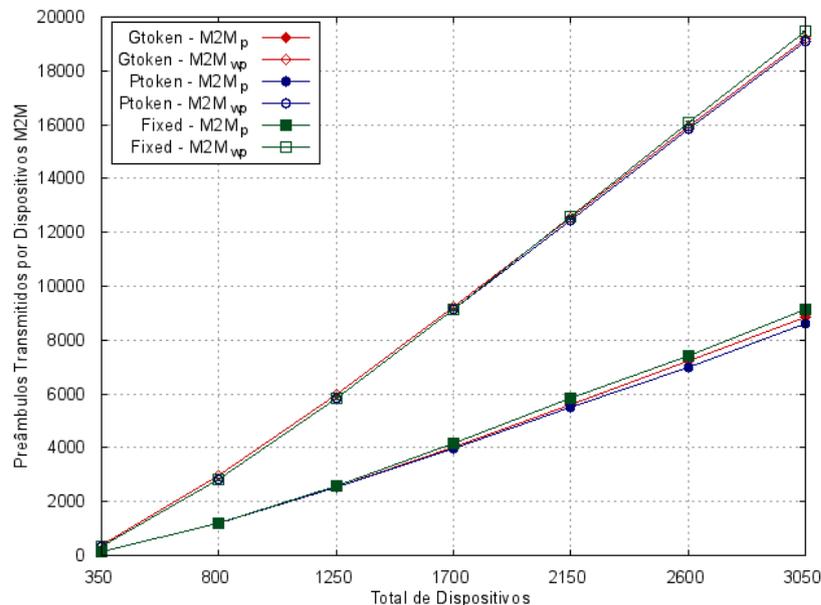


Figura 4.10: Solicitações de acesso enviadas por dispositivos M2M_p e M2M_{wp}.

Os três mecanismos apresentam poucas diferenças em relação à quantidade de preâmbulos transmitidos, como ilustrado na Figura 4.10. Contudo, o número de preâmbulos recebidos com sucesso, ou seja, sem colisão, é maior nos mecanismos Gtoken e Ptoken como ilustrado na Figuras 4.6 e 4.7. Assim, pode-se inferir que os mecanismos Gtoken e Ptoken

consomem menos energia do que o mecanismo Fixo.

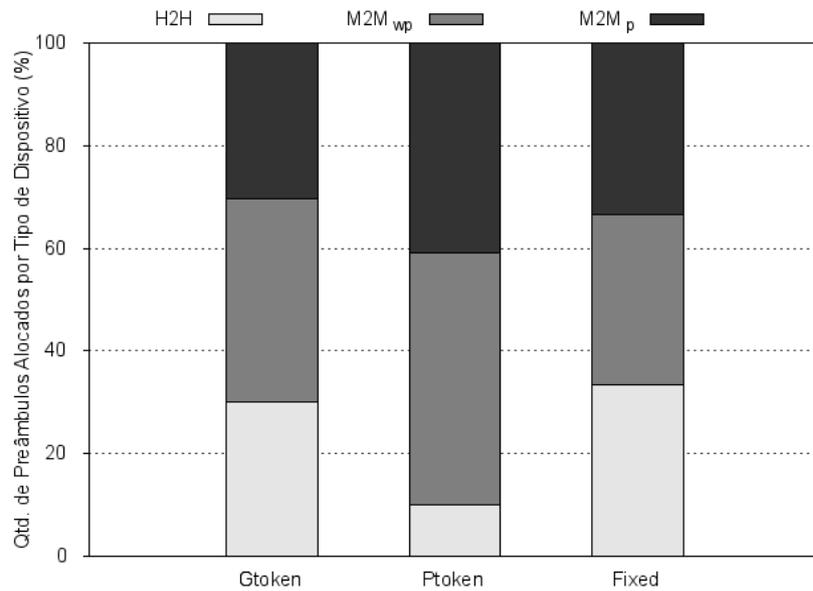


Figura 4.11: Alocação de recursos de cada mecanismo.

A Figura 4.11 complementa a informação apresentada na Figura 4.12, que apresentam a quantidade média de preâmbulos alocados para cada classe de dispositivos pelos mecanismos Gtoken, Ptoken e Fixo para o cenário com 3050 dispositivos. Como esperado, na abordagem fixa a quantidade de códigos alocados para as três classes permanece constante. Assim, enquanto há códigos reservados para os dispositivos H2H que não estão sendo utilizados, os dispositivos M2M_{wp} estão carecendo desses recursos.

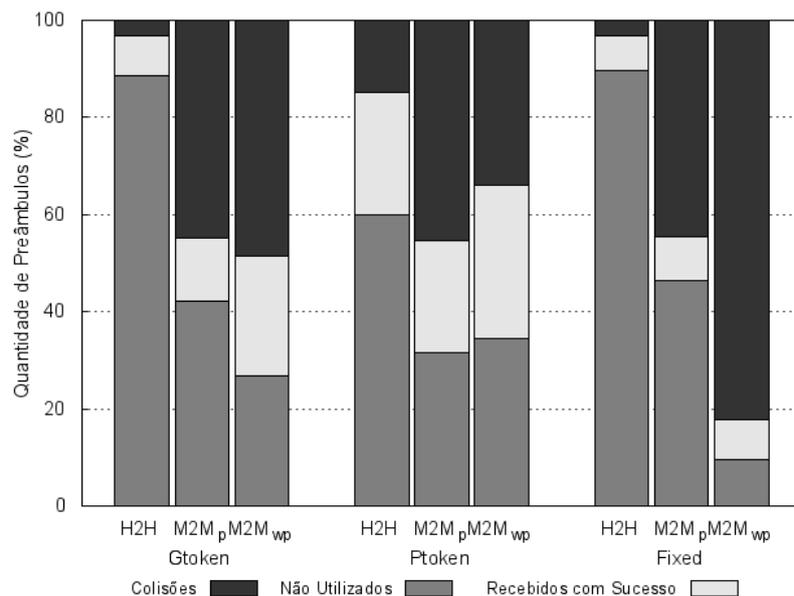


Figura 4.12: Nível de utilização de recursos alocados de cada mecanismo.

O mecanismo Gtoken, quando comparado com o Fixo, reserva mais recursos para os dispositivos M2M_{wp}. Através da Figura 4.11 é possível observar que o mecanismo Gtoken

divide os recursos alocados para o $M2M_{wp}$ de modo mais justo do que o mecanismo Fixo e sem comprometer os dispositivos H2H. Contudo, o número de preâmbulos alocados para H2H que não são utilizados ainda é bastante elevado. O mecanismo Ptoken aloca a maior quantidade de recursos para os dispositivos da classe M2M. Este mecanismo apresenta o gráfico de barra mais balanceado, mais precisamente em relação aos códigos alocados para os dispositivos H2H, como ilustrado na Figura 4.12. A redução dos recursos para o H2H contribui para o aumento do número de preâmbulos transmitidos pelos dispositivos (cf. Figuras 4.8 e 4.5).

4.2 Mecanismos de Alocação Dinâmica de Tempo de Backoff

A estratégia de criação de classes de *backoff* para os diferentes tipos de dispositivos apresenta-se como um problema de falência. A demanda por tempo das classes é proporcional ao número de dispositivos em cada classe, ou seja, as que possuem mais dispositivos demandam por intervalos de tempo maiores. Porém, devido ao número limitado de recursos de tempo, o somatório do intervalo demandado pelas classes ultrapassa o total de tempo disponível. Assim, tem-se um cenário de falência.

A priorização entre as diferentes classes de dispositivos também é um problema abordado pelos mecanismos Gcronos e Pcronos. Diferente dos mecanismos Gtoken e Ptoken, onde a priorização é alcançada através da atribuição de pesos ao vetor de demanda, nos mecanismos Gcronos e Pcronos a prioridade está associada a ordem em que os intervalos estão organizados, como ilustrado na Figura 4.13. Como apresentado na Figura 4.13, a prioridade atribuída aos dispositivos $M2M_p$ é menor do que a dos dispositivos H2H e maior do que a dos dispositivos $M2M_{wp}$, pois o tempo de espera dos dispositivos H2H está dentro do intervalo $(0, T_1)$ ms. Enquanto que para os dispositivos das classes $M2M_p$ e $M2M_{wp}$ os intervalos são respectivamente (T_1, T_2) e (T_2, T_3) .

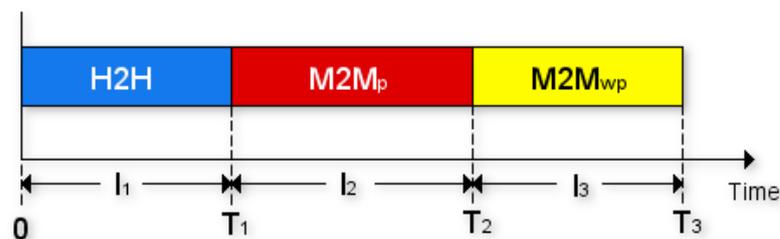


Figura 4.13: I_1 , I_2 e I_3 representam os intervalos de *backoff* atribuídos respectivamente as classes H2H, $M2M_p$ e $M2M_{wp}$. T é o tempo máximo de *backoff* que pode ser atribuído a uma classe.

Seja o tempo máximo de espera atribuível a uma determinada classe dado por $E = 200$ ms e $N = \{H2H, M2M_p, M2M_{wp}\}$ as classes de dispositivo indexadas por i . O vetor de demanda $c \in \mathbb{R}_+^n$, que identifica o intervalo de tempo solicitado por cada classe de dispositivos (i), é dado por:

$$c_i = q_i \times \delta \quad (4.17)$$

Onde δ o intervalo mínimo demandado por cada um dos q_i dispositivos da classe i . Para identificar o número de dispositivos que estão sendo simulados e consequentemente possibilitar que a estação base calcule o vetor de demanda das classes de dispositivos, os mecanismos Gcronos e Pcronos recebem como entrada a quantos dispositivos de cada classe estão sendo utilizados em cada experimento de simulação. De posse desta informação, a estação base, que age como árbitro, é capaz de calcular o vetor de demanda e determinar o tamanho dos intervalos para cada uma das classes. Deste modo, em um cenário de simulação hipotético que contenham 1000 dispositivos, sendo 300 da classe H2H, 200 da classe M2M_p e 500 da classe M2M_{wp}, $\delta = 1ms$, o vetor de reivindicações será igual a $c = (300, 200, 500)$.

4.2.1 Alocação de Intervalos de Backoff Usando Jogos Cooperativos (Gcronos)

O problema de definição do tamanho do intervalo das classes pode ser empregado utilizando uma abordagem de jogos $G(N, v_c(S))$, onde a quantidade de recursos disponível ($E = 200ms$) e o vetor com o intervalo de tempo demandado por cada dispositivo é dado por: $c_i = q_i \times \delta$, onde $\delta = 1 ms$ e q_i a quantidade de dispositivos que estão sendo simulado em cada uma das classes. Assim, tem-se que a quantidade de recursos tempo demandado por cada classe é dado por:

$$c_{H2H} = q_{H2H} \times 1.0 \quad (4.18)$$

$$c_{M2M_p} = q_{M2M_p} \times 1.0 \quad (4.19)$$

$$c_{M2M_{wp}} = q_{M2M_{wp}} \times 1.0 \quad (4.20)$$

Dado que a prioridade atribuída as classes segue a ordem Prioridade (H2H) > Prioridade (M2M_p) > Prioridade (M2M_{wp}), tem-se que o intervalo atribuído aos dispositivos H2H ocupará a primeira porção do tempo E , seguido pelos intervalos determinados para os dispositivos das classes M2M_p e M2M_{wp}, respectivamente. Deste modo, após uma colisão, os dispositivos H2H devem esperar um tempo menor do que os dispositivos das classes M2M_p e M2M_{wp} para poderem retransmitir uma nova solicitação. Assim, a prioridade de uma classe deve ser analisada a partir de dois elementos, sendo eles: o tamanho do intervalo e como estes intervalos estão ordenados/mapeados no intervalo E . Dada a função características $v_c(S) = \max\{E - \sum_{i \in N \setminus S} c_i, 0\}$, e $v_c(N) = E$. Tem-se que $v_c(S)$ pode ser dado por:

$$\begin{aligned}
v_c(H2H) &= \max\{200 - (c_{M2M_p} + c_{M2M_{wp}}), 0\} \\
v_c(M2M_p) &= \max\{200 - (c_{H2H} + c_{M2M_{wp}}), 0\} \\
v_c(M2M_{wp}) &= \max\{200 - (c_{H2H} + c_{M2M_p}), 0\} \\
v_c(H2H, M2M_p) &= \max\{200 - (c_{M2M_{wp}}), 0\} \\
v_c(M2M_p, M2M_{wp}) &= \max\{200 - (c_{H2H}), 0\} \\
v_c(H2H, M2M_p, M2M_{wp}) &= 200
\end{aligned}$$

O vetor de reivindicações é dado em milissegundos. Para finalizar, o tamanho do intervalo de cada uma das classe é calculado através do ϕ_i , (cf., Equação 2.6), ou seja, ϕ_{H2H} , ϕ_{M2M_p} , $\phi_{M2M_{wp}}$.

4.2.2 Alocação de Intervalos de Backoff Usando a Regra Proporcional (Pcronos)

Uma estratégia para modelar o problema da alocação de intervalos de *backoff* utilizando a regra proporcional é considerando os recursos de tempo disponíveis dado por $E = 200$ ms. A quantidade de tempo total reivindicada pelas classes de dispositivo é dado por $C = \sum_{i \in N} c_i$. Assim, C é dado por:

$$C = (q_{H2H} + q_{M2M_p} + q_{M2M_{wp}}) \times 1.0$$

Deste modo, tem-se que $\lambda = \frac{E}{C}$. A prioridade entre as classes utilizando a regra proporcional é a mesma apresentada no mecanismo Gcronos, ou seja, Prioridade(H2H) > Prioridade(M2M_p) > Prioridade (M2M_{wp}). O funcionamento dos mecanismos Gcronos e Pcronos está representado na Figura 4.14. A Tabela 4.1 apresenta a relação estabelecida neste trabalho entre os mecanismos Gtoken, Ptoken, Gcronos e Pcronos com o problema de falência.

Problema de Falência	Gtoken	Ptoken	Gcronos	Pcronos
Recursos em escassez	Código de Preâmbulos		Tempo	
Vetor de demanda	Equações: 4.11, 4.12, 4.13		Equações: 4.17	
Distribuição dos recursos	Valor de Shapley	Regra Proporcional	Valor de Shapley	Regra Proporcional
Jogadores	Classe de dispositivos: H2H, M2M _p e M2M _{wp}			

Tabela 4.1: Tabela comparativa entre os mecanismos Gtoken, Ptoken, Gcronos e Pcronos.

Além da relação com o problema de falência, a Tabela 4.1 também apresenta os pontos que diferem os mecanismos Gtoken, Ptoken, Gcronos e Pcronos.

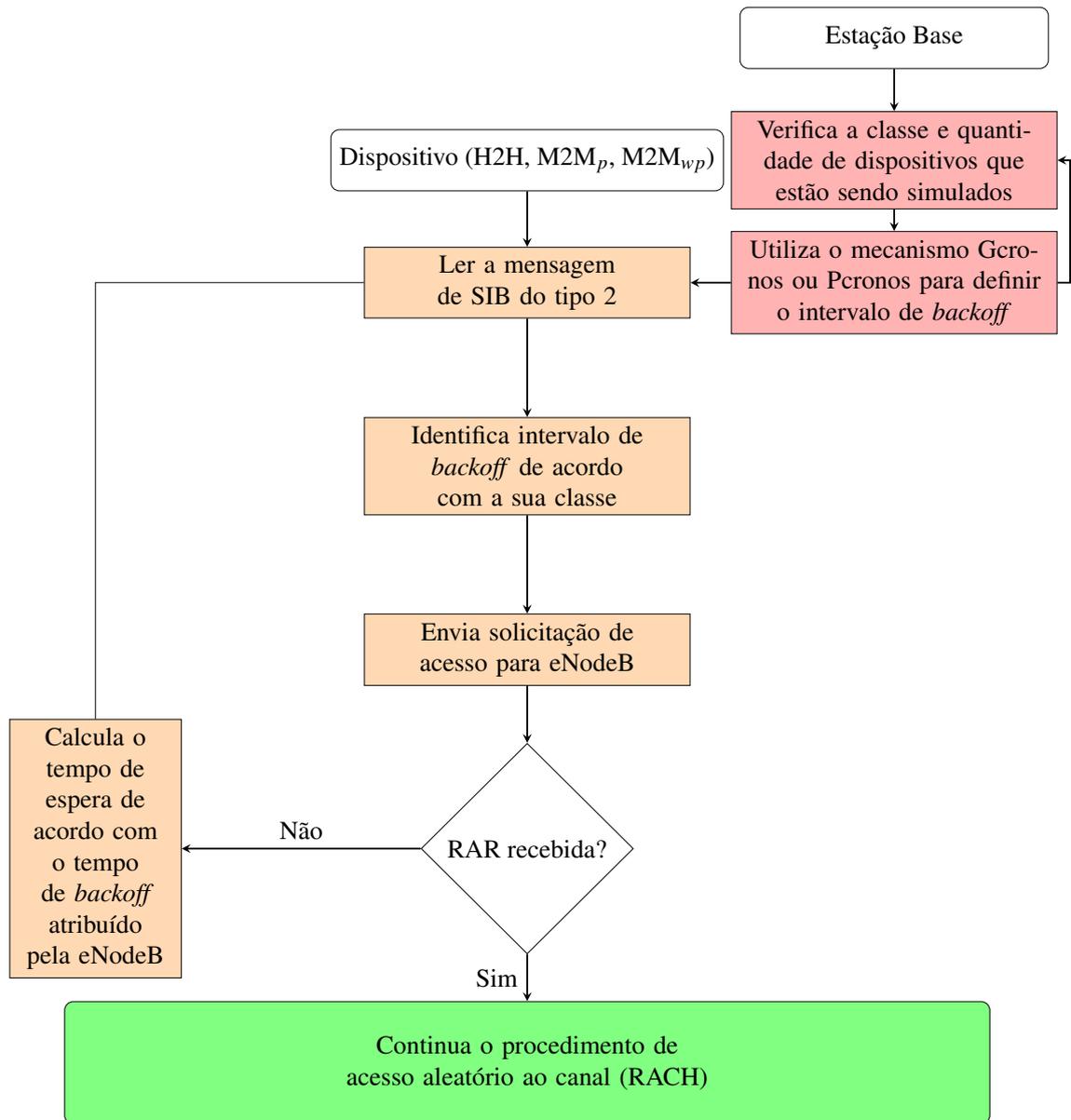


Figura 4.14: Etapas Comuns aos Mecanismos Gbrnos e Pbrnos.

4.2.3 Análise dos Resultados

Os resultados apresentados nesta seção comparam o desempenho dos mecanismos Gcronos, Pcronos e o mecanismo Fixo, que indiretamente representa a estratégia utilizada em outros trabalhos (e.g., (JIANG et al., 2014),(ZTE, 2011)). No mecanismo Fixo, os intervalos de *backoff* atribuídos as três classes são de mesmo tamanho. Os indicadores-chave de performance (*Key Performance Indicators - KPI's*) são: a latência no acesso, o total de acesso realizados com sucesso e a quantidade de preâmbulos transmitidos. Os principais aspectos analisados segundo os indicadores-chave supracitados são: o impacto sobre dispositivos H2H e o impacto sobre os dispositivos M2M e a eficiência energética.

4.2.3.1 Impacto Sobre os Dispositivos H2H

A Figura 4.15 apresenta a relação entre o número de acessos realizados com sucesso e número total de dispositivos solicitando acesso à rede. Em termo de tempo de acesso, a melhor e pior performance são apresentadas, respectivamente, pelo mecanismo Fixo e pelo mecanismo Gcronos. O mecanismo Fixo atribui o mesmo intervalo de *backoff* as três classes de dispositivos. Assim, o impacto sobre os dispositivos H2H causado pelos dispositivos M2M é menor do que o apresentado pelos mecanismos Gcronos e Pcronos. Isto ocorre porque o número de códigos reservados para H2H pelo mecanismo Fixo é suficiente para atender a demanda da classe. Nos mecanismos Gcronos e Pcronos, o intervalo de *backoff* atribuído aos dispositivos H2H está mais susceptível as variações do número de dispositivos M2M. A variação, aumento ou redução do intervalo de *backoff*, tem as seguintes implicações: se um intervalo pequeno é atribuído aos dispositivos, menor o intervalo de espera entre as sucessivas tentativas de acesso e menos dispersas serão as solicitações. Caso contrário, as solicitações ficam muito dispersas, elevando assim o tempo de acesso.

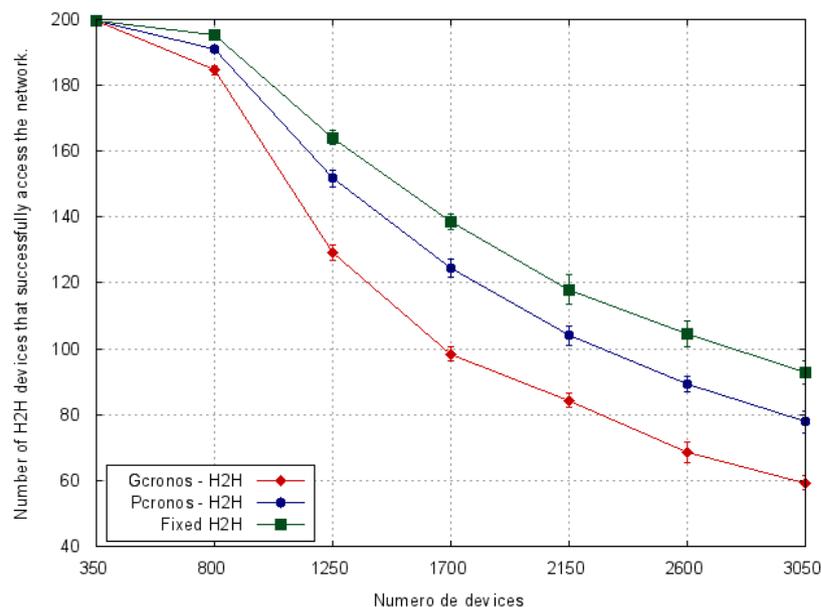


Figura 4.15: Acessos realizados com sucesso pelos dispositivos H2H.

O *tradeoff* entre tempo de acesso e total de acessos realizados com sucesso está bem representado pelas Figuras 4.16 e 4.15, respectivamente. Baseado nos resultados apresentados pelo mecanismo Fixo e Pcronos é possível observar que o Pcronos consegue reduzir o tempo de acesso, mas apresenta desempenho inferior ao mecanismo Fixo em relação ao número de acessos. Baseado no *tradeoff*, tempo de acesso e número total de acessos, espera-se que a baixa performance apresentada pelo mecanismo Gcronos em termos de tempo de acesso seja compensada apresentando melhor performance em termo de número total de acessos. Contudo, como apresentado na Figura 4.15, o mecanismo Gcronos não possui bons resultados em relação ao número de acessos realizados com sucesso. Uma possível explicação é o aumento do intervalo atribuído aos dispositivos da classe H2H que consequentemente dispersa o envio das solicitações de acesso, como apresentado na Figura 4.17.

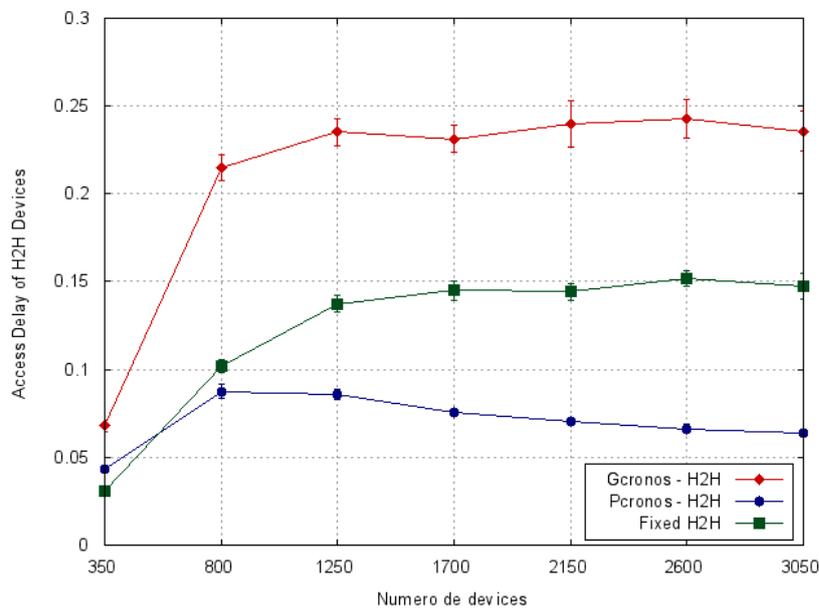


Figura 4.16: Tempo médio de acesso dos dispositivos H2H.

A Figura 4.17 apresenta o número de preâmbulos enviado por cada estratégia. Com o aumento do número de dispositivos solicitando acesso, mais colisões tendem a ocorrer. O número de preâmbulos pode ser utilizado como parâmetro para indicar o quão dispersas estão as solicitações no tempo. O baixo número de preâmbulos enviados pelo mecanismo Gcronos indica que as solicitações estão mais dispersas. Em contrapartida, o mecanismo Pcronos apresenta o menor intervalo de *backoff* e consequentemente menos disperso.

4.2.3.2 Impacto Sobre os Dispositivos M2M

As três abordagens apresentam bons resultados em relação ao número de dispositivos que conseguem acessar a rede com sucesso, como ilustrado na Figura 4.18. Apesar de haver poucas diferenças nos resultados, o mecanismo Pcronos apresenta o menor tempo de acesso para os dispositivos M2M prioritários, como ilustrado na Figura 4.19. Assim, baseado no número e tempo de acesso ilustrados respectivamente nas Figuras 4.18 e 4.19 é possível verificar que o mecanismo Pcronos apresenta a melhor performance. Isto é, além de sustentar o

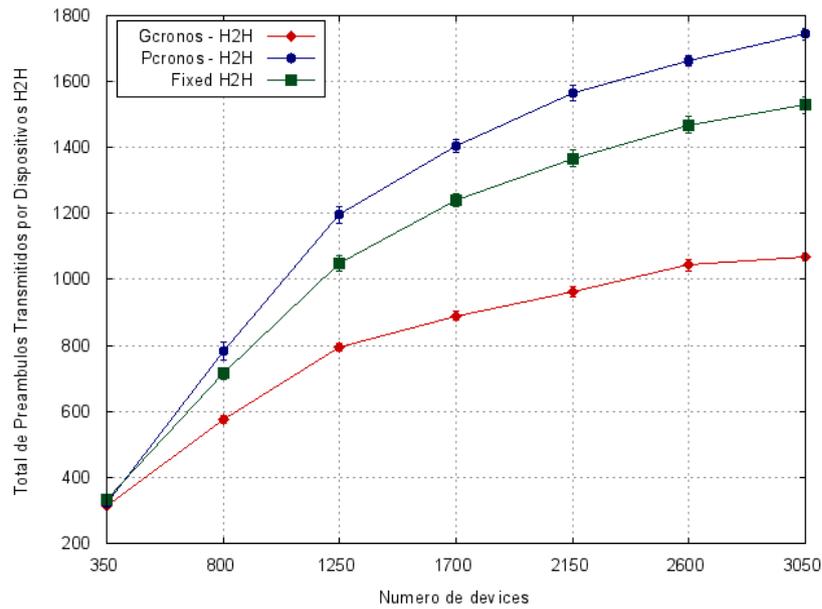


Figura 4.17: Solicitações de acesso enviadas por dispositivos H2H.

mesmo resultado dos outros mecanismos em relação ao total de acessos realizados com sucesso, o mecanismo Pcronos garante também o menor tempo de acesso.

A Figura 4.19 ilustra o tempo de acesso de cada abordagem. O melhor e pior desempenho são apresentados pelos mecanismos Gcronos e Pcronos, respectivamente. A relação entre o número de acessos efetuados com sucesso e o tempo de acesso é tal que quanto maior o número de dispositivos, maior o tempo de acesso.

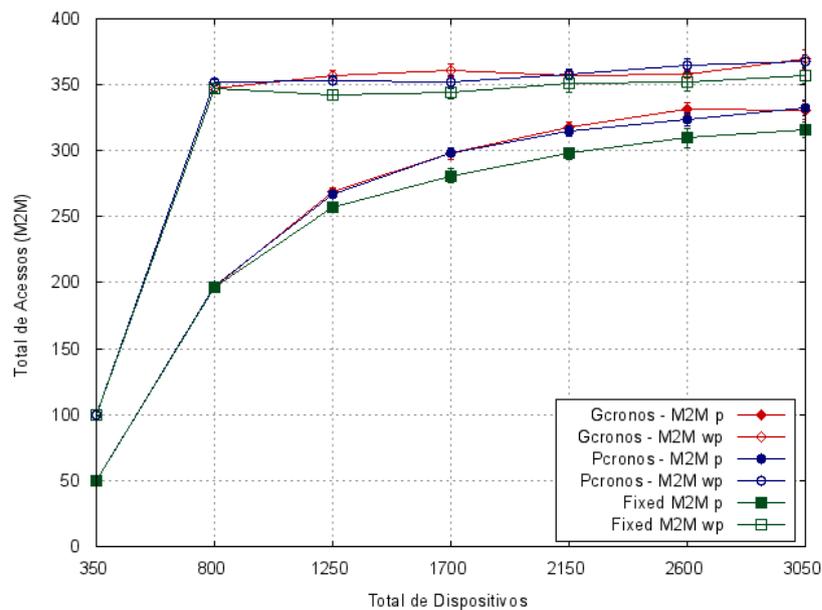


Figura 4.18: Acessos realizados com sucesso pelos dispositivos M2M_p e M2M_{wp}.

O número de solicitações de acesso enviadas pelos dispositivos M2M e H2H no mecanismo Pcronos é o maior dentre as estratégias simuladas, como ilustrado nas Figuras 4.20

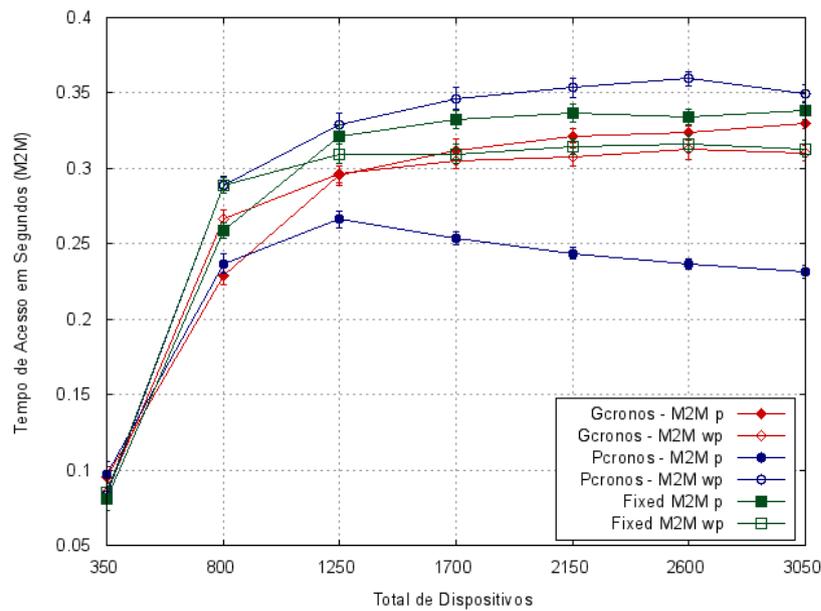


Figura 4.19: Tempo médio de acesso dos dispositivos $M2M_p$ e $M2M_{wp}$.

e 4.17. O aumento do número de preâmbulos alocados para a classe H2H resulta em mais colisões, aumentando assim o número de retransmissões e consequentemente o número de tentativas de acesso sem sucesso. Contudo, o aumento do número de preâmbulos reservados para os dispositivos M2M prioritários provê melhor desempenho em termos de tempo de acesso. Com base nesses resultados, podemos concluir que o mecanismo Pcronos aprimora o tempo de acesso para os dispositivos M2M prioritários sem impactar no acesso.

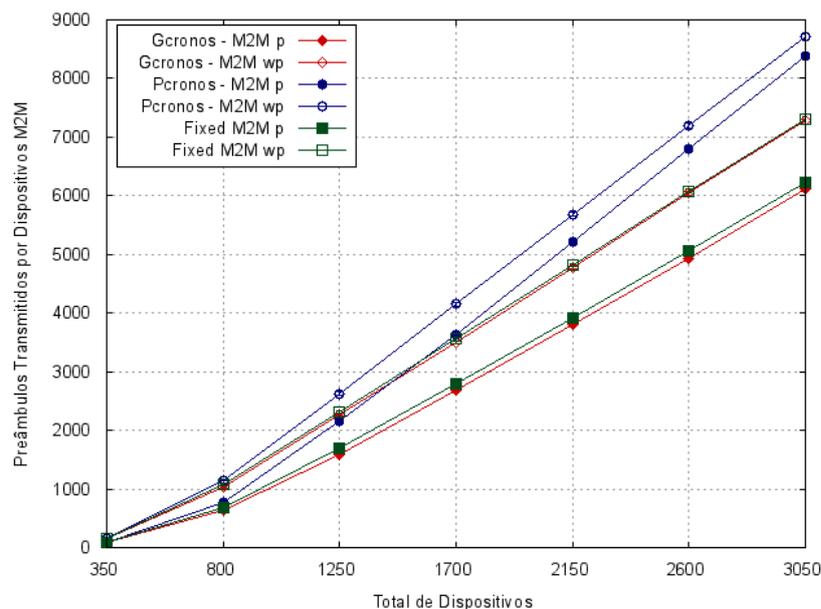


Figura 4.20: Solicitações de acesso enviadas por dispositivos $M2M_p$ e $M2M_{wp}$.

4.3 Recomendações

Os mecanismos propostos neste capítulo agem diretamente na estação base determinando como os recursos de tempo e código de preâmbulos devem ser alocados às diferentes classes de dispositivos. A partir dos resultados obtidos é possível verificar que as propostas que agem sobre a distribuição dos códigos de preâmbulos apresentam resultados melhores do que as propostas que agem sobre a reserva de tempo.

Os mecanismos Gtoken e Ptoken são indicados para cenários em que o controle do impacto sobre os dispositivos H2H é considerado um requisito de alta prioridade. Contudo, esses mecanismos também apresentam bons resultados em relação aos indicadores tempo e número de acessos realizados com sucesso para os dispositivos H2H e M2M.

A principal desvantagem dos Gtoken e Ptoken está relacionada a sua implementação em um cenário real, pois apesar de sugerida pela 3GPP, a identificação dos dispositivos e reserva de códigos de preâmbulos ainda não foram implementados no LTE.

A identificação dos dispositivos a partir do envio da primeira mensagem de RACH também apresenta-se como um empecilho para a implantação dos mecanismos Gcronos e Pcronos em um cenário real. Contudo, a estratégia de priorização dos dispositivos através da técnica de *backoff* utilizada por Gcronos e Pcronos é amplamente conhecida e utilizada. Em resumo, os mecanismos Gcronos e Pcronos devem ser aplicados quando a compatibilidade com a rede for um requisito prioritário.

5 CONCLUSÕES

A rede LTE e a comunicação M2M têm grande importância no contexto da Internet das coisas. Para suportar o número excessivo de dispositivos M2M estimado para os próximos anos é preciso implementar mecanismos que controlem a sobrecarga e o congestionamento ocasionado por esses dispositivos. Diante deste cenário, cinco mecanismos são apresentados neste trabalho, sendo um com o objetivo de garantir maior compatibilidade de implementação no LTE; dois com o objetivo de aprimorar a estratégia de alocação de recursos do acesso aleatório; e dois com o objetivo de aprimorar a técnica de *backoff*. A principal contribuição deste trabalho é expor como esses mecanismos podem contribuir para reduzir o problema do congestionamento durante o procedimento de RACH.

Visando manter a compatibilidade com a atual implementação do LTE, o PClass aprimora o algoritmo de acesso ao meio utilizado pelos dispositivos. O PClass utiliza como parâmetros o nível de congestionamento recebido da estação base, a classe do dispositivo e o número de tentativas de acesso já realizadas pelo dispositivo para coordenar o envio das solicitações de acesso. Os resultados obtidos mostram como a variação na probabilidade de acesso e tempo de *backoff* podem ser utilizados para priorizar as diferentes classes de dispositivos e mitigar o impacto ocasionado pelo excesso de solicitações realizadas pelos dispositivos M2M. O uso de uma estratégia condizente com a realidade da rede LTE, mesmo com o impacto esperado, apresenta-se como uma boa opção para este estudo. A implementação tem baixa complexidade, utilizando uma distribuição amplamente conhecida (Uniforme) em conjunto com a técnica de *backoff*. As principais diferenças deste mecanismo em relação aos demais apresentados neste trabalho é a priorização inter-classe, a baixa complexidade de implementação, a proximidade com a realidade das redes LTE, implementação essencialmente centrada na camada de acesso ao meio dos dispositivos.

Quatro mecanismos são propostos para tratar os problemas de falência dos recursos de tempo e código de preâmbulos. Os mecanismos Gtoken e Ptoken tratam o problema de falência que ocorre durante a alocação de recursos do RACH, enquanto os mecanismos Gcronos e Pcronos do problema de falência que ocorre durante a alocação de tempo para as classes de *backoff*. Dentre as soluções baseadas na alocação de códigos de preâmbulos, os mecanismos Gtoken e Ptoken apresentam vantagens em relação ao mecanismo Fixo em termos de controle de impacto sobre os dispositivos H2H, priorização entre os dispositivos M2M prioritário e não prioritários e tempo de acesso. Quando comparadas entre si, o mecanismo Ptoken apresenta mais vantagens do que o Gtoken. Dentre os mecanismos utilizados para determinar o intervalo de *backoff* destacam-se os mecanismos Pcronos e o Fixo. Em relação ao controle de impacto sobre o H2H, o mecanismo Pcronos apresenta tempo de acesso menor do que o Fixo, mas tem performance inferior ao Fixo em relação ao número total de acessos. Em relação aos dispositivos M2M, o mecanismo Pcronos apresenta resultados melhores do que o Fixo em termos de tempo de acesso e priorização entre os dispositivos da classe M2M.

Outra característica que se sobressai entre os mecanismos (Gtoken, Ptoken) e (Gcronos e Pcronos) é a técnica utilizada para o cálculo do vetor de demanda. A técnica empregada nos algoritmos Gtoken e Ptoken permite que os mecanismos estimem a quantidade de dispositi-

tivos sem depender de informações externas, dando assim maior capacidade de readaptação a esses mecanismos. Nos mecanismos Gcronos e Pcronos o número de dispositivos é repassado para o algoritmo no início da simulação.

5.1 Trabalhos Futuros

O desenvolvimento deste trabalho deixou, principalmente devido a falta de tempo para implementação, algumas questões e pontos para serem explorados em trabalhos futuros. Os trabalhos futuros relativos ao mecanismo PClass:

- A investigação de outras alternativas de distribuição para o cálculo da probabilidade de acesso que priorize as classes de dispositivos segundo o nível de congestionamento na rede de acesso via rádio (RAN) do LTE. A combinação de uma ou mais alternativas de distribuição podem aprimorar o cálculo da dispersão e conseqüente probabilidade de acesso dos dispositivos.
- Aprimorar o algoritmo utilizado na estação base pelo PClass para identificar o nível e congestionamento na rede de acesso via rádio.

Os trabalhos futuros relativos aos mecanismos Gtoken e Ptoken:

- Aprimorar a estratégia utilizada para definir o vetor de demanda. O vetor de demanda reflete a percepção do nível de congestionamento e influência diretamente no modo como os mecanismos Gtoken e Ptoken vão distribuir os códigos entre as diferentes classes de dispositivos. A implementação utilizada neste trabalho não faz distinção entre os preâmbulos recebidos com sucesso e os com falha (colisões) durante a contabilização desses preâmbulos. Contudo, uma colisão, por definição, é causada pela escolha de um mesmo código de preâmbulo por dois ou mais dispositivos, ou seja, o insucesso de, pelo menos, dois dispositivos. Deste modo, discriminar os preâmbulos com falhas, a começar com a atribuição de um peso duas vezes maior do que o de um preâmbulo com sucesso, apresenta-se como um ponto inicial de partida para o aprimoramento do cálculo do vetor de demanda que pode ser explorado em trabalhos futuros.

Os trabalhos futuros relativos aos mecanismos Gcronos e Pcronos:

- Aprimorar a técnica utilizada para o cálculo do vetor de demanda. Utilizar o nível de congestionamento, por exemplo, para derivar como os recursos devem ser alocados.

Além dos trabalhos já citados, outros trabalhos futuro que também podem ser explorados são: (i) aprimorar os mecanismos Gtoken, Ptoken, Gcronos e Pcronos para provê priorização intraclasse; (ii) combinar o mecanismo PClass com os mecanismos Gtoken, Ptoken, Gcronos e Pcronos.

REFERÊNCIAS BIBLIOGRÁFICAS

- 3GPP. *System Improvements for Machine-Type Communications*. [S.l.], 2012.
- 3GPP. *Study on RAN improvements for Machine-Type Communication*. [S.l.], 2014.
- AMOKRANE, A. et al. Congestion control for machine type communications. In: IEEE. *Communications (ICC), 2012 IEEE International Conference on*. [S.l.], 2012. p. 778–782.
- ARAGAO, D. B.; VIEIRA, D.; CASTRO, M. F. de. Uma proposta de controle de congestionamento ocasionado pela comunicação máquina-a-máquina em lte. In: *Proceedings of the 32th Brazilian Symposium on Computer Networks (SBRC 2015)*. [S.l.: s.n.], 2014.
- ARAGAO, D. B. et al. A dynamic backoff approach to control the congestion in m2m communication over lte networks. In: IEEE. *Global Information Infrastructure Symposium, 2015*. [S.l.], 2015.
- ARAGÃO, D. *Code improvements to the LTE-A module of the NS-3 simulator*. 2016. Disponível em: <<https://github.com/davidboto/ns3-lte>>.
- BALDO, N. et al. An open source product-oriented lte network simulator based on ns-3. In: ACM. *Proceedings of the 14th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems*. [S.l.], 2011. p. 293–298.
- CHENG, M.-Y. et al. Performance evaluation of radio access network overloading from machine type communications in lte-a networks. In: IEEE. *Wireless Communications and Networking Conference Workshops (WCNCW), 2012 IEEE*. [S.l.], 2012. p. 248–252.
- COX, C. *An introduction to LTE: LTE, LTE-advanced, SAE and 4G mobile communications*. [S.l.]: John Wiley & Sons, 2012.
- DAHLMAN, E.; PARKVALL, S.; SKOLD, J. *4G: LTE/LTE-advanced for mobile broadband*. [S.l.]: Academic press, 2013.
- FOUNDATION, P. S. *Python*. 2016. Disponível em: <<https://www.python.org/>>.
- GRIGORIK, I. *High Performance Browser Networking: What every web developer should know about networking and web performance*. [S.l.]: "O'Reilly Media, Inc.", 2013.
- HASAN, M.; HOSSAIN, E.; NIYATO, D. Random access for machine-to-machine communication in lte-advanced networks: issues and approaches. *Communications Magazine, IEEE*, v. 51, n. 6, p. 86–93, June 2013. ISSN 0163-6804.
- HERRERO, C.; VILLAR, A. The three musketeers: four classical solutions to bankruptcy problems. *Mathematical Social Sciences*, Elsevier, v. 42, n. 3, p. 307–328, 2001.
- HUSSAIN, F.; ANPALAGAN, A.; VANNITHAMBY, R. Medium access control techniques in m2m communication: survey and critical review. *Transactions on Emerging Telecommunications Technologies*, p. n/a–n/a, 2014. ISSN 2161-3915. Disponível em: <<http://dx.doi.org/10.1002/ett.2869>>.

JIAN, X. et al. A novel class-dependent back-off scheme for machine type communication in lte systems. In: IEEE. *Wireless and Optical Communication Conference (WOCC), 2013 22nd.* [S.l.], 2013. p. 135–140.

JIANG, T. et al. Evolutionary game based access class barring for machine-to-machine communications. In: *Advanced Communication Technology (ICACT), 2014 16th International Conference on.* [S.l.: s.n.], 2014. p. 832–835.

LAYA, A.; ALONSO, L.; ALONSO-ZARATE, J. Is the random access channel of lte and lte-a suitable for m2m communications? a survey of alternatives. *Communications Surveys Tutorials, IEEE*, v. 16, n. 1, p. 4–16, First 2014. ISSN 1553-877X.

LEE, K.-D.; KIM, S.; YI, B. Throughput comparison of random access methods for m2m service over lte networks. In: *GLOBECOM Workshops (GC Wkshps), 2011 IEEE.* [S.l.: s.n.], 2011. p. 373–377.

LIEN, S.-Y. et al. Cooperative access class barring for machine-to-machine communications. *Wireless Communications, IEEE Transactions on*, v. 11, n. 1, p. 27–32, January 2012. ISSN 1536-1276.

LO, A. et al. Enhanced lte-advanced random-access mechanism for massive machine-to-machine (m2m) communications. In: *27th World Wireless Research Forum (WWRF) Meeting.* [S.l.: s.n.], 2011. p. 1–5.

NS-3. *The network simulator ns-3.* [S.l.], 2014. Disponível em: <<http://www.nsnam.org>>.

O'NEILL, B. A problem of rights arbitration from the talmud. *Mathematical Social Sciences*, v. 2, n. 4, p. 345–371, 1982.

OSBORNE, M. J.; RUBINSTEIN, A. *A course in game theory.* [S.l.]: MIT press, 1994.

OSBORNE, M. J.; RUBINSTEIN, A. *A Course in game theory.* Cambridge, MA, London: MIT press, 1997. ISBN 0-262-65040-1. Disponível em: <<http://books.osborne.economics.utoronto.ca/>>.

PANG, Y.-C. et al. Network access for m2m/h2h hybrid systems: a game theoretic approach. *Communications Letters, IEEE*, v. 18, n. 5, p. 845–848, 2014. ISSN 1089-7798.

PERERA, C. et al. Context aware computing for the internet of things: A survey. *Communications Surveys Tutorials, IEEE*, v. 16, n. 1, p. 414–454, First 2014. ISSN 1553-877X.

PERERA, C. et al. Context aware computing for the internet of things: A survey. *Communications Surveys & Tutorials, IEEE, IEEE*, v. 16, n. 1, p. 414–454, 2014.

PHUYAL, U. et al. Controlling access overload and signaling congestion in m2m networks. In: *Signals, Systems and Computers (ASILOMAR), 2012 Conference Record of the Forty Sixth Asilomar Conference on.* [S.l.: s.n.], 2012. p. 591–595. ISSN 1058-6393.

PIRO, G. et al. Simulating lte cellular systems: an open-source framework. *Vehicular Technology, IEEE Transactions on, IEEE*, v. 60, n. 2, p. 498–513, 2011.

SESIA, S.; TOUFIK, I.; BAKER, M. *LTE, the UMTS long term evolution: from theory to practice*, John Wiley & Sons. [S.l.]: Feb, 2011.

WEYRICH, M.; SCHMIDT, J.-P.; EBERT, C. Machine-to-machine communication. *Software, IEEE*, v. 31, n. 4, p. 19–23, July 2014. ISSN 0740-7459.

ZHENG, K. et al. Challenges of massive access in highly dense lte-advanced networks with machine-to-machine communications. *Wireless Communications, IEEE, IEEE*, v. 21, n. 3, p. 12–18, 2014.

ZTE. *Backoff Enhancements for RAN Overload Control*. [S.l.], 2011.

APÊNDICE A – CONFIGURAÇÃO DOS EXPERIMENTOS

A simulação dos mecanismos propostos neste trabalho foram realizadas no simulador de redes baseado em eventos NS-3 (NS-3, 2014). A escolha pelo NS-3 deve-se a sua popularidade, suporte a rede LTE e por apresentar-se como uma solução de código aberto. Os cenários de simulação no NS-3 podem ser descritos em C++ ou em Python (FOUNDATION, 2016). O suporte à rede LTE foi incorporado ao NS-3 pelo projeto LENA (LTE-EPC network simulator) que propõem um módulo para a parte rádio (módulo LTE) e um módulo para o núcleo da rede (módulo EPC) (BALDO et al., 2011), que implementam os seguintes aspectos da rede LTE:

- **Módulo LTE.** Contém a pilha de protocolos de rádio (RRC - *Radio Resource Control*, PDCP - *Packet Data Convergence Protocol*, RLC - *Radio Link Control*, MAC, PHY) que são utilizados pelos equipamentos de usuários e estação base.
- **Módulo EPC.** Contém interfaces, protocolos e entidades do núcleo da rede e são utilizados pelo S-GW, P-GW e MME e parcialmente nas estações base.

Um outro simulador também muito conhecido é o LTE-Sim (PIRO et al., 2011). O LTE-Sim é um simulador de código fonte aberto voltado para rede LTE que também implementa aspectos do núcleo e da interface de rádio do LTE. O LTE-Sim tem sido muito utilizado no estudo de algoritmos de escalonamento da rede LTE. Contudo, apesar de suportar o procedimento de RACH para alguns processos (e.g., *handover*), o LTE-Sim considera que todos os dispositivos simulados já passaram pelo procedimento de RACH para o estabelecimento de um canal.

Os parâmetros de configuração comuns aos cenários de simulação são:

- **Largura de Banda (*Bandwidth*):** Identificar a largura de banda utilizada para os canais de *uplink* e *download*. A partir da largura de banda é possível identificar quantos RBs (*Resource Blocks*) são disponibilizados por quadro LTE. Para a largura de banda de 5 MHz, tem-se 25 RBs (3GPP, 2014).
- **Número de Estações Base (eNB):** Quantidade de estações base utilizadas existentes no cenário de simulação.
- **Índice de Configuração do PRACH:** O índice de configuração do PRACH é um parâmetro informado pela estação base e identifica/mapeia quais *slots* podem ser utilizados para o envio do preâmbulo de PRACH, ou seja, envio da primeira mensagem do procedimento de RACH.
- **Número de Máximo de Retransmissões de Preâmbulo (*L*):** Outro parâmetro associado ao PRACH. O número máximo de retransmissões de preâmbulos identifica o número máximo de tentativas de acesso que podem ser realizadas pelos dispositivos.

- **Códigos de Preâmbulo Disponíveis:** Identifica quais códigos de preâmbulos estão disponíveis para o procedimento de RACH, mais precisamente, quais códigos de preâmbulos podem ser enviados durante a primeira mensagem do procedimento de RACH.
- **Tempo de Janela de Resposta (RAR):** O tempo máximo de espera pela mensagem de resposta a primeira mensagem do procedimento de RACH.
- **Intervalo de Chegada:** Intervalo contido dentro do intervalo de simulação que identifica quando os dispositivos devem enviar uma solicitação de acesso.
- **Tempo de Simulação:** Tempo total de simulação.
- **Dispositivos H2H:** Identifica a quantidade de dispositivos H2H simulados.
- **Dispositivos M2M de alta prioridade ($M2M_p$):** Identifica a quantidade de dispositivos M2M prioritários simulados.
- **Dispositivos M2M de baixa prioridade ($M2M_{wp}$):** Identifica a quantidade de dispositivos M2M prioritários simulados.
- **Taxa de Chegada:** Identifica o intervalo de chegada das solicitações das diferentes classes de dispositivos.

A.1 Simulação do Mecanismo Pclass

O cenário de configuração utilizado contém uma única estação base e três classes de dispositivos, a saber, H2H, M2M de alta prioridade ($M2M_p$) e M2M de baixa prioridade ($M2M_{wp}$). A quantidade de dispositivos H2H é fixa e igual a 100. A quantidade de dispositivos $M2M_p$ e $M2M_{wp}$ simulados apresenta uma taxa de crescimento de respectivamente 125 e 500 dispositivos a cada cenário de simulação. No total tem-se oito cenários de simulação que diferem entre si em relação ao número de dispositivos simulados. As condições de parada utilizada foram o tempo de simulação e o número de acesso, incluindo as tentativas de acesso sem sucesso.

O tempo de simulação considerado foi de 5s e o intervalo de chegada dos dispositivos ficou limitado ao primeiro segundo de simulação. A largura de banda utilizada na rede foi de 5 MHz, disponibilizando assim 50 blocos de recursos por milissegundo ou (25 RBs a cada 0.5ms). A configuração de índice do PRACH utilizada determina que dois *slots* devem ser semialocados por quadro LTE. Detalhes dos mecanismos simulados estão sumarizados na Tabela A.1.

Parâmetros Gerais da Rede	
Bandwidth	5 MHz (25 RBs)
Número de Estações Base (eNB)	1
Índice de Configuração do PRACH	6 ($n = 2$)
Número de Máximo de Retransmissões de Preâmbulo (L)	10
Tempo de Janela de Resposta (RAR)	5 ms
Códigos de Preâmbulo Disponíveis	54
Intervalo de Chegada	[0, ..., 1000] ms
Parâmetros do algoritmo PClass	
Tempo de Simulação	5000 ms
Dispositivos H2H	{ 100, 100, 100, ..., 100 }
Dispositivos M2M de alta prioridade ($M2M_p$)	{ 2, 125, 250, ..., 1000 }
Dispositivos M2M de baixa prioridade ($M2M_{wp}$)	{ 10, 500, 1000, ..., 4000 }
Taxa de Chegada (H2H) - Poisson	$\lambda_{H2H} = 1/300$
Taxa de Chegada ($M2M_p$) - Poisson	$\lambda_{M2M_p} = 1/900$
Taxa de Chegada ($M2M_{wp}$) - Poisson	$\lambda_{M2M_{wp}} = 1/900$
$i = 0$ para dispositivos H2H, $i = 1$ para dispositivos M2M de alta prioridade, $i = 2$ para dispositivos M2M de baixa prioridade.	
Parâmetros do algoritmo Jian	
$T = 67.0$, $C = 2$ (classes), $k = 1$	

Tabela A.1: Parâmetros utilizados nas simulações.

A.2 Simulação dos Mecanismos Gtoken, Ptoken, Gcronos e Pcronos

O desempenho dos mecanismos Gtoken, Ptoken, Gcronos e Pcronos são analisados segundo o impacto sobre os dispositivos H2H, a priorização entre dispositivos H2H e M2M, a priorização interclasse dos dispositivos M2M. A simulação dos mecanismos propostos neste trabalho foram realizadas no simulador de redes baseado em eventos NS-3 (NS-3, 2014). A quantidade de dispositivos H2H é fixa em 200 e a dos dispositivos $M2M_p$ e $M2M_{wp}$ apresentam uma taxa de crescimento de respectivamente 50 e 300 dispositivos a cada cenário de simulação. No total tem-se seis cenários de simulação que diferem entre si em relação ao número de dispositivos simulados. Para cada cenário foram realizadas 30 simulações.

O intervalo de confiança considerado foi de 95%. O tempo de simulação considerado foi de 1s. A largura de banda utilizada foi de 5 MHz, ou seja, um total de 50 blocos de recursos são disponibilizados a cada 1ms (25 RBs a cada 0.5ms). A configuração de índice do PRACH define que dois *slots* de acesso são disponibilizados por quadro. A configuração do cenário de simulação baseia-se no cenário apresentado pela 3GPP em (3GPP, 2014) e encontra-se resumido na Tabela A.2.

Parâmetros Gerais da Rede	
Bandwidth	5 MHz (25 RBs)
Número de Estações Base (eNB)	1
Índice de Configuração do PRACH	6 ($n = 2$)
Número de Máximo de Retransmissões de Preâmbulo (L)	10
Tempo de Janela de Resposta (RAR)	5 ms
Códigos de Preâmbulo Disponíveis	52
Intervalo de Chegada	[0,...,1000] ms
Parâmetros comuns aos algoritmos <i>Gtoken</i>, <i>Ptoken</i>, <i>Gcronos</i> e <i>Pcronos</i>	
Tempo de Simulação	1000 ms
Dispositivos H2H	200
Dispositivos M2M de alta prioridade ($M2M_p$)	{50,200,350, ...,950}
Dispositivos M2M de baixa prioridade ($M2M_{wp}$)	{100,400,700 ...,1900}
Taxa de Chegada (H2H) - Poisson	$\lambda_{H2H} = 1/300$
Taxa de Chegada ($M2M_p$) - Poisson	$\lambda_{M2M_p} = 1/200$
Taxa de Chegada ($M2M_{wp}$) - Poisson	$\lambda_{M2M_{wp}} = 1/500$
Parâmetros dos algoritmos <i>Gcronos</i> e <i>Pcronos</i>	
Recursos Disponíveis: 200 ms, $\delta = 1$ ms	
Parâmetros dos algoritmos <i>Gtoken</i> e <i>Ptoken</i>	
Recursos Disponíveis: 54 códigos; $j = 4$ na Eq. 4.10; SIB2 = 10 ms(periodicidade), $\delta = 0.37$; Pesos $\alpha = 3$, $\beta = 2$, $\gamma = 1$ - <i>Ptoken</i> / $\alpha = 3$, $\beta = 2$, $\gamma = 0.265$ - <i>Gtoken</i> nas Equações 4.11, 4.12, 4.13	

Tabela A.2: Parâmetros utilizados nas simulações.

APÊNDICE B – CONTRIBUIÇÕES

Algumas funcionalidades foram implementadas para contornar as limitações do NS-3. A primeira limitação está relacionada ao número máximo de dispositivos simulados. O NS-3 limita número máximo de dispositivos por cenário de simulação em 320 dispositivos. Para viabilizar a simulação de cenários com mais de 3000 dispositivos somente o procedimento de RACH é considerado nos experimentos. A limitação quanto ao número de dispositivos está associado a periodicidade do envio do SRS(*Sounding Reference Signal*). O SRS é enviado pela estação base e solicita dos dispositivos informações sobre a qualidade do canal de *uplink*(CQI - *Channel Quality Indicator*).

O procedimento de RACH baseado em contenção implementado no módulo LTE não contempla a configuração do parâmetro *PRACH Configuration Index*. O *PRACH Configuration Index* é difundido para os dispositivos na mensagem de SIB do tipo dois e identifica a disposição/mapeamento dos *slots* disponíveis para o envio do preâmbulo de PRACH (mensagem 1) do procedimento de RACH. O LTE dispõem de um total de 64 (0-63) mapeamentos que são gerados a partir da combinação de diferentes valores para os atributos descritos abaixo:

- **Formato do Preâmbulo.** O padrão LTE suporta quatro formatos de preâmbulos, que diferem entre si em relação ao tamanho CP, tamanho da sequência do preâmbulo e tamanho do tempo de guarda. O formato do preâmbulo empregado varia segundo o raio de abrangência da estação base.
- **System Frame Number- (SFN).** Sistema de numeração dos quadros que permite aos equipamentos de usuários/dispositivos identificarem, dentre outros aspectos, a paridade do quadro, sendo este um dos parâmetros utilizados para identificar quais *slots* estão disponíveis.
- **Subquadros disponíveis.** Identificar, dentro do quadro LTE, quais subquadros de um determinado quadro podem ser utilizados para transferência do preâmbulo PRACH.

Outra funcionalidade também não implementado pelo módulo LTE é o parâmetro *Backoff Indicator* (BI), enviado na mensagem RAR (mensagem 2) do procedimento de RACH baseado em contenção.

Para viabilizar a simulação dos mecanismos propostos neste trabalho de acordo com o cenário apresentado pela 3GPP em (3GPP, 2014), foram implementadas as funcionalidades de configuração do índice de configuração do PRACH e do indicador de *backoff* no módulo LTE no simulador. O código fonte da implementação pode ser encontrado em (ARAGÃO, 2016).