



Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Programa de Pós-Graduação em Engenharia Elétrica

# Gerenciamento Dinâmico de Energia com Política do Tempo Ótimo de Desligamento

Bruna Maria Justino Cruz Vilar

Tese de Doutorado apresentada à Coordenadoria do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Campina Grande - Campus de Campina Grande como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências no Domínio da Engenharia Elétrica.

Área de Concentração: Processamento da Informação

Angelo Perkusich, D. Sc.

Orientador

Saulo Oliveira Dornellas Luiz, D. Sc.

Orientador

Campina Grande, Paraíba, Brasil

©Bruna Maria Justino Cruz Vilar, 05 de Julho de 2019

V697g

Vilar, Bruna Maria Justino Cruz.

Gerenciamento dinâmico de energia com política do tempo ótimo de desligamento / Bruna Maria Justino Cruz Vilar. – Campina Grande, 2019.

96 f. : il. color.

Tese (Doutorado em Engenharia Elétrica) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2019.

"Orientação: Prof. Dr. Angelo Perkusich, Prof. Dr. Saulo Oliveira Dornellas Luiz".

Referências.

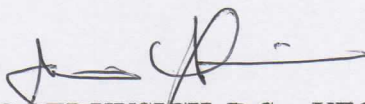
1. Processamento de Energia. 2. Gerenciamento Dinâmico de Energia. 3. Política do Tempo Ótimo de Desligamento. I. Perkusich, Angelo. II. Luiz, Saulo Oliveira Dornellas. III. Título.

CDU 621.311(043)

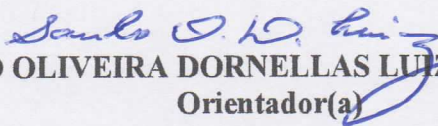
**"GERENCIAMENTO DINÂMICO DE ENERGIA COM POLÍTICA DO TEMPO ÓTIMO DE  
DESLIGAMENTO "**

**BRUNA MARIA JUSTINO CRUZ VILAR**

**TESE APROVADA EM 05/07/2019**

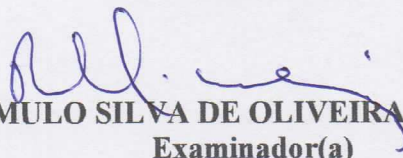


**ANGELO PERKUSICH, D.Sc., UFCG**  
**Orientador(a)**

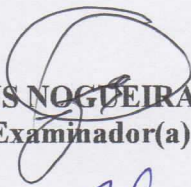


**SAULO OLIVEIRA DORNELLAS LUIZ, D.Sc., UFCG**  
**Orientador(a)**

**PAULO SÉRGIO SAUSEN, D.Sc., UNIJUÍ**  
**Examinador(a)**



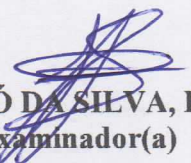
**RÔMULO SILVA DE OLIVEIRA, Dr., UFSC**  
**Examinador(a)**



**ANTONIO MARCUS NOGUEIRA LIMA, Dr., UFCG**  
**Examinador(a)**



**JOSÉ SÉRGIO DA ROCHA NETO, D.Sc., UFCG**  
**Examinador(a)**



**JAIDILSON JÓ DA SILVA, D.Sc., UFCG**  
**Examinador(a)**

**CAMPINA GRANDE - PB**

Aos queridos Pablo e Agatha,  
e aos meus pais José Roberto e Zélia, *dedico*.

## Agradecimentos

Meu primeiro sentimento ao concluir esta longa jornada é gratidão. Em primeiro lugar, eu agradeço a Deus e Nossa Senhora por todas as bênçãos e oportunidades concedidas.

Agradeço aos professores Saulo Dornellas e Angelo Perkusich, pela incansável orientação e colaboração sem as quais não seria possível realizar este trabalho.

Agradeço aos meus irmãos, sobrinhos e cunhados por todo o carinho. De forma muito especial, agradeço à minha irmã Jocelma. Sem ela esse caminho não teria sido trilhado. Muito obrigada por sempre ter me visto com olhos tão gentis e amorosos. Agradeço aos meus amados pais, pelo amor incondicional e por sempre acreditarem em mim. Sem a colaboração deles eu também não teria conseguido concluir este trabalho.

Toda a minha gratidão ao meu amado esposo Pablo. São longos anos de companheirismo, amizade, cumplicidade, respeito e amor. Obrigada por ser minha sanidade nos dias de desespero e por todos os momentos maravilhosos divididos desde os tempos de escola.

De maneira muito especial, eu gostaria de agradecer à Agatha. Minha filha amada, eu não sei como vivi tantos anos sem você. Obrigada por ter chegado na hora certa, por ter me dado o incentivo e a força que eu precisava para concluir este trabalho.

Aos funcionários da COPELE, em especial a Ângela e Pedro, por todo o apoio, paciência e diligência durante toda a pós-graduação.

Ao CNPq pelo suporte financeiro desta pesquisa.

# Resumo

Os avanços constantes nas tecnologias de *hardware* e *software* têm permitido que os sistemas computacionais móveis se tornem cada vez mais necessários no nosso dia a dia. O desenvolvimento de aplicativos cada vez mais complexos para esses sistemas computacionais, tais como os de edição de vídeos e jogos, exigem maior capacidade de processamento. O aumento do número e da complexidade desses aplicativos aliado à necessidade de conectividade por meio de alguma interface de rede, implicam em maior consumo de energia. Entretanto, os dispositivos móveis dispõem de fonte de energia limitada, e por isso, o uso eficiente da energia é um fator determinante para uma boa experiência do usuário.

O Gerenciamento Dinâmico de Energia (GDE) é uma maneira eficaz de reduzir o consumo de energia e é amplamente utilizado em sistemas reais. A ação mais comum no gerenciamento é o desligamento de dispositivos que estejam ociosos. Embora a ação de desligar indique redução no consumo, a energia necessária para religar é muitas vezes alta o bastante para justificar que nem sempre a melhor opção, durante um intervalo ocioso, seja desligar dispositivo.

Neste trabalho de doutorado é desenvolvida uma política para o gerenciamento dinâmico de energia, denominada Política do Tempo Ótimo de Desligamento (TOD). A carga de trabalho é modelada como uma cadeia de Markov de dois estados, *ativo* e *ocioso*. Considera-se neste trabalho que exista um *tempo ótimo* para o qual o dispositivo com gerenciamento dinâmico de energia deverá permanecer desligado ou em um estado de baixo consumo. Entretanto, esse tempo não é necessariamente todo o intervalo ocioso, pois as durações dos intervalos ociosos variam ao longo do tempo. São desenvolvidos modelos analíticos para penalidade de desempenho e para o consumo de energia. Um problema de otimização é formulado e a tabela de políticas é definida. De forma que, para cada carga de trabalho conhecida, é determinado o *tempo ótimo de desligamento*,  $T$ , de acordo com métricas de economia de energia e penalidade de desempenho. Em tempo de execução, o modelo da carga de trabalho é identificado, e a partir de uma consulta à tabela de políticas, é definido qual o  $T$  ótimo para a situação, o que permitirá o religamento preditivo, minimizando o consumo de energia, e atendendo uma restrição de penalidade de desempenho.

# Abstract

The constant advances in hardware and software technologies allowed mobile computing systems to become increasingly necessary in our everyday lives. The development of increasingly complex applications for these computer systems, such as video editing and gaming, requires more processing power. The increase in number and complexity of these applications coupled with the need for connectivity by means of some network interface means higher power consumption. However, mobile devices have a limited power source, so efficient use of energy is a determining factor for a good user experience.

The Dynamic Power Management (DPM) is an effective way to reduce power consumption and is widely used in real systems. The most common action in power management is the shutdown of idle devices. Although the turn off action indicates a reduction in consumption, the turn on power consumption is often high enough to justify that not always it is the best option in an idle interval to change the state of consumption of the that are not being used effectively.

In this doctoral work, a policy for dynamic energy management, named the Optimal Shutdown Time Policy (TOD) is developed. The workload is modeled as a two state Markov chain, *active* and *idle*. It is considered in this work that there is an *optimal time* for which the device with dynamic power management should remain off or in a low power state. However, this time is not necessarily the entire idle interval, since the lengths of the idle intervals vary over time. Analytical models are developed for performance penalty and for energy consumption. An optimization problem is formulated and the policy table is defined. So, for each known workload, the *optimal shutdown time*,  $T$ , is determined based on energy saving metrics and performance penalty. At runtime, the workload model is identified and, using a policy table, it is defined the optimal  $T$  for this situation, which will allow predictive wake-up, minimizing energy consumption and meeting a certain performance penalty constraint.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Motivação . . . . .	1
1.2	Objetivo geral . . . . .	3
1.3	Objetivos específicos . . . . .	4
1.4	Contribuição . . . . .	4
1.5	Organização do texto . . . . .	6
<b>2</b>	<b>Revisão Bibliográfica</b>	<b>7</b>
2.1	Introdução . . . . .	7
2.2	Políticas de gerenciamento de energia . . . . .	9
2.2.1	Políticas de timeout . . . . .	9
2.2.2	Políticas preditivas . . . . .	12
2.2.3	Políticas estocásticas . . . . .	14
2.3	Métricas de penalidade . . . . .	15
2.4	Considerações finais . . . . .	18
<b>3</b>	<b>Gerenciamento dinâmico de energia</b>	<b>19</b>
3.1	Gerenciamento dinâmico de energia para cargas de trabalho estacionárias . .	19
3.1.1	Componentes do sistema . . . . .	20
3.2	Gerenciamento dinâmico de energia para cargas de trabalho não estacionárias	27
3.2.1	Identificação da carga de trabalho . . . . .	28
3.2.2	Multi-Janelas . . . . .	29
3.3	Considerações finais . . . . .	31



<b>4</b>	<b>Política do tempo ótimo de desligamento</b>	<b>32</b>
4.1	Introdução . . . . .	32
4.2	Política do tempo ótimo de desligamento . . . . .	32
4.2.1	Modelos de energia . . . . .	35
4.3	Política do tempo ótimo de desligamento . . . . .	39
4.3.1	Modelo analítico de energia . . . . .	41
4.3.2	Modelo analítico de penalidade de desempenho . . . . .	43
4.3.3	Obtenção do tempo ótimo de desligamento $T$ . . . . .	44
4.3.4	Tabela de políticas . . . . .	46
4.3.5	Simulação da política de energia . . . . .	47
4.4	Projeto de experimento . . . . .	48
4.4.1	Determinando o intervalo de confiança para $\rho = 95\%$ . . . . .	49
4.4.2	Determinando o valor de $n$ para obter $\rho = 95\%$ . . . . .	50
4.5	Considerações finais . . . . .	58
<b>5</b>	<b>Estudo de Caso</b>	<b>60</b>
5.1	Modelo de energia para a interface de rede . . . . .	60
5.2	Carga de trabalho para interfaces de rede . . . . .	63
5.3	Obtenção da política do tempo ótimo de desligamento . . . . .	64
5.4	Comparação de políticas . . . . .	68
5.4.1	Comparações no cenário estacionário . . . . .	68
5.4.2	Comparações no cenário não-estacionário . . . . .	75
5.4.3	Cargas de trabalho experimentais . . . . .	85
5.5	Implementação da política . . . . .	87
5.6	Considerações finais . . . . .	89
<b>6</b>	<b>Conclusão</b>	<b>90</b>
6.1	Sugestões para Trabalhos Futuros . . . . .	92
	<b>Referências bibliográficas</b>	<b>93</b>

# Glossário

BRA *Buffer* da Requisição Anterior

FS Fila de Serviços

GDE Gerenciamento Dinâmico de Energia

GE Gerenciador de energia

JMC Janelas Deslizantes de Múltiplos Comprimentos

PO Problema de Otimização

PS Provedor de Serviços

RS Requisitante de Serviços

TOD Tempo Ótimo de Desligamento

USB *Universal Serial Bus*

# Lista de Tabelas

4.1	Tabela de políticas. . . . .	46
5.1	Dados da interface de rede. . . . .	62
5.2	Tabela de políticas. . . . .	67

# Lista de Figuras

2.1	Esquema de gerenciamento para carga de trabalho e dispositivo, com intervalo ocioso longo. . . . .	7
2.2	Esquema de gerenciamento para carga de trabalho e dispositivo, com intervalo ocioso curto. . . . .	8
2.3	Ilustração do valor do $T_{be}$ necessário para compensar a energia adicional consumida na transição de estados. . . . .	9
2.4	Ilustração do valor da potência quando o intervalo ocioso é menor que o <i>timeout</i> selecionado . . . . .	10
2.5	Ilustração do valor da potência quando o intervalo ocioso é maior ou igual ao <i>timeout</i> selecionado. . . . .	11
2.6	Ilustração de intervalos ociosos versus intervalos ativos, denominada curva L.	13
2.7	Representação dos períodos de atrasos devido ao uso de uma técnica de GDE.	16
2.8	Ilustração da insatisfação do usuário. . . . .	17
3.1	Componentes do modelo de um sistema com gerenciamento de energia. . . .	20
3.2	Modelo da cadeia de Markov para o requisitante de serviços. . . . .	21
3.3	Exemplo de um RS não-estacionário. . . . .	28
3.4	Valor de $R_i$ ao longo do tempo para uma carga de trabalho não-estacionária.	28
3.5	Diagrama de blocos para a técnica de gerenciamento dinâmico de energia proposta. . . . .	29
3.6	Exemplo de uso de multi-janela para RS de dois estados. . . . .	30
3.7	O erro de amostragem (linha de maior espessura) após a transição, em $t_s$ , de uma carga de trabalho estacionária $u_r$ para outra $u_s$ . . . . .	30

4.1	Exemplo de um modelo para um requisitante de serviços com dois estados, <i>ativo e ocioso</i> . . . . .	33
4.2	Exemplo de um intervalo ocioso. . . . .	33
4.3	Estados de consumo de um provedor de serviços com gerenciamento de energia, onde $h$ é o tempo mínimo necessário para identificação de um intervalo ocioso: (a) nenhum desligamento (b) religamento preditivo; (c) religamento preditivo (antecipado); (d) religamento realizado pelo usuário; (e) quando o desligamento ocorre no momento da transição do intervalo ocioso para o intervalo ativo. . . . .	36
4.4	Consumo de energia para um provedor de serviços quando o gerenciador de energia emite um comando para desligar e o intervalo ocioso é maior ou igual ao tempo ótimo de desligamento. . . . .	40
4.5	Consumo de energia para um provedor de serviços quando o gerenciador de energia emite um comando para desligar e o intervalo ocioso é menor que o tempo ótimo de desligamento. . . . .	40
4.6	Exemplo do valor esperado da energia versus a penalidade de desempenho. . . . .	45
4.7	Interpolação linear para uma penalidade de desempenho especificada. . . . .	47
4.8	Diagrama de blocos para simulação de uma política de energia. . . . .	47
4.9	Valor $n$ de repetições para cada $T$ , para a carga de trabalho com $(q = 0, 1)$ . . . . .	55
4.10	Valor $n$ de repetições para cada $T$ , para a carga de trabalho com $(q = 0, 3)$ . . . . .	55
4.11	Valor $n$ de repetições para cada $T$ , para a carga de trabalho com $(q = 0, 6)$ . . . . .	56
4.12	Valor $n$ de repetições <i>versus</i> cargas de trabalho. . . . .	57
4.13	Ilustração da composição dos intervalos ociosos e intervalos ativos ao longo do intervalo de tempo $N$ . . . . .	57
5.1	Diagrama de blocos para medição da tensão e corrente demandados pelo adaptador Wireless USB. . . . .	61
5.2	Imagens do Micro Adaptador USB - Wi-Fi 150Mps . . . . .	61
5.3	Diagrama elétrico do circuito elétrico para monitoramento da potência consumida. . . . .	61

5.4	Representação da potência média consumida pela interface ativa, $P_{on}$ , e desligada, $P_{off}$ . . . . .	62
5.5	Representação da potência média consumida pela interface desligada, $P_{off}$ , e ativa, $P_{on}$ . . . . .	63
5.6	Intervalo de tempo de uma carga de trabalho gerada pelo uso de uma interface de rede. . . . .	64
5.7	Função custo para a seleção do limiar de ociosidade da interface de rede. . .	65
5.8	Resultados de energia e penalidade de desempenho para a carga de trabalho leve, ( $q = 0, 1$ ). . . . .	66
5.9	Resultados de energia e penalidade de desempenho para a carga de trabalho intermediária, ( $q = 0, 4$ ). . . . .	66
5.10	Resultados de energia e penalidade de desempenho para a carga de trabalho intensa, ( $q = 0, 6$ ). . . . .	67
5.11	Curva da Energia <i>versus</i> Penalidade de desempenho para a carga de trabalho ( $q = 0, 1$ ). . . . .	69
5.12	Curva da Energia <i>versus</i> Penalidade de desempenho para carga de trabalho ( $q = 0, 4$ ). . . . .	70
5.13	Estimação da carga de trabalho: (a) carga de trabalho com ( $q = 0, 4$ ); (b) $\hat{q}$ estimado; (c) comprimentos de janelas utilizados; (d) valor estimado do intervalo ocioso $1/\hat{q}$ . . . . .	71
5.14	Curva da Energia <i>versus</i> Penalidade de desempenho para carga de trabalho ( $q = 0, 6$ ). . . . .	72
5.15	Estimação da carga de trabalho: (a) carga de trabalho com ( $q = 0, 6$ ); (b) $\hat{q}$ estimado; (c) comprimentos de janelas utilizados; (d) valor estimado do intervalo ocioso $1/\hat{q}$ . . . . .	72
5.16	Curva da Energia <i>versus</i> Penalidade de desempenho para carga de trabalho ( $q = 0, 8$ ). . . . .	73
5.17	Estimação da carga de trabalho: (a) carga de trabalho com ( $q = 0, 8$ ); (b) $\hat{q}$ estimado; (c) comprimentos de janelas utilizados; (d) valor estimado do intervalo ocioso $1/\hat{q}$ . . . . .	74

5.18	Curva da Energia <i>versus</i> Penalidade de desempenho para carga de trabalho ( $q = 0,9$ ). . . . .	75
5.19	Curva da Energia <i>versus</i> Penalidade de desempenho para a carga de trabalho não-estacionária (5.2). . . . .	76
5.20	Estimativas para a carga de trabalho (5.2) utilizando a técnica JMC. . . . .	77
5.21	Curva da Energia <i>versus</i> Penalidade de desempenho para a carga de trabalho não-estacionária (5.3). . . . .	78
5.22	Estimativas para a carga de trabalho (5.3) utilizando a técnica JMC. . . . .	79
5.23	Curva da Energia <i>versus</i> Penalidade de desempenho para a carga de trabalho não-estacionária (5.4). . . . .	79
5.24	Estimativas para a carga de trabalho (5.4) utilizando a técnica JMC. . . . .	80
5.25	Curva da Energia <i>versus</i> Penalidade de desempenho para a carga de trabalho não-estacionária (5.5). . . . .	81
5.26	Curva da Energia <i>versus</i> Penalidade de desempenho para a carga de trabalho não-estacionária (5.6). . . . .	82
5.27	Estimativas para a carga de trabalho (5.6) utilizando a técnica JMC. . . . .	82
5.28	Curva da Energia <i>versus</i> Penalidade de desempenho para a carga de trabalho não-estacionária (5.7). . . . .	83
5.29	Estimativas para a carga de trabalho (5.7) utilizando a técnica JMC. . . . .	84
5.30	Curva da Energia <i>versus</i> Penalidade de desempenho para a carga de trabalho não-estacionária (5.8). . . . .	84
5.31	Estimativas para a carga de trabalho (5.8) utilizando a técnica JMC. . . . .	85
5.32	Curva da Energia <i>versus</i> Penalidade de desempenho para carga de trabalho (5.9) . . . . .	86
5.33	Estimação da carga de trabalho: (a) carga de trabalho não-estacionária; (b) $q$ estimado; (c) comprimentos de janelas utilizados; (d) valor estimado do intervalo ocioso $1/\hat{q}$ . . . . .	87
5.34	Arquitetura do <i>software</i> proposto para o gerenciamento dinâmico de energia. . . . .	88
5.35	Fluxograma da técnica proposta para o gerenciamento dinâmico de energia. . . . .	88

# Capítulo 1

## Introdução

### 1.1 Motivação

Dispositivos móveis, tais como *smartphones*, tem sua autonomia limitada pelo tamanho e capacidade de armazenamento de energia em suas fontes de alimentação. Por isso, o consumo de energia é uma preocupação em tais dispositivos [1]. Uma opção para prolongar a autonomia é utilizar uma mesma quantidade de energia durante um intervalo de tempo maior, isto é, reduzir o consumo de energia por meio de alterações no desempenho do dispositivo [2].

O *Gerenciamento Dinâmico de Energia* (GDE) permite reduzir o consumo de energia a partir de alterações no estado de consumo de dispositivos que encontram-se ociosos ou sub-utilizados. Uma ação de GDE simples e que é aplicável para quase todos os dispositivos é o desligamento [3]. Entretanto, religar um dispositivo é uma ação que pode demandar tempo e energia extra. De forma que, quando um desligamento desnecessário é realizado, é possível até aumentar o consumo de energia [4].

As abordagens para o gerenciamento dinâmico de energia podem ser classificadas em três grandes categorias: baseadas em *timeout*, preditivas e estocásticas [5–8]. Alternativamente, as abordagens podem também ser classificadas em duas categorias: heurísticas (*timeout* e preditivas) e estocásticas [9, 10].

A política de *timeout* é amplamente utilizada em muitas aplicações, tais como: micro-processadores, monitores e disco rígido, por causa de sua simplicidade [5, 8, 11]. Existem



duas estratégias dessa política: *timeout* fixo e *timeout* dinâmico [8].

Em uma política de *timeout* fixo, o comando para desligar ocorre quando é verificado que o dispositivo encontra-se ocioso por um intervalo de tempo específico denominado *timeout*. Com essa política, o usuário pode definir qual o melhor *timeout* manualmente. Uma política de *timeout* dinâmico pode ser mais eficiente, pois o valor do *timeout* é alterado de acordo com outras informações, como por exemplo, a partir da observação dos intervalos ociosos passados. A principal vantagem desse tipo de política é a simplicidade, visto que se resume à contagem do intervalo de tempo ocioso. A principal desvantagem dessa política é que deixa-se de economizar energia enquanto o *timeout* não expira.

Políticas preditivas podem ser classificadas em duas categorias: desligamento preditivo e religamento preditivo. Essas políticas foram propostas com o objetivo de resolver as desvantagens das políticas de *timeout* [8]. Em uma política de desligamento preditivo, a predição do próximo intervalo ocioso é baseada em observações feitas sobre o histórico de uso do dispositivo. Caso a previsão de intervalo ocioso seja maior que um determinado *threshold*, o estado de consumo do dispositivo é alterado para um estado de baixo consumo tão cedo ele inicie um intervalo de ociosidade. De forma semelhante, em uma política de religamento preditivo, se é previsto que o intervalo ocioso está prestes a acabar, o dispositivo é religado, mesmo que não tenham chegado requisições da carga de trabalho. Ambas as categorias da política preditiva podem ser penalizadas pela imprecisão na predição do intervalo ocioso.

Uma característica comum entre as políticas de *timeout* e preditivas é o fato de serem formuladas heurísticamente e, em seguida, testadas por meio de simulações ou experimentos para avaliar seu desempenho. Políticas estocásticas modelam as chegadas das requisições e as mudanças no estado de energia do dispositivo como processos estocásticos, tais como processos de decisão de Markov. Nesse caso, o problema de minimizar o consumo de energia é formulado como um problema de otimização estocástico.

No contexto de políticas estocásticas, Benini et al. [12] modelaram a chegada de requisições e as transições de estado como cadeias de Markov estacionárias e discretas no tempo. Observando as limitações ao considerar apenas o caso estacionário, Chung et al. [7] propuseram uma solução mais abrangente para estimação de cargas de trabalho variantes no tempo, baseada em janelas deslizantes. Entretanto, essa solução apresenta alguns problemas como o erro de amostragem e o tempo de adaptação de uma carga estacionária para outra. Ren et

al. [13] modificaram a abordagem proposta por Chung et al. [7] particularizando o problema de identificação da carga de trabalho, propondo um modelo no qual o gerenciador de energia solicita do sistema operacional a informação sobre a carga de trabalho. Luiz et al. [14] desenvolveram uma solução mais abrangente que a proposta por Chung et al. [7] ao utilizar janelas deslizantes de múltiplos comprimentos para estimação da carga de trabalho, onde os problemas com erro de amostragem e tempo de adaptação são reduzidos significativamente.

Neste trabalho de doutorado é desenvolvida uma política para o gerenciamento dinâmico de energia, denominada Política do Tempo Ótimo de Desligamento (TOD). A carga de trabalho é modelada como uma cadeia de Markov de dois estados, *ativo* e *ocioso*. Considera-se neste trabalho que existe um *tempo ótimo* para o qual o dispositivo com gerenciamento dinâmico de energia deverá permanecer desligado/estado de baixo consumo. Entretanto, esse tempo não é necessariamente todo o intervalo ocioso, pois as durações dos intervalos ociosos variam ao longo do tempo. São desenvolvidos modelos analíticos para penalidade de desempenho e para consumo de energia. Um problema de otimização é formulado e a tabela de políticas é definida. De forma que, para cada carga de trabalho conhecida à priori, é determinado o *tempo ótimo de desligamento*,  $T$ , atendendo métricas de economia de energia e penalidade de desempenho. Em tempo de execução, é identificado o modelo da carga de trabalho e a partir de uma consulta à tabela de políticas, é definido qual o  $T$  ótimo para tal situação, que permitirá o religamento preditivo, minimizando o consumo de energia, e atendendo uma determinada restrição de penalidade de desempenho.

## 1.2 Objetivo geral

O objetivo geral neste trabalho é minimizar o consumo de energia em dispositivos que suportem pelo menos dois estados de consumo, como *on* e *off* (respectivamente ligado e desligado), e experimentem cargas de trabalho que se aderem a modelos de cadeias de Markov. Para atender esse objetivo foi desenvolvida uma política para o gerenciamento dinâmico de energia.

## 1.3 Objetivos específicos

Os objetivos específicos são:

- desenvolvimento do modelo analítico para o consumo de energia;
- desenvolvimento do modelo analítico para penalidade de desempenho;
- formulação e resolução de um problema de otimização;
- desenvolvimento de uma política de gerenciamento dinâmico de energia;
- validação da política proposta através de comparações com outras políticas, por meio de simulações.

## 1.4 Contribuição

O gerenciamento dinâmico de energia permite reduzir o consumo de energia a partir de alterações no estado de consumo de dispositivos que encontram-se ociosos ou sub-utilizados. Diversas abordagens já foram propostas, como as políticas de *timeout*, preditivas e estocásticas.

Quando um dispositivo tem seu estado de energia alterado, isto é, no desligamento ou religamento, o consumo medido está relacionado ao tempo em que tal dispositivo permanece no estado anterior, às energias de transição entre os estados e ao estado atual de consumo que o dispositivo se encontra. Por exemplo, se uma política de *timeout* emitir um comando de *desligar* para um dispositivo que esteja atendendo uma requisição da carga de trabalho, isto é, não encontra-se ocioso, além dessa ação computar um penalidade de desempenho, ela também contribuirá para aumentar o consumo de energia. Dessa forma, cada situação é analisada individualmente para compor um modelo analítico de energia e um modelo analítico de penalidade de desempenho para um dispositivo genérico que esteja sob gerenciamento dinâmico de energia e suporte dois estados de consumo: *on* e *off*.

O religamento preditivo é uma ação que contribui para a redução da penalidade de desempenho no sistema, na medida que religa o dispositivo sob gerenciamento antes que este seja solicitado. Entretanto, religamentos demasiadamente preditivos não contribuem de

forma significativa para a economia de energia. Determinar por quanto tempo um dispositivo deve permanecer em estado de baixo consumo é tarefa difícil de ser definida de forma empírica. Ao modelar a carga de trabalho como uma cadeia de Markov de dois estados, *ativo* e *ocioso*, e analisar extensivamente o comportamento dessa carga de trabalho, observa-se que, cada carga apresenta um comportamento padrão, caracterizado pelas probabilidades de transição da cadeia de Markov. Tais probabilidades estão relacionadas com a intensidade da carga de trabalho, e mais que isso, fornecem informações sobre o comportamento dos intervalos ociosos para tal carga.

Neste trabalho busca-se determinar o *tempo ótimo de desligamento* para um dispositivo sob gerenciamento de energia. Esse tempo é obtido a partir do modelo da carga de trabalho e atende a métricas de redução de consumo de energia e penalidade de desempenho. A proposta é desenvolver uma política considerando um modelo estocástico da carga de trabalho e modelos analíticos da energia e da penalidade de desempenho, a partir dos quais é formulado e resolvido o problema de otimização, mas com característica preditiva quanto ao religamento do dispositivo.

O problema de otimização definido estabelece os valores do *tempo ótimo de desligamento*,  $T$ , para cada carga de trabalho, que minimizam o consumo de energia, para uma dada restrição de penalidade de desempenho, onde  $L$  é a penalidade máxima aceitável e uma restrição de consumo definida por  $\mathcal{E}_{on}$ , que é o consumo máximo aceitável. A política é validada através de simulações para cargas de trabalho diversas e com durações definidas a partir do teorema do limite central [15].

É apresentado um estudo de caso, onde a política proposta foi aplicada para um dispositivo de rede, Wi-Fi, e comparada através de simulações com as políticas de *timeout* e preditiva. A política proposta não está restrita ao estudo de caso apresentado neste trabalho, abrindo possibilidades para ser utilizada em outros dispositivos dentro de um sistema computacional, desde que se verifiquem variações da carga de trabalho ao longo do tempo e o dispositivo suporte ao menos dois estados de consumo, como por exemplo: *on* e *off*.

## 1.5 Organização do texto

O texto deste trabalho está organizado da seguinte forma:

Neste Capítulo foi realizada uma introdução da tese, sendo apresentados a motivação, os objetivos e as contribuições.

No Capítulo 2 é apresentada a revisão bibliográfica sobre trabalhos que abordam o gerenciamento dinâmico de energia utilizando políticas de *timeout*, preditivas e estocásticas. No fim do Capítulo é apresentada uma breve consideração sobre penalidade de desempenho do ponto de vista do usuário.

No Capítulo 3 é apresentada uma revisão da abordagem básica de gerenciamento dinâmico de energia introduzida por Benini et al. [12] e Chung et al. [7], que é eficiente para sistemas computacionais que podem ser modelados por meio de processos de decisão de Markov. Motivados pela necessidade de estimar uma carga de trabalho inicialmente desconhecida e não-estacionária, é apresentada a técnica de estimação da carga de trabalho baseada em janelas deslizantes de múltiplos comprimentos (JMC), introduzida por Luiz et al. [14, 16].

No Capítulo 4 a carga de trabalho é modelada como uma cadeia de Markov de dois estados, *ativo* e *ocioso*. São formuladas as expressões analíticas para o modelo de energia e o modelo de penalidade de desempenho para um dispositivo que suporte dois estados de consumo, *on* e *off*. O problema de otimização é formulado e resolvido. É definida a tabela de políticas, o interpolador de políticas e apresentado o algoritmo para simulação da política proposta.

No Capítulo 5 é apresentado um estudo de caso, onde a política proposta é aplicada para um dispositivo de rede, Wi-Fi. São apresentados a plataforma experimental, o dispositivo de rede, e o processo de aquisição da carga de trabalho. O problema de otimização é formulado e resolvido e a política proposta é então comparada com uma política de *timeout* e uma política preditiva, em vários cenários, para validar seu desempenho.

No Capítulo 6 são apresentadas as conclusões gerais em relação ao trabalho realizado e também são apresentadas as sugestões para trabalhos futuros.

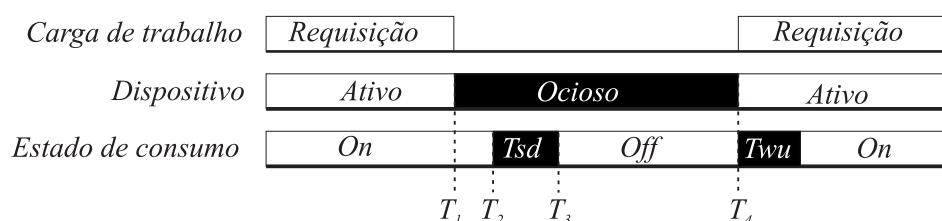
# Capítulo 2

## Revisão Bibliográfica

### 2.1 Introdução

O *Gerenciamento Dinâmico de Energia* (GDE) permite reduzir o consumo de energia a partir de alterações no estado de consumo de dispositivos que encontram-se ociosos ou subutilizados [8]. O conceito de um Gerenciador de Energia (GE) pode ser ilustrado com a Figura 2.1.

Figura 2.1: Esquema de gerenciamento para carga de trabalho e dispositivo, com intervalo ocioso longo.



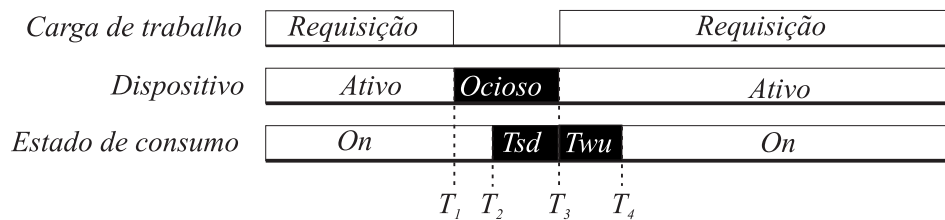
Fonte: Adaptado de Lu e De Micheli [5].

A carga de trabalho consiste de requisições. Em discos rígidos, requisições são comandos de leitura e escrita; em interfaces de rede, são pacotes de dados enviados e recebidos. Quando ocorre uma requisição, o dispositivo encontra-se ativo, caso contrário, ocioso [5]. Entre  $T_1$  e  $T_4$  não há requisição da carga de trabalho e o dispositivo encontra-se ocioso. Durante esse intervalo de tempo é possível utilizar alguma técnica de gerenciamento de energia para

alterar o estado de consumo do dispositivo. Um gerenciador de energia monitora e controla os estados de energia do sistema e dos componentes. Esse procedimento é chamado de política de gerenciamento de energia [17], cujos estados de energia são selecionados em função da carga de trabalho do sistema.

Em  $T_2$  o dispositivo é desligado e em  $T_4$  ligado. Os tempos  $T_{sd}$  e  $T_{wu}$  são os atrasos associados com alterações de estado do dispositivo. Há situações em que não é viável realizar uma alteração no estado de consumo. Como exemplo, na Figura 2.2, a duração do intervalo ocioso não justifica o desligamento.

Figura 2.2: Esquema de gerenciamento para carga de trabalho e dispositivo, com intervalo ocioso curto.



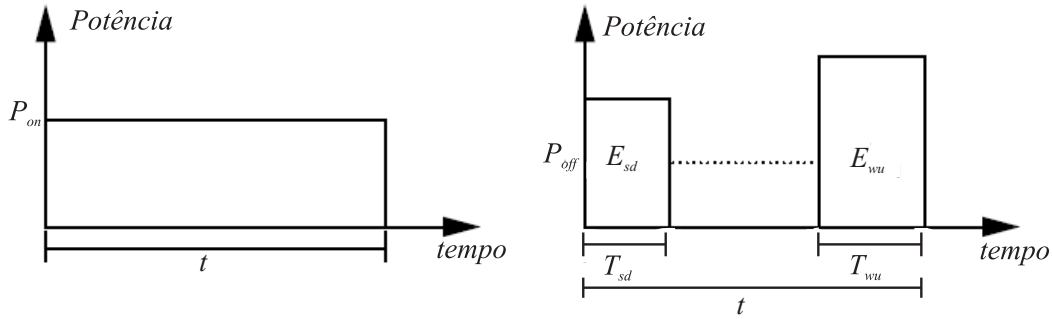
Fonte: Adaptado de Lu e De Micheli [5].

Alterações no estado de consumo do dispositivo, tais como o desligamento e o religamento, provocam um acréscimo no consumo de energia do sistema [18]. Dessa forma, quando uma transição acontece é necessário que o dispositivo permaneça no estado de baixo consumo por um certo tempo, de forma a compensar o acréscimo de energia causado pela transição de estados do mesmo [19].

Quando o intervalo ocioso é longo o suficiente para amortizar a energia adicional devido à transição de estado, o dispositivo pode ser colocado em estado de baixo consumo ou ser desligado para economizar energia. Assim, pode-se definir um limiar, denominado *Break-even time*,  $T_{be}$ , que é o tempo ocioso necessário para igualar a economia de energia à energia de transição [8]. O valor do  $T_{be}$  é definido em (2.1). Uma ilustração do valor do  $T_{be}$  necessário para compensar a energia adicional consumida na transição de estados é apresentada na Figura 2.3. As constantes  $E_{sd}$  e  $E_{wu}$ , representam as energias para desligar e religar, respectivamente, um dispositivo. Os tempos  $T_{sd}$  e  $T_{wu}$  representam os tempos necessários para que um dispositivo desligue e religue, respectivamente. A constante  $P_{on}$

representa a potência média do dispositivo ligado, e  $P_{off}$  é a potência média do dispositivo desligado.

Figura 2.3: Ilustração do valor do  $T_{be}$  necessário para compensar a energia adicional consumida na transição de estados.



Fonte: Adaptado de Lu e De Micheli [5].

$$\begin{aligned}
 P_{on} \cdot t &\geq E_{sd} + E_{wu} + P_{off} \cdot (t - T_{sd} - T_{wu}) \\
 t &\geq \frac{E_{sd} + E_{wu} - P_{off} \cdot (T_{sd} + T_{wu})}{P_{on} - P_{off}} \\
 T_{be} &= \frac{E_{sd} + E_{wu} - P_{off} \cdot (T_{sd} + T_{wu})}{P_{on} - P_{off}}
 \end{aligned} \tag{2.1}$$

## 2.2 Políticas de gerenciamento de energia

A mais simples e também mais conhecida ação utilizada no GDE é o desligamento de dispositivos eletrônicos após o recebimento de um comando [20]. Muitos esforços em pesquisas já foram dedicados a explorar diferentes técnicas de gerenciamento dinâmico de energia. Tais técnicas podem ser classificadas em três grandes categorias: baseadas em *timeout*, preditivas e estocásticas [5–8].

### 2.2.1 Políticas de timeout

Políticas de gerenciamento dinâmico de energia com base em *timeout* são amplamente aplicadas para dispositivos eletrônicos de consumo, tais como: o tempo de espera para desligar

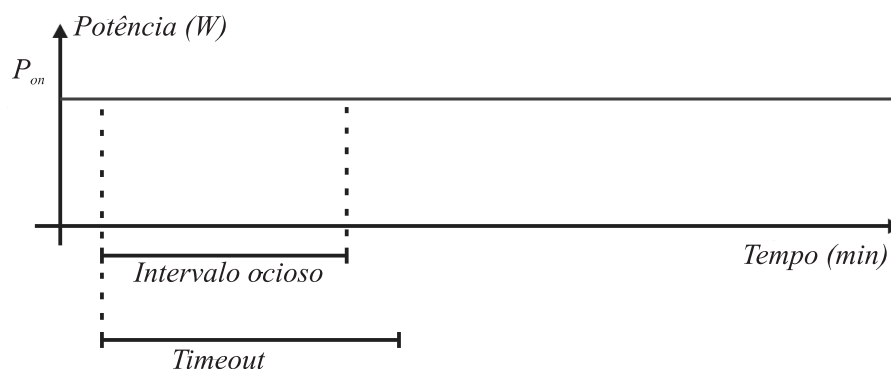


a tela ou o tempo para suspender todo o sistema. Uma política de *timeout* fixo é uma estratégia simples de gerenciamento de energia [21]. Quando um dispositivo permanece ocioso por um tempo maior que *timeout*, a política coloca o dispositivo em um estado de menor consumo de energia e o dispositivo volta à atividade quando ocorre uma requisição da carga de trabalho. Uma política de *timeout* dinâmico modifica o valor do *timeout* de forma a otimizar o seu desempenho. Existem evidências que mesmo uma política simples de *timeout* dinâmico [21] pode apresentar um melhor desempenho que uma política de *timeout* fixo. A política de *timeout* [22] é a mais amplamente utilizada em muitas aplicações, tais como microprocessadores, monitores e discos rígidos, por causa de sua simplicidade. O valor do *timeout* pode ser fixado (*timeout* estático) ou pode ser alterado ao longo do tempo.

Para o caso de interfaces de rede, o objetivo de uma política de gerenciamento de energia com base em *timeout* é minimizar tanto o consumo de energia quanto a penalidade de desempenho causada quando a interface de rede está desligada mas o usuário tem interesse em utilizá-la. Nesse contexto, Luiz et al. [23] propuseram uma política de *timeout* fixo para interfaces de rede, considerando a taxa total de transferência de dados (*download* e *upload*) e um *timeout*.

A potência média durante um intervalo ocioso  $t_i$  é função do *timeout* utilizado. Se  $t_i$  é menor que o *timeout*, como apresentado na Figura 2.4, o gerenciador de energia não enviará o comando para desabilitar a interface de rede. Dessa forma, a potência média consumida durante o intervalo ocioso será igual à potência correspondente ao estado ativo,  $P_{on}$ .

Figura 2.4: Ilustração do valor da potência quando o intervalo ocioso é menor que o *timeout* selecionado

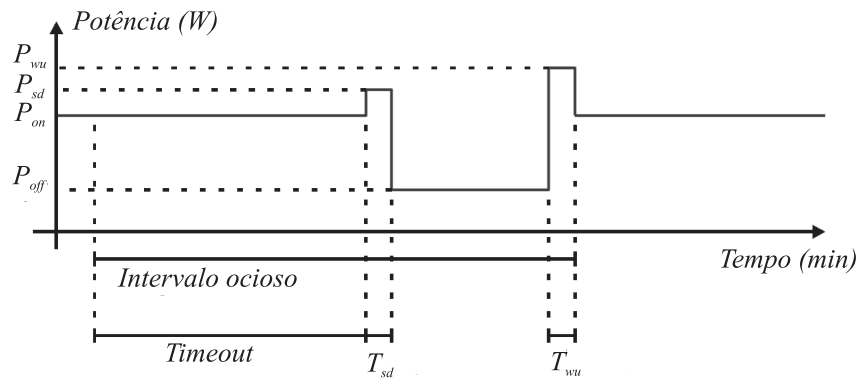


Fonte: Adaptado de Luiz et al. [23].

Se o intervalo  $t_i$  é maior ou igual ao valor do *timeout*,  $\delta$ , selecionado, como apresentado na Figura 2.5, o gerenciador de energia envia o comando para colocar o sistema num estado de baixo consumo de energia após o *timeout*. A constante  $T_s$  representa a soma dos intervalos de tempo que a interface leva para desligar,  $P_{wu}$ , e religar,  $P_{sd}$ . A energia consumida no processo de desligar e religar o dispositivo é representada por  $E_s$ , ou seja  $(E_{sd} + E_{wu})$ . A potência durante o intervalo ocioso será:

- a potência da interface ligada durante o *timeout*  $P_{on}$
- a potência para desligar a interface  $P_{sd}$  no tempo  $T_{sd}$
- a potência durante o tempo em que o dispositivo permanece em estado de baixo consumo  $P_{off}$
- a potência para religar a interface  $P_{wu}$  durante o intervalo de tempo  $T_{wu}$ .

Figura 2.5: Ilustração do valor da potência quando o intervalo ocioso é maior ou igual ao *timeout* selecionado.



Fonte: Adaptado de Luiz *et al.* [23].

Portanto, considerando os casos apresentados nas Figuras 2.4 e 2.5 a potência média durante um intervalo ocioso é dada em (2.2) e é válida para  $T_s$  menor que o período de amostragem.

$$p_{med}(\delta, t_i) = \begin{cases} P_{on}, & \text{se } t_i < \delta \\ \frac{P_{on} \cdot \delta + P_{off} \cdot (t_i - \delta - T_s) + E_s}{t_i}, & \text{se } t_i > \delta \end{cases} \quad (2.2)$$

A penalidade no desempenho também pode ser considerada. Se o usuário desejar utilizar a interface de rede e ela estiver desabilitada, o mesmo terá de esperar um tempo de transição  $T_{off/on}$  para poder utilizá-la. A probabilidade desse evento ocorrer é igual a  $Pr[T_i \geq \delta]$ .

Vilar et al. [11] apresentaram uma comparação entre políticas de *timeout* dinâmico e *timeout* estático para interfaces de rede. O trabalho mostrou que a política de *timeout* dinâmico superava a de *timeout* fixo tanto na economia de energia quanto na penalidade de desempenho.

### 2.2.2 Políticas preditivas

Numa política preditiva, um intervalo ocioso futuro é estimado com base nos intervalos ociosos anteriores. O objetivo disso é que não seja necessário aguardar o *timeout* para desligar o dispositivo. Assim, em políticas preditivas, o período de espera pelo *timeout* é eliminado. Se é previsto que o dispositivo ficará ocioso por um tempo maior que o  $T_{be}$ , o dispositivo será desligado ao iniciar seu intervalo de inatividade [24].

A questão crítica em políticas preditivas é a precisão nas estimativas dos intervalos ociosos. É com base nessa previsão que o GE atua, desligando ou não o dispositivo. No entanto, em sistemas onde a carga de trabalho não é conhecida a priori, como é o caso de sistemas computacionais de propósito geral, onde se tem pouco conhecimento a respeito das requisições futuras na carga de trabalho, as políticas preditivas baseiam-se na premissa que o histórico de intervalos ociosos passados irá se repetir em um futuro próximo, devido ao comportamento repetitivo de algumas aplicações [18].

Lu et al. [25] propõe uma política que é uma combinação de *timeout* adaptativo e sistemas de previsão. Nessa política, as requisições que ocorrem no disco rígido são divididas em sessões. Um algoritmo prevê, a partir do comprimento das sessões anteriores, quando ocorrerá uma nova requisição. No período entre requisições, não há atividade no disco, e é esse o momento ideal para alterar o estado de energia do disco, economizando energia e minimizando penalidades.

Outra política preditiva utilizada baseia-se na duração dos intervalos ociosos passados para prever a duração dos intervalos ociosos futuros, chamada de Árvore de Aprendizado Adaptativo. Essa política pode ser aplicada em sistemas ou componentes com um número

arbitrário de estados de baixo consumo de energia, e com ela é possível prever qual o estado mais indicado para um dispositivo que inicia um período de inatividade. Segundo Chung et al. [26], essa técnica de gerenciamento dinâmico de energia foi usada em discos rígidos, e os resultados mostraram que a mesma é consistente e supera em termos de potência consumida e penalidade de desempenho outras políticas preditivas e de *timeout*.

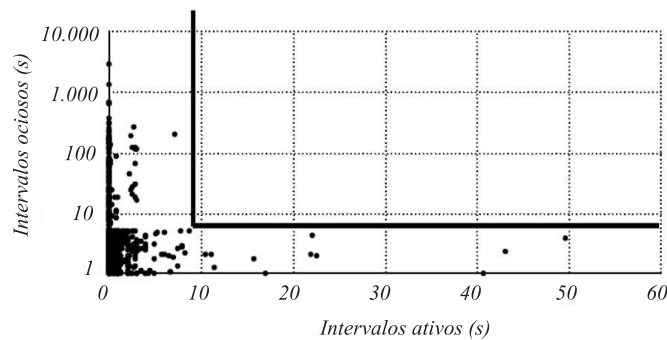
Hwang e Wu [27] utilizam uma soma ponderada do último intervalo ocioso e da última estimativa para prever a duração do próximo intervalo ocioso. A fórmula para a estimativa recursiva é:

$$I_{n+1} = a \cdot i_n + (1 - a) \cdot I_n \quad (2.3)$$

Em (2.3),  $I_{n+1}$  é o novo valor previsto,  $I_n$  é o último valor previsto,  $i_n$  é o último intervalo ocioso medido, e  $a \in (0, 1)$  é uma constante. Esse esquema propõe religamento preditivo para evitar atrasos de desempenho. Utilizando (2.3) é possível prever o próximo intervalo ocioso, que é função do último intervalo ocioso e do valor previsto passado. O parâmetro  $a$  controla o peso relativo entre o histórico recente e passado na previsão. Se  $a = 0$ , então  $I_{n+1} = I_n$ , ou seja, a história recente não tem nenhum efeito sobre o valor previsto. Por outro lado, se  $a = 1$ , então  $I_{n+1} = i_n$ , e a previsão considera apenas o intervalo ocioso mais recente.

Segundo Lu, Benini e Micheli [5], existe uma relação entre os intervalos ociosos e ativos de uma carga de trabalho. Essa relação entre os intervalos forma uma curva que possui a forma de  $L$ , como apresentado na Figura 2.6.

Figura 2.6: Ilustração de intervalos ociosos versus intervalos ativos, denominada curva  $L$ .



Fonte: Adaptado de Lu e De Micheli [5].

De acordo com Srivastava [28], se é verificado que para uma dada carga de trabalho a curva  $L$  é válida, é possível aplicar uma técnica de gerenciamento de energia baseada em um *threshold*. Se o período ativo é menor que o *threshold* considera-se que o intervalo ocioso subsequente será maior que o  $T_{be}$  e dessa forma o sistema pode ser desligado. A razão dessa política é que curtos períodos ativos são frequentemente seguidos por longos períodos ociosos. Claramente, a escolha do *threshold* é difícil. Para isso, é necessária a análise cuidadosa da curva de intervalos ociosos *versus* intervalos ativos para que se possa chegar a um valor correto. Assim, a escolha de um *threshold* é um método *offline*, ou seja, baseado na extensiva coleta e análise de dados. Além disso, tal método não é aplicável a uma carga de trabalho cuja curva dos intervalos ativos e ociosos não possua a forma de um  $L$ .

### 2.2.3 Políticas estocásticas

A função do GE é reduzir o consumo de energia, selecionando componentes que estão ociosos ou sub-utilizados para serem desligados ou entrarem em um estado de baixo consumo. Em tempo de execução, o GE observa as requisições do usuário, o estado de consumo do provedor de serviços, e o tamanho da fila. Se todas as requisições foram atendidas, o GE pode escolher alterar o estado de energia do dispositivo. Essa escolha é tomada baseada na política utilizada. Em uma política estocástica, cada componente do sistema é modelado probabilisticamente. A partir dos modelos de cadeia de Markov para o provedor, a fila, e o requisitante de serviços, é possível formular um problema de otimização de política que consiste na escolha da política que minimiza o consumo de energia do provedor de serviços para uma dada restrição de penalidade de desempenho.

Modelos estocásticos são utilizados para obter políticas ótimas para o gerenciamento dinâmico de energia. Benini et al. [12] apresentaram uma técnica que consiste de um modelo estocástico baseado em processos de decisão de Markov para a formulação da política de energia. É obtida uma solução exata do problema de otimização da política de energia em tempo polinomial resolvendo-se um problema de otimização linear. Entretanto, a solução requer que todas as transições de estados obedeçam a distribuições geométricas estacionárias, o que não é verdade em muitos casos práticos [29]. Ren et al. [13] particularizaram

o problema de identificação da carga de trabalho, propondo um modelo em que o gerenciador de energia solicita do sistema operacional a informação sobre qual o processo que gera as requisições, cuja carga de trabalho é conhecida. Cada tipo de carga de trabalho é denominada de modo do sistema. Essa técnica é restrita a sistemas em que o gerenciador de energia pode solicitar do sistema o modo da carga de trabalho. Chung et al. [7] generalizam a solução para o ambiente não-estacionário, utilizando janelas de comprimento fixo para estimar cargas de trabalho não-estacionárias. Uma limitação da abordagem no tempo discreto é o consumo extra de energia devido ao fato da decisão do GE ser reavaliada em cada período de amostragem, mesmo quando o sistema está em um intervalo ocioso [30]. A vantagem é que, por reavaliar periodicamente as decisões, elas podem ser reconsideradas, se adaptando melhor ao comportamento da carga de trabalho.

Uma alternativa à formulação baseada em decisões de Markov no tempo discreto são os modelos baseados em decisões de Markov no tempo contínuo [31]. Nesse esquema, o GE emite comandos uma vez que eventos ocorram. Como resultado, uma maior quantidade de energia pode ser economizada pois não há a necessidade de reavaliar a política durante os intervalos inativos. Os resultados são ótimos para os casos em que uma distribuição exponencial descreve bem os tempos de transição [30, 32]. Entretanto, na maioria dos casos práticos, os tempos de transição podem seguir uma outra distribuição [33], e por consequência, apresentar resultados muito distantes do caso ótimo.

Todas as técnicas de modelagem e otimização anteriormente mencionadas assumem que o gerenciador de energia possui a informação perfeita do estado atual do sistema. Com base nessas informações, o gerenciador de energia encontra a melhor ação de gerenciamento de energia a partir de uma tabela pré-calculada armazenada na memória.

## 2.3 Métricas de penalidade

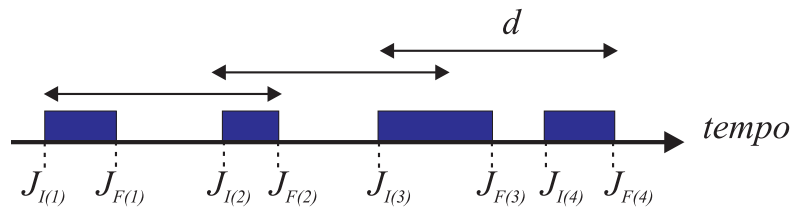
No GDE procura-se uma solução de compromisso entre economia de energia e penalidade de desempenho. Muitas abordagens já foram utilizadas na tentativa de quantificar a penalidade de desempenho, como medir o atraso médio ou total, por exemplo.

Entretanto, essas métricas podem ser enganadoras. Por exemplo, um sistema que causa um total de 50 segundos de atraso em cinco horas de observação é melhor que um que causa

60 segundos de atraso nas mesmas cinco horas. No entanto, se o primeiro exige que um usuário aguarde 40 segundos seguidos, enquanto o último requer pequenos atrasos de no máximo 10 segundos, a maioria dos usuários pode pensar que o segundo sistema possui melhor desempenho. O tempo de atrasos total não reflete essa discrepância [34]. Lu et al. [34] usam duas métricas de desempenho objetivas: atrasos longos e atrasos consecutivos (dentro de curtos intervalos de tempo) para refletir a percepção de degradação de desempenho.

Na Figura 2.7 está representado um intervalo de tempo onde ocorreram quatro atrasos. Um atraso ( $J$ ) inicia em  $J_{I(n)}$  e termina em  $J_{F(n)}$ , onde  $n = 1, 2, 3, \dots, n_d$ . Uma métrica para o desempenho pode ser considerar o maior atraso dentro uma janela de comprimento  $d$ , chamado  $J_d$  [34]. Esse valor pode ser obtido através da soma de todos os atrasos de tempo dentro das janelas de comprimento  $d$ . Dependendo do tamanho da janela, nela podem caber vários atrasos, como os atrasos  $J_3$  e  $J_4$  caberiam na terceira janela  $d$ , como também pode conter apenas partes de atrasos como  $J_2$  e  $J_3$  na segunda janela. Utilizando (2.4) é possível calcular o valor de  $J_d$ .

Figura 2.7: Representação dos períodos de atrasos devido ao uso de uma técnica de GDE.



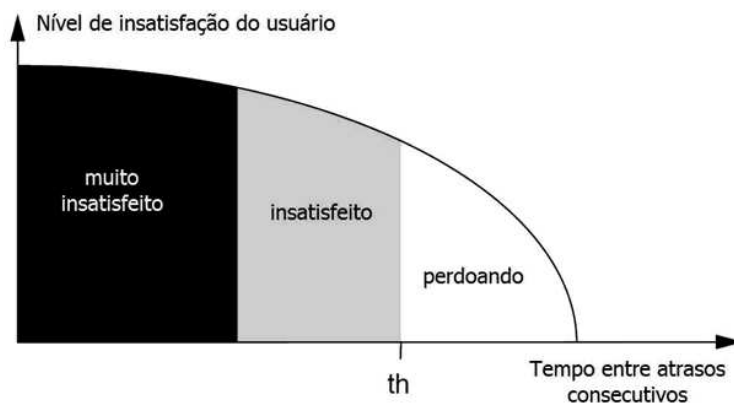
Fonte: Adaptado de Lu *et al.* [34].

$$J_d = \sum_{n=1}^{n_d} (J_{F(n)} - J_{I(n)}) \quad (2.4)$$

A segunda métrica de desempenho proposta por Lu et al. [34] é medir o número de atrasos consecutivos que estão próximos e causam penalidades repetitivamente. Muitos usuários compreendem que alguns atrasos são inevitáveis, e quando ocorrem com baixa frequência podem ser relevados ou tolerados. No entanto, se os atrasos ocorrem frequentemente, a insatisfação do usuário cresce drasticamente. Lu e De Micheli [5] chamam esse fenômeno

de “efeito de memória curta”, pois usuários esquecem atrasos que ocorreram a muito tempo atrás. No entanto, se o intervalo de tempo entre os atrasos é demasiadamente curto, o usuário irá lembrar de atrasos consecutivos em um pequeno intervalo de tempo e seu sentimento de insatisfação tenderá a crescer. Conseqüentemente, se o tempo entre os atrasos é demasiadamente grande, o usuário tende a aceitar melhor o próximo atraso. Essa relação entre o intervalo de tempo entre os atrasos e a insatisfação do usuário é ilustrada na Figura 2.8.

Figura 2.8: Ilustração da insatisfação do usuário.



Fonte: Adaptado de Lu e De Micheli [5].

Em um determinado problema de gerenciamento de energia, é interessante portanto, buscar um tempo  $th$  entre os atrasos, de forma que o usuário possa “perdoar” a ocorrência de penalidades consecutivas. No gerenciamento dinâmico de energia o objetivo é minimizar as penalidades de desempenho e o consumo de energia. Esse objetivo pode ser alcançado com um gerenciador de energia ideal, que contemple o conhecimento do presente, passado e futuro da carga de trabalho [7]. No entanto, em sistemas computacionais de propósito geral, o comportamento da carga de trabalho é desconhecido a priori. Então, é necessário estimar o comportamento da carga de trabalho, para que se possa tomar decisões de gerenciamento.



## 2.4 Considerações finais

O Gerenciamento Dinâmico de Energia desempenha um papel importante para reduzir o consumo de energia de maneira eficaz tanto na fase de projeto quanto na operacional dos sistemas computacionais, a partir da alteração do estado de consumo de dispositivos que estão ociosos ou sub-utilizados e tornou-se uma técnica popular para redução do consumo de energia no nível do sistema. A eficácia do gerenciamento de energia depende de quão preciso é o modelo da carga de trabalho e das técnicas de otimização de políticas [35, 36]. No trabalho de Hwang e Wu [27], a previsão dos intervalos ociosos é feita utilizando uma abordagem de média exponencial. No trabalho apresentado por Srivastava [28], uma função de regressão é usada para prever a próxima requisição da carga de trabalho. Entretanto, a natureza da carga de trabalho de um sistema computacional é complexa e incerta porque é determinada pelo contexto do usuário e pela interação *hardware/software*. Os métodos estocásticos são naturalmente selecionados pelas ferramentas já existentes para modelagem e otimização das políticas [35].

No capítulo seguinte apresenta-se inicialmente a abordagem de Benini et al. [12] baseada em modelos estocásticos e adequada apenas para cargas de trabalho estacionárias. Em seguida, são apresentadas as abordagens que lidam com cargas de trabalho desconhecidas e não-estacionárias [7, 14].

# Capítulo 3

## Gerenciamento dinâmico de energia

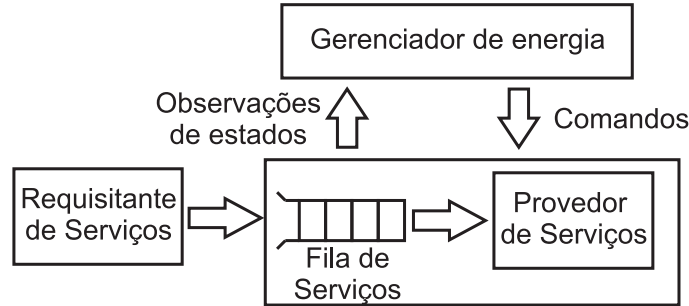
Nesse capítulo será apresentado um modelo de sistema com gerenciamento de energia, onde os componentes são modelados como cadeias de Markov estacionárias e discretas no tempo. Essa abordagem é baseada em modelos estocásticos e é adequada apenas para cargas de trabalho estacionárias. Em seguida são apresentadas abordagens que lidam com cargas de trabalho não-estacionárias utilizando janelas deslizantes de múltiplos comprimentos [14, 16, 37].

### 3.1 Gerenciamento dinâmico de energia para cargas de trabalho estacionárias

Um sistema com gerenciamento de energia é modelado por Benini et al. [12] como um *provedor de serviços* (PS), um *requisitante de serviços* (RS), uma *fila de serviços* (FS), e um *gerenciador de energia* (GE), como apresentado na Figura 3.1. O provedor de serviços processa as requisições de um único *requisitante de serviços* (RS), e possui diversos estados de operação. Cada estado é caracterizado por um nível de consumo de energia e uma taxa de serviço, que é proporcional à quantidade de requisições atendidas em uma unidade de tempo. Alguns estados podem ter taxa de serviço nula, por exemplo: desligado, ocioso. Enquanto estados com taxa de serviço não nula são chamados de *estados ativos*.

O *gerenciador de energia* (GE) é a unidade que controla as transições de estados de energia do provedor de serviços. É possível que haja uma *fila de serviços* (FS) que arma-

Figura 3.1: Componentes do modelo de um sistema com gerenciamento de energia.



Fonte: Adaptado de Benini *et al.* [12].

zena requisições que ainda não puderam ser atendidas pelo provedor de serviços, porque o provedor de serviços estava ocupado atendendo uma outra requisição, ou porque estava com taxa de serviço nula.

### 3.1.1 Componentes do sistema

Como apresentado por Benini *et al.* [12], os componentes do sistema com gerenciamento de energia, apresentados na Figura 3.1, podem ser modelados como cadeias de Markov estacionárias e discretas no tempo. O gerenciador de energia toma decisões com respeito ao provedor de serviços com base na história do provedor de serviços, do requisitante de serviços e da fila de serviços. O tempo é considerado discreto:  $t = 0, 1, 2, 3, \dots$ , e o intervalo  $h$  entre dois instantes consecutivos de tempo é denominado *período de amostragem*.

**Definição 1.** *Um requisitante de serviços (RS) é um par  $(\mathcal{M}_{RS}, z(r))$  onde:*

- $\mathcal{M}_{RS}$  é uma cadeia de Markov estacionária com conjunto de estados  $\mathcal{R} = \{r_i | i = 0, 1, \dots, (R-1)\}$  e matriz de transição  $\mathbf{P}^{RS}$ ;
- $z(r)$  é uma função  $z : \mathcal{R} \rightarrow \mathbb{N}$ .

A função  $z(r)$  representa o número de requisições emitidas por intervalo de tempo pelo requisitante de serviços quando no estado  $r$ . O processo de Markov para a geração de requisições é completamente autônomo e não depende do comportamento do sistema: ele

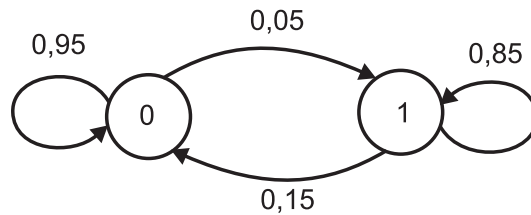
representa o ambiente externo sobre o qual o sistema não tem controle. Os intervalos de tempos entre as requisições seguem uma distribuição geométrica.

**Exemplo 1.** Considere um RS com dois estados  $r_0$  e  $r_1$ , onde  $z(r)$  é definido como  $z(r_0) = 0$ ,  $z(r_1) = 1$ . Em qualquer tempo  $t$ , apenas duas possibilidades existem: nenhuma requisição é recebida ou uma requisição é recebida. Um exemplo de matriz de transição para o RS é apresentado em (3.1)

$$P^{RS} = \begin{matrix} & \begin{matrix} r_0 & r_1 \end{matrix} \\ \begin{matrix} r_0 \\ r_1 \end{matrix} & \begin{bmatrix} 0,95 & 0,05 \\ 0,15 & 0,85 \end{bmatrix} \end{matrix} \quad (3.1)$$

A cadeia de Markov para esse RS é apresentada na Figura 3.2. O RS modela uma carga de trabalho de uso intensivo. Existe uma alta probabilidade (0,85) de receber uma requisição no intervalo de tempo  $n + 1$  se uma requisição foi recebida no intervalo de tempo  $n$ .

Figura 3.2: Modelo da cadeia de Markov para o requisitante de serviços.



Fonte: Adaptado de Benini *et al.* [12].

**Definição 2.** Um provedor de serviços (PS) é descrito por uma tripla  $(\mathcal{M}_{PS}(a), b(s, a), c(s, a))$  onde:

- $\mathcal{M}_{PS}(a)$  é uma cadeia de Markov estacionária controlável com conjunto de estados  $\mathcal{S} = \{s_i | i = 1, 2, \dots, S\}$ , conjunto de comandos  $\mathcal{A} = \{a_i | i = 1, 2, \dots, A\}$  e matriz de probabilidades de transição  $\mathbf{P}^{PS}(a)$ ;
- $b(s, a)$  é uma função  $b : \mathcal{S} \times \mathcal{A} \rightarrow [0, 1]$ ;
- $c(s, a)$  é uma função  $c : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ .

A taxa de serviço  $b(s, a)$  está associada com cada estado  $s \in \mathcal{S}$  e comando  $a \in \mathcal{A}$ , e representa a probabilidade do PS completar o atendimento de uma requisição de serviço num intervalo de tempo, dado que o mesmo encontra-se no estado  $s$  e que o comando  $a$  foi emitido no início do intervalo de tempo. Cada par  $(s, a)$  é caracterizado por uma taxa de serviço e uma consumo de potência  $c(s, a)$ . O valor de  $b(s, a)$  satisfaz  $0 \leq b \leq 1$ . Taxa de serviço  $b(s, a) = 0$  significa que nenhuma requisição pode ser atendida e que o PS não está ativo. Taxa de serviço  $b(s, a) = 1$  significa que as requisições serão prontamente atendidas em cada intervalo de tempo. A função  $c$  representa uma métrica de consumo em unidades arbitrárias, tal como  $W$ .

**Exemplo 2.** Considere um PS com dois estados  $\mathcal{S} = \{on, off\}$ . Os comandos são  $\mathcal{A} = \{s_{on}, s_{off}\}$ , que significam ligar e desligar, respectivamente. Quando o comando é emitido, o PS irá para o novo estado com probabilidade dependente apenas do comando  $a$ , do estado do qual ele irá partir e para o qual ele irá. A matriz de transição  $P^{PS}(a)$  para esse PS, pode ser representada por duas matrizes, uma para cada comando, como apresentada em (3.2) e (3.3).

$$P^{PS}(s_{on}) = \begin{array}{cc} & \begin{array}{cc} on & off \end{array} \\ \begin{array}{c} on \\ off \end{array} & \begin{bmatrix} 1 & 0 \\ 0,1 & 0,9 \end{bmatrix} \end{array} \quad (3.2)$$

$$P^{PS}(s_{off}) = \begin{array}{cc} & \begin{array}{cc} on & off \end{array} \\ \begin{array}{c} on \\ off \end{array} & \begin{bmatrix} 0,8 & 0,2 \\ 0 & 1 \end{bmatrix} \end{array} \quad (3.3)$$

Por exemplo, se o PS estiver no estado *on* quando for emitido um comando  $s_{on}$ , o PS permanecerá no estado *on* com probabilidade 1. Outro exemplo, se o PS estiver no estado *on* quando for emitido o comando  $s_{off}$ , o PS transitará para o estado *off* com probabilidade 0,2. Ou seja, se esse comando permanecer nos próximos períodos o PS transitará para o estado  $s_{off}$  em um tempo médio de  $1/0,2 = 5$  períodos de amostragem.

A taxa de serviço  $b(s, a)$  e a potência consumida  $c(s, a)$  podem ser representadas por duas matrizes com uma entrada para cada par de estado-comando.

$$b(s, a) = \begin{array}{c} \\ \\ \end{array} \begin{array}{cc} s_{on} & s_{off} \\ \begin{bmatrix} 0,8 & 0 \\ 0 & 0 \end{bmatrix} \end{array} \quad (3.4)$$

$$c(s, a) = \begin{array}{c} \\ \\ \end{array} \begin{array}{cc} s_{on} & s_{off} \\ \begin{bmatrix} 3 & 4 \\ 4 & 0 \end{bmatrix} \end{array} \quad (3.5)$$

O consumo de energia é nulo no estado desligado, mas ligar ou desligar tem um custo de energia considerável: o consumo de energia do PS durante os tempos de comutação é maior que o do estado ligado.

**Definição 3.** Uma fila de serviço (FS) é descrita como uma cadeia de Markov estacionária controlável  $\mathcal{M}_{FS}(a, s, r)$  com conjunto de estados  $\mathcal{Q} = \{q_i | i = 0, 1, \dots, (Q - 1)\}$ , conjunto de comandos  $\Theta = \mathcal{A} \times \mathcal{S} \times \mathcal{R}$  e matriz de transição  $\mathbf{P}^{FS}(a, s, r)$ .

Quando as requisições chegam durante um intervalo de tempo, elas são armazenadas em uma fila de tamanho  $(Q - 1)$ . A fila está no estado  $q_i$  quando  $i$  requisições estão esperando para serem atendidas. Se uma nova requisição chega quando a fila já está na posição  $q_{Q-1}$ , a requisição é perdida. O estado  $q_{Q-1}$  significa que a lista está completamente cheia e  $q_0$  lista vazia. O PS controla quão rápido a lista é esvaziada, e o RS controla quão rápido a lista será preenchida. Dada a tripla  $(a, s, r)$  sabe-se que  $b(s, a)$  é a taxa de serviço para o número de requisições emitidas no estado  $r$ ,  $z(r)$ . A probabilidade de uma requisição ser atendida é  $b(s, a)$ , enquanto que a probabilidade de nenhuma requisição da lista ser atendida é  $1 - b(s, a)$ .

Se  $q = q_0$  e  $z(r) = 0$ , a lista está vazia e nenhuma requisição chegou, então a lista permanece no estado  $q_0$  com probabilidade 1. Se a lista está cheia, os estados irão mudar com probabilidade  $b(s, a)$  se  $z(r) = 0$ , caso contrário, em que  $z(r) > 0$ , a lista permanece no estado  $q_{Q-1}$  com probabilidade 1. Se o número de requisições emitidas excede o comprimento da fila, i.e.,  $i + z(r) \geq Q - 1$ , para uma lista que encontra-se no estado  $q_i$ , então o novo estado será  $q_{Q-1}$  com probabilidade 1. As demais probabilidades de transição são apresentadas em (3.6)

$$p_{q_i, q_j}^{FS}(a, s, r) = \begin{cases} 1 - b(s, a), & \text{se } j = i + z(r) \text{ e } 0 < i + z(r) < Q \\ b(s, a), & \text{se } j = i + z(r) - 1 \text{ e } 0 \leq i + z(r) < Q + 1 \\ 0, & \text{caso contrário} \end{cases} \quad (3.6)$$

**Exemplo 3.** Considere o modelo do PS e do RS introduzidos nos exemplos anteriores, assumindo que o comprimento máximo da fila é 1,  $\mathcal{Q} = \{q_0, q_1\}$ . A matriz  $\mathbf{P}^{FS}(a, s, r)$  pode ser descrita por oito matrizes 2x2, uma para cada valor da tripla  $(a, s, r)$ .

$$P^{FS}(s_{on}, on, 0) = \begin{matrix} & 0 & 1 \\ 0 & \begin{bmatrix} 1 & 0 \\ 0,8 & 0,2 \end{bmatrix} \\ 1 & \end{matrix} \quad (3.7)$$

$$P^{FS}(s_{on}, on, 1) = \begin{matrix} & 0 & 1 \\ 1 & \begin{bmatrix} 0,8 & 0,2 \\ 0 & 1 \end{bmatrix} \\ 0 & \end{matrix} \quad (3.8)$$

$$P^{FS}(s_{on}, off, 0) = \begin{matrix} & 0 & 1 \\ 0 & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ 1 & \end{matrix} \quad (3.9)$$

$$P^{FS}(s_{on}, off, 1) = \begin{matrix} & 0 & 1 \\ 0 & \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \\ 1 & \end{matrix} \quad (3.10)$$

$$P^{FS}(s_{off}, on, 0) = \begin{matrix} & 0 & 1 \\ 0 & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ 1 & \end{matrix} \quad (3.11)$$

$$P^{FS}(s_{off}, on, 1) = \begin{matrix} & 0 & 1 \\ 0 & \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \\ 1 & \end{matrix} \quad (3.12)$$

$$P^{FS}(s_{off}, off, 0) = \begin{matrix} & 0 & 1 \\ 0 & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ 1 & \end{matrix} \quad (3.13)$$

$$P^{FS}(s_{off}, off, 1) = \begin{matrix} & 0 & 1 \\ 0 & \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \\ 1 & \end{matrix} \quad (3.14)$$

Pode-se observar que o PS está atendendo as requisições com taxa 0,8 apenas quando encontra-se no estado  $on$  e o comando emitido é  $s_{on}$ . Em todas as outras combinações, a taxa de serviço é zero. Se chegarem requisições, essas só poderão ser armazenadas. Entretanto, se a fila já estiver cheia, a requisição será perdida.

**Definição 4.** O gerenciador de energia (GE) é o procedimento de controle que emite um comando  $a \in \mathcal{A}$  no tempo  $t$  de acordo com o estado do sistema  $x = (s, r, q)$ , que é a composição dos estados do provedor de serviços, requisitante de serviços e fila de serviços, respectivamente. Então, o sistema tem o conjunto de estados  $\mathcal{X} = \mathcal{S} \times \mathcal{R} \times \mathcal{Q}$  com  $X = S \cdot R \cdot Q$  estados.

**Exemplo 4.** Nos exemplos anteriores foi assumido que  $\mathcal{A} = \{s_{on}, s_{off}\}$ . Um gerenciador de energia pode em qualquer instante de tempo  $t$  emitir o comando  $s_{on}$  ou  $s_{off}$ , dependendo da política adotada. Um exemplo de uma política determinística é uma que toma decisões a partir da fila de serviços. Se a fila estiver vazia, é emitido o comando  $s_{off}$ , caso contrário é emitido o comando  $s_{on}$ . Um exemplo de política randomizada é uma que toma decisões a partir dos três últimos estados observados. Se nenhuma requisição chegou nos últimos três intervalos de tempo, o comando  $s_{off}$  é emitido com probabilidade 0,8.

O sistema pode ser visto como uma composição de cadeias de Markov para o provedor de serviços, o requisitante de serviços e a fila de serviços,  $x = (s, r, q)$ , respectivamente. A matriz de transição do sistema  $\mathbf{P}(a)$  é função do comando  $a$ , onde  $a$  é uma variável aleatória com distribuição de probabilidade discreta  $\delta^{(t)} = (p_{a_0}, p_{a_1}, \dots, p_{a_{A-1}})$ , denominada a *decisão* do gerenciador de energia, que associa um valor de probabilidade  $p_a$  para cada comando  $a \in \mathcal{A}$  no tempo  $t$ . A matriz de transição  $\mathbf{P}_{\delta^{(t)}}$  do sistema quando a decisão  $\delta^{(t)}$  é tomada,



é apresenta em (3.15):

$$\mathbf{P}_{\delta^{(t)}} = \sum_{p_a \in \delta^{(t)}} p_a \mathbf{P}(a) \quad (3.15)$$

De forma geral, as métricas relevantes no problema de otimização de uma política são função do estado  $x$  e do comando  $\delta_x$  tomado quando no estado  $x$ . A primeira métrica é o consumo esperado de potência  $c(x, \delta_x) = \sum_{p_a \in \delta_x^{(t)}} p_a c(s, a)$ , onde  $c(s, a)$  é o consumo de potência quando o provedor de serviços está no estado  $s$  e o comando  $a$  é emitido. A segunda métrica é a penalidade de desempenho  $d(x)$ , que pode ser definida como a quantidade de requisições pendentes na fila. Dessa forma, o consumo esperado e a penalidade de desempenho esperada para o provedor de serviços sob a política  $\pi$  são apresentados em (3.16) e (3.17), respectivamente, em que  $E$  é o operador esperança matemática,  $c_\delta$  é uma métrica de consumo e  $d_\delta$  é uma métrica de penalidade de desempenho.

$$E_\pi[\mathbf{c}_\delta(t)] = p(t)c_\delta \quad (3.16)$$

$$E_\pi[\mathbf{d}_\delta(t)] = p(t)d_\delta \quad (3.17)$$

O *problema de otimização de política (PO)*, expresso em (3.18), consiste na escolha da política  $\pi$  que minimiza o consumo esperado para uma dada restrição de penalidade de desempenho, onde  $D$  é a máxima penalidade de desempenho aceitável. A resolução do problema de otimização da política gera uma tabela de decisão, que é uma tabela com número de linhas igual ao número de estados do sistema e número de colunas igual ao de comandos passados pelo GE para o PS, e cada elemento é a probabilidade de o GE passar um dado comando para o estado atual do sistema.

$$\begin{aligned} \text{PO :} \quad & \min_{\pi} \sum_{t=0}^{\infty} E_{\pi} [\mathbf{c}_{\delta}(t)] \\ & \text{tal que } \sum_{t=0}^{\infty} E_{\pi} [\mathbf{d}_{\delta}(t)] \leq D \end{aligned} \quad (3.18)$$

No trabalho de Benini et al. [12], o PS e o RS são modelados como processos estacionários, e as probabilidades do RS são invariantes no tempo, o que corresponde a uma carga de trabalho estacionária. Com essas considerações, todo o modelo de sistema não inclui nenhum parâmetro variante no tempo. Portanto, o sistema completo pode ser descrito por uma cadeia de Markov estacionária controlada. A política de energia para tal sistema pode ser otimizada sob restrições de consumo de potência ou desempenho, resolvendo um problema de programação linear. Em situações práticas, não se possui conhecimento a priori da carga de trabalho e ela apresenta comportamento não-estacionário. Para superar essas limitações, o modelo apresentado por Benini et al. [12] é estendido por Chung et al. [7] para o ambiente não-estacionário.

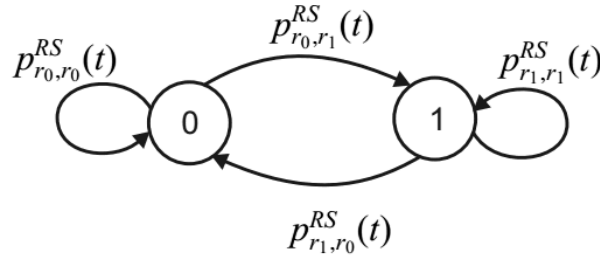
## 3.2 Gerenciamento dinâmico de energia para cargas de trabalho não estacionárias

A identificação do modelo da carga de trabalho não-estacionária é essencial para técnicas de *GDE* que dependem de modelos da carga de trabalho, tais como *timeout* [5], desligamento preditivo [27], árvore de aprendizado adaptativo [26] e abordagens de aprendizado da carga de trabalho baseadas em janelas deslizantes de comprimento fixo [7].

Em muitas situações práticas, a estacionaridade do RS não é verificada. Por exemplo: em um sistema computacional a carga de trabalho varia com o tempo e com a utilização. No trabalho de Chung et al. [7] uma carga de trabalho não-estacionária, como a ilustrada na Figura 3.3 é denotada por  $U^l$ , e modelada como uma série de cargas de trabalho estacionárias que possuem diferentes probabilidades de requisição do usuário. Cada carga de trabalho estacionária é denotada por  $u_s$ ,  $s = 0, 1, \dots, N_u - 1$ , onde  $N_u$  é o número total de cargas de

trabalho estacionárias que compõem a carga não-estacionária,  $U^l$ .

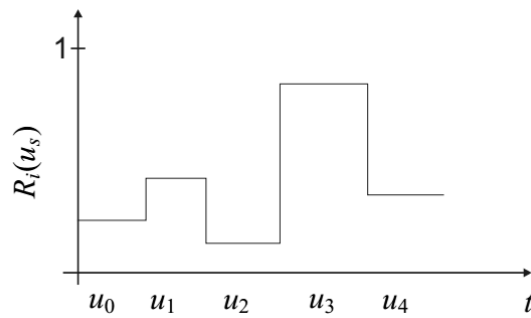
Figura 3.3: Exemplo de um RS não-estacionário.



Fonte: Adaptado de Luiz *et al.* [14].

Em um ambiente não-estacionário como o ilustrado na Figura 3.4, a política ótima é a que toma as decisões baseadas em uma tabela de decisão otimizada com o  $R_i$  para cada  $u_s$ , sendo  $R_i$  a probabilidade de um RS permanecer no estado  $r_i$ , i.e.,  $R_i = p_{r_i,r_i}^{RS}$ . Chung *et al.* [7] definiram essa política como *política adaptativa ótima* que não pode ser implementada em situações reais pois requer o perfeito conhecimento da carga de trabalho  $u_s$ . Por isso há necessidade da estimação das probabilidades de transição dos estados do requisitante de serviços em cargas de trabalho não-estacionárias.

Figura 3.4: Valor de  $R_i$  ao longo do tempo para uma carga de trabalho não-estacionária.



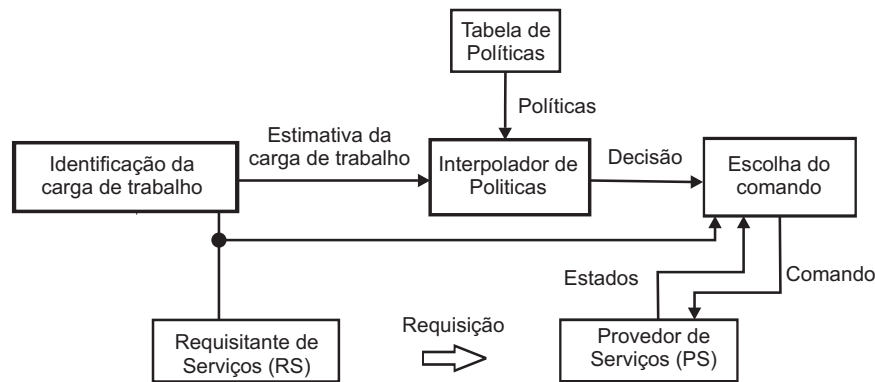
Fonte: Adaptado de Chung *et al.* [7].

### 3.2.1 Identificação da carga de trabalho

Na Figura 3.5 está representado um sistema com gerenciamento de energia para o ambiente não-estacionário, para o qual se pretende obter resultados próximos da política adaptativa ótima. No bloco *identificação da carga de trabalho* são observados os estados do requisitante

de serviços para se obter uma estimativa da carga de trabalho. Tal estimativa é obtida utilizando a técnica das janelas deslizantes de múltiplos comprimento (JMC), apresentada por Luiz et al. [14, 16]. A tabela de políticas é construída na fase de projeto com valores pré determinados da carga de trabalho. Como a estimativa pode ser diferente das cargas de trabalho utilizadas na construção da tabela de políticas, é necessário fazer uma interpolação. No bloco *interpolador de políticas*, é escolhido dinamicamente, dentro da tabela de políticas, a política mais adequada para a estimativa da carga de trabalho, e é tomada uma decisão. No bloco *escolha do comando*, são observados os estados do requisitante de serviços e, do provedor de serviços e a decisão para definir o comando que será passado para o provedor de serviços.

Figura 3.5: Diagrama de blocos para a técnica de gerenciamento dinâmico de energia proposta.



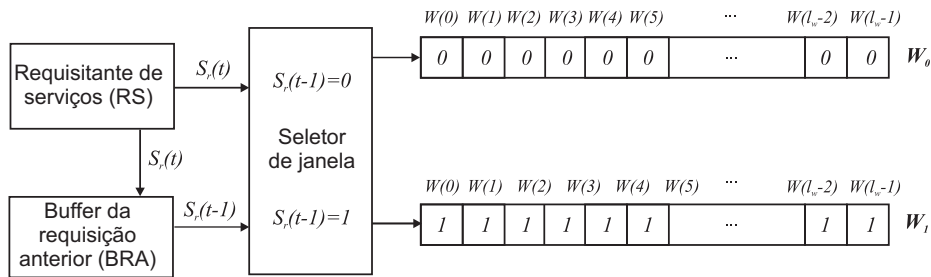
Fonte: Adaptado de Luiz [16].

### 3.2.2 Multi-Janelas

A técnica de estimação das probabilidades de transição do requisitante de serviços a partir de multi-janelas armazena o histórico das requisições da carga de trabalho a fim de prever as futuras requisições [7]. A estrutura básica do modelo de multi-janelas é apresentada na Figura 3.6. A quantidade de janelas corresponde à quantidade de estados do RS. A janela  $W_i$  é dedicada ao estado  $s_i$  do RS, e consiste de  $l_w$  compartimentos, onde cada compartimento  $W_i(j), j = 0, 1, \dots, l_w - 1$ , armazena transições anteriores a partir do estado  $i$  da carga de trabalho. A ideia básica para a operação das janelas é que, para qualquer intervalo de tempo  $t$ , o *Buffer da Requisição Anterior* (BRA) armazene  $s_r(t - 1)$  e escolha a janela  $W_i$ ,

caso  $s_r(t - 1) = i$ . A janela selecionada deve deslocar seu conteúdo de uma posição para a direita, ou seja,  $W_i(j + 1) \leftarrow W_i(j), i = 1, 2, \dots, l_w - 2$  e  $W_i(0)$  deve armazenar o estado mais recente da carga de trabalho  $s_r(t)$ .

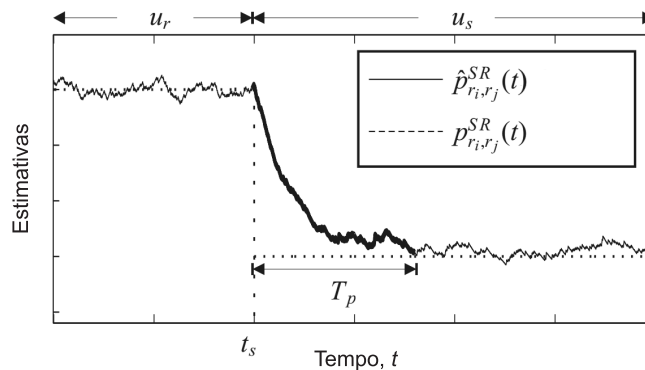
Figura 3.6: Exemplo de uso de multi-janela para RS de dois estados.



Fonte: Adaptado de Chung *et al.* [7].

As fontes de erros na estimação pelo método de multi-janelas são: o erro de amostragem e o tempo de adaptação. Na Figura 3.7 está ilustrado o atraso na identificação de uma carga de trabalho estacionária  $u_r$  para outra  $u_s$ . Após o atraso de identificação, as estimativas oscilam em torno dos valores de probabilidades de transição verdadeiros  $p_{r_i,r_j}^{RS}(t)$ . Chung *et al.* [7] denominaram esse efeito como *erro de amostragem*. Luiz *et al.* [14, 16] evidenciaram que quanto maior o número de amostras  $l_w$  para identificação de  $p_{r_i,r_j}^{RS}(t)$ , menor o erro de amostragem, pois se reduz a variância, entretanto o tempo inicial para obter a estimativa aumenta, acarretando num atraso de identificação tanto maior quanto maior o número de amostras.

Figura 3.7: O erro de amostragem (linha de maior espessura) após a transição, em  $t_s$ , de uma carga de trabalho estacionária  $u_r$  para outra  $u_s$ .



Fonte: Luiz [16].

Nesse trabalho, será utilizada a técnica *JMC*, apresentada por Luiz et al. [14, 16] que combina erro de amostragem pequeno para comprimentos de janela grandes, e pequenos atrasos de identificação para comprimentos de janela pequenos. Por meio da técnica proposta por esses autores, verificou-se também a possibilidade de adaptação quando a carga de trabalho de um sistema é estacionária.

### 3.3 Considerações finais

Uma abordagem de gerenciamento dinâmico de energia baseada em modelos estocásticos pode levar a um melhor desempenho e redução do consumo de energia do que uma técnica de gerenciamento de energia estática [3]. Com as ferramentas apresentadas nesse capítulo, é possível modelar um sistema como cadeias de Markov estacionárias e discretas no tempo. A técnica descrita na Seção 3.2 pode ser vista como uma extensão da abordagem descrita na Seção 3.1, com a consideração que as probabilidades de requisição do usuário variam com o tempo.

No próximo Capítulo será apresentada a técnica desenvolvida para o gerenciamento dinâmico de energia, que a partir da estimação da carga de trabalho não-estacionária, toma a decisão de desligar ou não o sistema, e qual o tempo ótimo que o mesmo deverá permanecer desligado para alcançar a maior economia de energia atendendo a uma dada restrição de penalidade. São formuladas as expressões analíticas para o consumo de energia e penalidade de desempenho, e então é formulado um problema de otimização com o compromisso de minimizar o consumo de energia respeitando restrições de penalidade de desempenho.

# Capítulo 4

## Política do tempo ótimo de desligamento

### 4.1 Introdução

Nesse capítulo é apresentada a construção da política de gerenciamento dinâmico de energia. A carga de trabalho é modelada como uma cadeia de Markov de dois estados. Será apresentado o desenvolvimento de modelos analíticos para o consumo de energia e para a penalidade de desempenho. A partir desses modelos é construído um problema de otimização com a intenção de minimizar o consumo de energia atendendo a níveis de penalidade de desempenho, considerando que existe um tempo ótimo de desligamento ( $T$ ) para cada modelo de carga de trabalho.

### 4.2 Política do tempo ótimo de desligamento

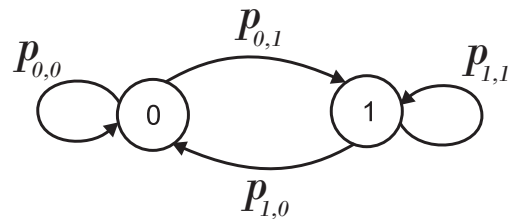
Nesse trabalho, a abordagem é voltada para um *requisitante de serviços* (RS) que possui apenas dois estados, aqui denominados *ativo* e *ocioso*. O RS pode ser representado por uma cadeia de Markov de dois estados  $\mathcal{R} = \{0, 1\}$  e uma matriz de transição  $\mathbf{P}^{RS}$ , como apresentado em (4.1). Nesse conjunto de estados, 0 e 1 representam os estados *ocioso* e *ativo*, respectivamente.

$$P^{RS} = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} \quad (4.1)$$

Os dispositivos que suportam alterações nos estados de energia serão aqui denominados

*provedores de serviços*. Para a apresentação da política proposta, será considerada inicialmente uma carga de trabalho estacionária. O estado do RS é representado por  $s_r(kh)$ , onde,  $k = 0, 1, 2, 3, \dots$  e  $h$  é o intervalo de amostragem. Cada elemento da matriz de transição é uma probabilidade de transição  $p_{x,y} = Prob[s_r(kh) = y | s_r(kh - 1) = x]$ . Um exemplo é apresentado na Figura 4.1.

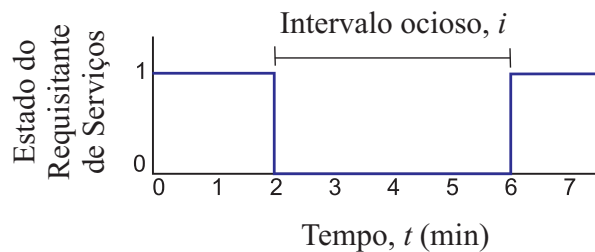
Figura 4.1: Exemplo de um modelo para um requisitante de serviços com dois estados, *ativo* e *ocioso*.



Fonte: Adaptado de Luiz *et al.* [38].

Os intervalos de tempo em que o RS está continuamente ocioso são denominados de *intervalos ociosos*, como apresentado na Figura 4.2. A duração de um intervalo ocioso será aqui representada por  $i$ .

Figura 4.2: Exemplo de um intervalo ocioso.



Fonte: Adaptado de Luiz *et al.* [38].

Considerando que os instantes de transição de estados em cadeias de Markov seguem uma distribuição de probabilidade geométrica [12], define-se a variável aleatória  $I$  como o número de repetições da transição  $(0 \rightarrow 0)$  necessárias para obter a primeira transição  $(0 \rightarrow 1)$ . Assim,  $I$  assume os valores possíveis  $1, 2, 3, \dots$ . Como  $I = i$  se, e somente se, as



primeiras  $(i - 1)$  repetições apresentarem a transição  $(0 \rightarrow 0)$ , enquanto a  $i$ -ésima apresentar a transição  $(0 \rightarrow 1)$ , tem-se:

$$Pr[I = i] = p_{0,0}^{i-1} \cdot p_{0,1}, \text{ para } i=1,2,3,\dots \quad (4.2)$$

A distribuição de probabilidade apresentada em (4.2) é denominada de distribuição geométrica [39]. Como  $p_{0,0} + p_{0,1} = 1 \Rightarrow p_{0,0} = 1 - p_{0,1}$ . A probabilidade de transição  $p_{0,1}$  será aqui denotada por  $q$  e a probabilidade de ocorrência de um intervalo ocioso  $i$  é apresentada em (4.3).

$$Pr[I = i] = (1 - q)^{i-1} \cdot q, \text{ para } i=1,2,3,\dots \quad (4.3)$$

Se  $I$  tiver uma distribuição geométrica como apresentada em (4.3), o valor esperado de  $I$  será dado em (4.4) [39].

$$E[I] = \frac{1}{q} \quad (4.4)$$

O PS aqui considerado tem apenas dois estados, denominados *on* e *off*. No estado *off*, o provedor de serviço encontra-se em baixo consumo e não é capaz de atender a nenhuma requisição. A potência média consumida é  $P_{off}$ . No estado *on*, o provedor de serviços pode atender às requisições, e a potência média consumida é  $P_{on}$ . O consumo de potência quando o provedor de serviços está *on* e atendendo às requisições do requisitante de serviços é maior ou igual a potência do mesmo quando no estado *on* e não há requisições para serem atendidas. Nesse trabalho não será abordada a situação em que o provedor de serviço está ativo e atendendo uma requisição, pois nesses casos não há interesse em alterar o estado de consumo do dispositivo. Dessa forma, a potência consumida no estado *on* é  $P_{on}$ , e assume-se que o provedor de serviço encontra-se ocioso esperando a chegada de uma requisição.

As transições de estado, como sair de um estado de baixo consumo para um estado de alto consumo, e vice-versa, são ações que demandam tempo e consumo extra de energia. A energia para desligar, ou seja, o consumo de energia para transitar do estado *on* para

o estado *off*, é  $E_{sd}$ , e ocorre durante um intervalo de tempo  $T_{sd}$ . A energia para religar, isto é, a energia para transitar do estado *off* para o estado *on*, é  $E_{wu}$ , e ocorre durante um intervalo de tempo  $T_{wu}$ . O consumo de energia durante o chaveamento é definido por  $E_s = E_{sd} + E_{wu}$ , e o tempo total do chaveamento é  $T_s = T_{sd} + T_{wu}$ .

### 4.2.1 Modelos de energia

Durante um intervalo ocioso, dependendo da política implementada, o GE poderá tomar a decisão de desligar ou não o provedor de serviços. Caso o desligamento ocorra, ele pode durar tempo suficiente para compensar a energia de transição ou não. E em algum momento o provedor de serviços será religado. O religamento poderá ser preditivo (se a política implementada for de religamento preditivo), ou realizado pelo usuário. Essas situações estão ilustradas na Figura 4.3<sup>1</sup>.

Os casos da Figura 4.3 (b) e (c) são situações típicas em políticas que atuam com desligamento e religamento preditivo. Já os casos apresentados na Figura 4.3 (d) e (e) são mais frequentes em políticas que atuam apenas com o desligamento preditivo e em políticas de *timeout*.

A seguir serão descritos os casos ilustrados na Figura 4.3. As expressões para o consumo de energia serão formulados apenas para os intervalos ociosos, pois a energia consumida nos intervalos ativos pode variar com a carga de trabalho. Por exemplo, para Wi-Fi a potência consumida em intervalos de tempo ativos é função da taxa de *download* e *upload* como descrita por Neves et al. [40].

#### Caso A

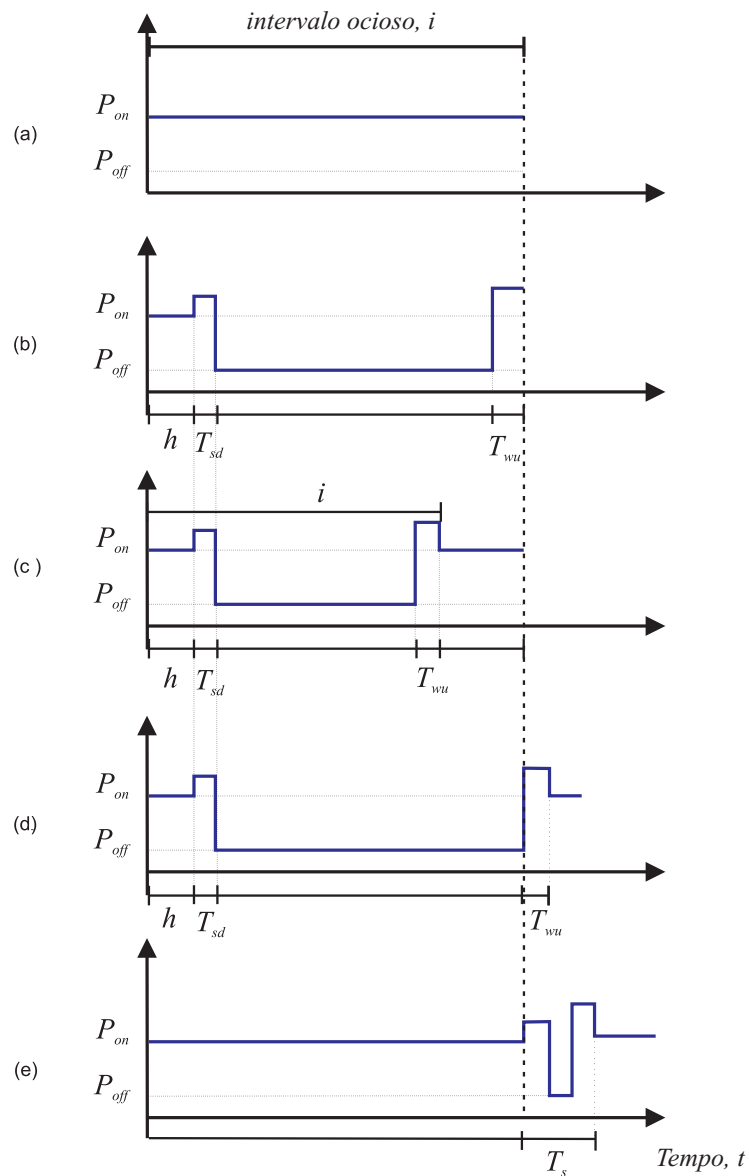
Nessa situação o provedor de serviços permanece ligado durante todo o intervalo ocioso. O consumo de energia para esse caso é dado em (4.5):

$$\mathcal{E} = P_{on}i \tag{4.5}$$

---

<sup>1</sup>As demais figuras que constam nesta tese e que não estejam referenciadas, foram elaboradas pela autora da mesma.

Figura 4.3: Estados de consumo de um provedor de serviços com gerenciamento de energia, onde  $h$  é o tempo mínimo necessário para identificação de um intervalo ocioso: (a) nenhum desligamento (b) religamento preditivo; (c) religamento preditivo (antecipado); (d) religamento realizado pelo usuário; (e) quando o desligamento ocorre no momento da transição do intervalo ocioso para o intervalo ativo.



Fonte: Elaborado pela Autora.

**Caso B**

No GDE essa situação poderia ser definida como ideal. No instante em que é identificado o início de um intervalo ocioso, o provedor de serviços é desligado. O religamento ocorre imediatamente antes da chegada de uma requisição da carga de trabalho, o que permite que o provedor de serviços esteja pronto para atender as requisições geradas, sem causar penalidade de desempenho. Esse caso é observado em políticas com religamento preditivo.

Para o *Caso B*, o consumo de energia é dado em (4.6):

$$\begin{aligned}\mathcal{E} &= P_{on}h + E_{sd} + P_{off}(i - h - T_{sd} - T_{wu}) + E_{wu} \\ &= P_{on}h + P_{off}(i - h - T_s) + E_s\end{aligned}\tag{4.6}$$

**Caso C**

Nesse caso, o desligamento ocorre no início do intervalo ocioso. O religamento depende da estimativa de intervalo ocioso  $\hat{i}$ , que é a previsão de duração do intervalo ocioso. Se esse valor for menor que o *intervalo ocioso* em curso, o provedor de serviços será religado antes da chegada estimada de uma requisição da carga de trabalho. Quanto antes aconteça esse religamento, menor terá sido a economia de energia durante o intervalo ocioso.

Para o *Caso C*, o consumo de energia é dado em (4.7):

$$\begin{aligned}\mathcal{E} &= P_{on}h + E_{sd} + P_{off}(\hat{i} - h - T_{sd} - T_{wu}) \\ &\quad + E_{wu} + P_{on}(i - \hat{i}) \\ &= P_{on}h + P_{off}(\hat{i} - h - T_s) + P_{on}(i - \hat{i}) \\ &\quad + E_s\end{aligned}\tag{4.7}$$

**Caso D**

Esse caso pode acontecer em qualquer política de energia que atue realizando desligamentos. O desligamento ocorre quando é identificado o início de um intervalo ocioso, ou no caso de uma *política de timeout*, quando o valor do *timeout* expira. O religamento é feito de forma manual pelo usuário. Existe uma diferença em como esse caso é visto do ponto de vista de *política de timeout* e do ponto de vista de *políticas preditivas*. Em uma *política de timeout*, esse é o comportamento padrão: quando a política atua desligando o provedor de serviços, será sempre o usuário quem deverá religar. Em uma política preditiva o comportamento representado pelo *Caso D* significa que a estimativa para o intervalo ocioso foi maior que o *intervalo ocioso* em curso, e a política não conseguirá religar o provedor de serviços antes da chegada de uma requisição da carga de trabalho, sendo essa uma exceção ao correto funcionamento desse tipo de política.

Para o *Caso D*, o consumo de energia é dado em (4.8):

$$\begin{aligned}\mathcal{E} &= P_{on}h + E_{sd} + P_{off}(i - h - T_{sd}) + E_{wu} \\ &= P_{on}h + P_{off}(i - h - T_{sd}) + E_s\end{aligned}\tag{4.8}$$

**Caso E**

Esse último caso pode ocorrer em duas situações: quando o intervalo ocioso tem o mesmo tamanho do *timeout*, i.e.,  $i = \delta$  ou, quando  $i = h$ . O comando para desligar a interface de rede será enviado no instante em que a requisição da carga de trabalho chegar.

Para o *Caso E*, o consumo de energia é dado em (4.9):

$$\mathcal{E} = P_{on}i + E_s\tag{4.9}$$

### 4.3 Política do tempo ótimo de desligamento

A predição do intervalo ocioso  $i$  é essencial em uma política eficiente no gerenciamento de energia. Entretanto, a distribuição e variação de  $i$  podem ser fortemente dependentes do comportamento do usuário, do tipo de aplicações e do sistema operacional [27].

Uma técnica de gerenciamento de energia pode realizar desligamentos durante os intervalos ociosos, por exemplo. Para que essa técnica seja eficaz quanto à economia de energia, é preciso que o intervalo ocioso tenha um comprimento mínimo, denominado de Break-Even Time,  $T_{be}$ , que é o tempo ocioso necessário para igualar a economia de energia à energia de transição [8]. Além disso, para que a técnica não comprometa o desempenho do sistema, é necessário que satisfaça critérios de penalidade de desempenho. Nesse trabalho será implementada uma política preditiva que minimiza o consumo de energia atendendo uma restrição de penalidade de desempenho.

A ideia básica para a política proposta é apresentada na Figura 4.4. Se um intervalo ocioso inicia e o valor médio dos intervalos ociosos anteriores é maior ou igual ao tempo ótimo de desligamento  $T$ , então o gerenciador de energia emite o comando de desligar o provedor de serviços, como apresentado na Figura 4.4. Nesse caso, nenhuma penalidade de desempenho é computada, porque o provedor de serviços será religado antes da chegada de uma requisição da carga de trabalho. A energia consumida para o caso apresentado na Figura 4.4, onde o intervalo ocioso é maior ou igual ao tempo ótimo de desligamento, é apresentado em (4.10).

$$\mathcal{E}(T, i) = P_{on}h + P_{off}(T - h - T_s) + P_{on}(i - T) + E_s, \text{ se } i \geq T \quad (4.10)$$

Na Figura 4.5 é apresentado um caso onde o intervalo ocioso  $i$  é menor que o tempo ótimo de desligamento,  $T$ , i.e., foi estimado que o intervalo ocioso seria maior do que realmente foi, e não haverá tempo para realizar um religamento preditivo. Esse caso é associado ao cenário apresentado na Figura 4.3 (d), onde não há religamento preditivo do provedor de serviços, é computada uma penalidade de desempenho, e a energia consumida é dada por (4.11).

Figura 4.4: Consumo de energia para um provedor de serviços quando o gerenciador de energia emite um comando para desligar e o intervalo ocioso é maior ou igual ao tempo ótimo de desligamento.

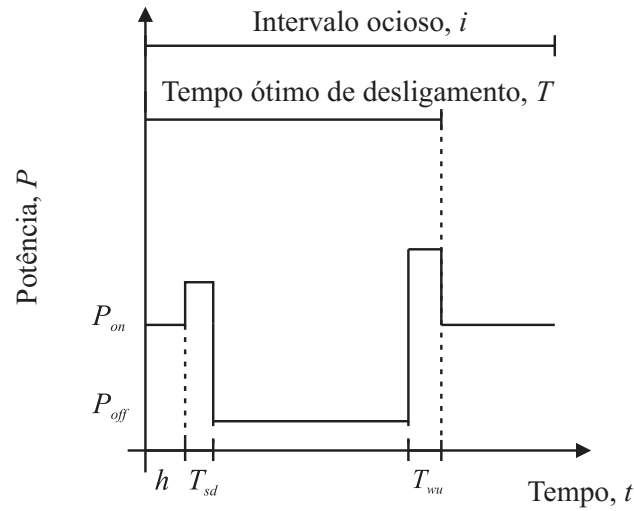
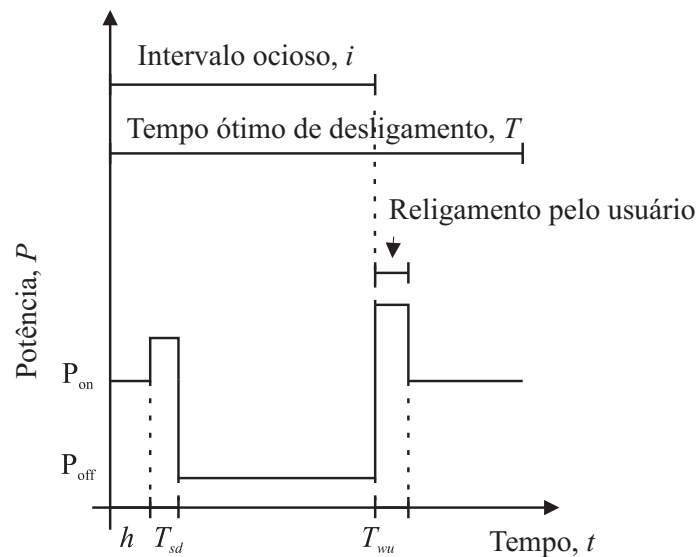


Figura 4.5: Consumo de energia para um provedor de serviços quando o gerenciador de energia emite um comando para desligar e o intervalo ocioso é menor que o tempo ótimo de desligamento.



$$\mathcal{E}(T, i) = P_{on}h + P_{off}(i - h - T_{sd}) + E_s, \text{ se } i < T \quad (4.11)$$

Para a técnica de gerenciamento dinâmico proposta nesse trabalho, será assumido que existe um tempo ótimo de desligamento,  $T$ , que minimiza o consumo de energia enquanto atende uma determinada restrição de penalidade. Em seguida, serão apresentados os modelos analíticos de energia e penalidade de desempenho para o provedor de serviços.

### 4.3.1 Modelo analítico de energia

A energia consumida por um PS, com tal GE, durante um intervalo ocioso, é apresentado em (4.12), onde  $E[I]$  é o valor estimado do intervalo ocioso  $i$ .

$$\mathcal{E}(T, i) = \begin{cases} P_{on}, & \text{se } E[I] < T, \forall i \\ P_{on}h + P_{off}(T - h - T_s) + P_{on}(i - T) + E_s, & \text{se } E[I] \geq T \text{ e } i \geq T \\ P_{on}h + P_{off}(i - h - T_{sd}) + E_s, & \text{se } E[I] \geq T \text{ e } i \leq T \\ P_{on}h + E_s, & \text{se } E[I] \geq T \text{ e } i = 1 \end{cases} \quad (4.12)$$

O valor esperado da energia consumida pelo provedor de serviços para um intervalo ocioso  $i$ , e um dado  $T$ , é apresentado em (4.13).

$$E[\mathcal{E}(T)] = \sum_{i=1}^{\infty} Pr[I = i] \cdot \mathcal{E}(T, i) \quad (4.13)$$

O gerenciador de energia irá desligar o provedor de serviços apenas se, o valor esperado dos intervalos ociosos  $E[I]$  é maior ou igual ao tempo ótimo de desligamento,  $T$ . Substituindo (4.12) em (4.13), o valor esperado na energia é dado em (4.14).



$$\begin{aligned}
 E[\mathcal{E}(T)] &= Pr[I = 1] \cdot \mathcal{E}(T, 1) + \sum_{i=2}^{T-1} Pr[I = i] \cdot \mathcal{E}(T, i) + \sum_{i=T}^{\infty} Pr[I = i] \cdot \mathcal{E}(T, i) \\
 &= Pr[I = 1] \cdot (P_{on}h + E_s) + \sum_{i=2}^{T-1} Pr[I = i] \cdot [P_{on}h + P_{off}(i - h - T_{sd}) + E_s] \\
 &\quad + \sum_{i=T}^{\infty} Pr[I = i] \cdot [P_{on}h + P_{off}(T - h - T_s) + P_{on} \cdot (i - T) + E_s] \\
 &= Pr[I = 1] \cdot (P_{on}h + E_s) + \sum_{i=2}^{T-1} Pr[I = i] \cdot [P_{on}h - P_{off}(h + T_{sd}) + E_s] + \\
 &\quad + \sum_{i=T}^{\infty} Pr[I = i] \cdot [P_{on}(h - T) + P_{off}(T - h - T_s) + E_s] + \\
 &\quad + P_{off} \cdot \sum_{i=2}^{T-1} Pr[I = i] \cdot i + P_{on} \sum_{i=T}^{\infty} Pr[I = i] \cdot i
 \end{aligned} \tag{4.14}$$

Utilizando (4.3), algumas simplificações podem ser feitas em (4.14):

$$Pr[I = i] = (1 - q)^{i-1} \cdot q \tag{4.15}$$

$$Pr[I = 1] = q \tag{4.16}$$

$$\begin{aligned}
 \sum_{i=2}^{T-1} Pr[I = i] &= \sum_{i=1}^{T-1} Pr[I = i] - Pr[I = 1] \\
 &= 1 - (1 - q)^{T-1} - q \\
 &= (1 - q) - (1 - q)^{T-1}
 \end{aligned} \tag{4.17}$$

$$\begin{aligned}
 \sum_{i=T}^{\infty} Pr[I = i] &= 1 - \sum_{i=1}^{T-1} Pr[I = i] \\
 &= (1 - q)^{T-1}
 \end{aligned} \tag{4.18}$$

$$\begin{aligned}
 \sum_{i=T}^{\infty} Pr[I = i] \cdot i &= \sum_{i=1}^{\infty} Pr[I = i] \cdot i - \sum_{i=1}^{T-1} Pr[I = i] \cdot i \\
 &= \frac{1}{q} - \sum_{i=1}^{T-1} (1 - q)^{i-1} \cdot q \cdot i
 \end{aligned} \tag{4.19}$$

Substituindo (4.16), (4.17), (4.18) e (4.19) em (4.14), o valor esperado da energia é definido em (4.20).

$$\begin{aligned}
 E[\mathcal{E}(T)] &= q \cdot (P_{on} \cdot h + E_s) + [(1 - q) - (1 - q)^{T-1}] \cdot [P_{on} \cdot h - P_{off} \cdot (h + T_{sd}) + E_s] \\
 &\quad + [P_{on} \cdot (h - T) + P_{off} \cdot (T - h - T_s) + E_s] \cdot [(1 - q)^{T-1}] \\
 &\quad + P_{on} \cdot \left[ \frac{1}{q} - \sum_{i=1}^{T-1} (1 - q)^{i-1} \cdot q \cdot i \right] + P_{off} \cdot \left[ \sum_{i=1}^{T-1} (1 - q)^{i-1} \cdot q \cdot i - q \right] \tag{4.20}
 \end{aligned}$$

### 4.3.2 Modelo analítico de penalidade de desempenho

O nível de degradação do desempenho do sistema é aqui avaliado em termos da penalidade de desempenho  $l(T, i)$ :

- Se  $i \geq T$ , o gerenciador de energia irá religar o provedor de serviços ao final do tempo ótimo de desligamento  $T$ , ainda durante o intervalo ocioso  $i$ . O usuário encontrará o provedor de serviços ligado, e não haverá nenhuma penalidade no desempenho do sistema:  $l(T, i) = 0$ .
- Se  $i < T$ , o intervalo inativo  $i$  termina antes de completar o tempo ótimo de desli-

gamento  $T$ . O usuário encontrará o provedor de serviços desligado e precisará religar manualmente, nesse caso:  $l(T, i) = 1$ .

O valor esperado da penalidade de desempenho é apresentado em (4.21):

$$\begin{aligned}
 E[l(T)] &= \sum_{i=1}^{\infty} Pr[I = i] \cdot l(T, i) \\
 &= \sum_{i=1}^{T-1} Pr[I = i] \cdot 1 + \sum_{i=T}^{\infty} Pr[I = i] \cdot 0 \\
 &= 1 - (1 - q)^{T-1}
 \end{aligned} \tag{4.21}$$

### 4.3.3 Obtenção do tempo ótimo de desligamento $T$

No gerenciamento dinâmico de energia, é importante considerar tanto o consumo de energia quanto a penalidade de desempenho. Quando o consumo de energia é reduzido, é muitas vezes necessário limitar a penalidade de desempenho a níveis aceitáveis. Além disso, só é viável a implementação de uma política de gerenciamento de energia se o consumo de energia com a mesma for menor que a energia consumida quando nenhuma política é utilizada  $\mathcal{E}_{on}$ , caso contrário, seria melhor não ter nenhuma política. Portanto, para resolver o *problema de otimização* (PO), proposto em (4.22), é preciso encontrar os valores ótimos de  $T$  que minimizam o consumo de energia, para uma dada restrição de penalidade de desempenho, onde  $L$  é a penalidade de desempenho máxima aceitável e uma restrição de consumo, onde  $\mathcal{E}_{on}$  é o consumo máximo aceitável, definido em (4.23).

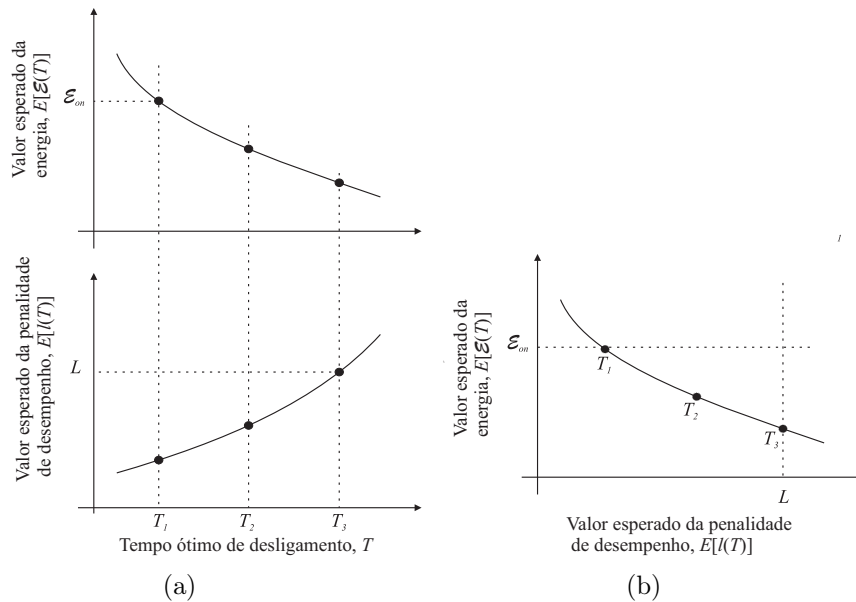
$$\begin{aligned}
 \mathbf{PO} : \quad & \min_T E[\mathcal{E}(T)] \\
 \text{tal que} \quad & E[l(T)] \leq L \\
 & E[\mathcal{E}(T)] \leq \mathcal{E}_{on}
 \end{aligned} \tag{4.22}$$

$$\begin{aligned}
 \mathcal{E}_{on} &= E[P_{on}I] = \sum_{i=1}^{\infty} Pr[I = i]P_{on}i \\
 &= P_{on} \sum_{i=1}^{\infty} Pr[I = i]i \\
 &= P_{on}E[I]
 \end{aligned} \tag{4.23}$$

A partir das expressões (4.20) e (4.21), tanto  $E[\mathcal{E}(T)]$  quanto  $E[l(T)]$  são funções de  $q$ , i.e., dependem da carga de trabalho. Assim, o problema de otimização é resolvido para cada carga de trabalho.

Por exemplo, considere a energia média e a penalidade de desempenho média mostradas na Figura 4.6 (a). Esses valores podem ser apresentados como a curva Valor esperado da energia *versus* a penalidade de desempenho, conforme representado na Figura 4.6 (b).

Figura 4.6: Exemplo do valor esperado da energia versus a penalidade de desempenho.



Para o tempo de desligamento  $T_1$ , o provedor de serviços consome o valor máximo de energia  $\mathcal{E}_{on}$  permitido pela restrição de energia. Para o tempo de desligamento  $T_2$ , o consumo de energia do provedor de serviços é menor que a restrição de energia  $\mathcal{E}_{on}$ , e a penalidade de desempenho é menor que a restrição de penalidade de desempenho  $L$ , mas o consumo

de energia ainda pode ser diminuído, aumentando a penalidade de desempenho. Isto é conseguido para o tempo de desligamento  $T_3$ , para o qual o valor mínimo de energia média é alcançado, enquanto satisfaz as restrições de energia e desempenho. Assim, para este exemplo, a solução do problema de otimização em (4.22) é  $T_3$ .

#### 4.3.4 Tabela de políticas

Em tempo de execução, recalculando uma política ótima a cada período de amostragem não é viável. É preciso que o gerenciador de energia tome decisões rápidas e que o gasto de energia possa ser desconsiderado. Para tanto, no trabalho de Chung et al. [7] é proposta uma tabela de consulta (look-up table) com interpolação linear para a obtenção da política de energia em tempo de execução.

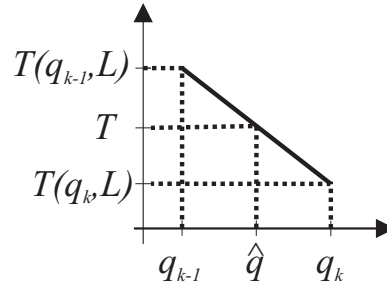
Na Tabela 4.1, é apresentado um exemplo de Tabela de Políticas. Cada linha corresponde a um parâmetro da carga de trabalho,  $q$ , onde  $q = 1 - p_{00}$ . Cada coluna,  $L$ , corresponde a uma especificação do problema de otimização. Para cada par  $(q, L)$ , o problema de otimização em (4.22) é resolvido por meio da metodologia apresentada no trabalho de Luiz [38], resultando em um  $T(q, L)$  ótimo que minimiza o consumo de energia respeitando a restrição máxima de penalidade de desempenho  $L$  e de energia  $\mathcal{E}_{on}$ .

Tabela 4.1: Tabela de políticas.

	Restrição de penalidade de desempenho, L		
<b>Carga, q</b>	0,1	...	1,0
0,1	$T(0,1, 0,1)$	...	$T(0,1, 1,0)$
⋮	⋮	⋮	⋮
0,9	$T(0,9, 0,1)$	...	$T(0,9, 1,0)$

Em tempo de execução, num dado período de amostragem, uma estimativa da carga de trabalho,  $\hat{q}$ , é obtida. Como tal estimativa pode não ser exatamente igual aos valores de  $q$  usados para a construção da tabela de políticas, então são determinados os valores de  $q$  na tabela de políticas mais próximos da estimativa, ou seja,  $q_{k-1} < \hat{q} < q_k$ . De acordo com a restrição de penalidade adotada, uma técnica de interpolação linear é aplicada para que um  $T(\hat{q}, L)$  ótimo seja escolhido, como representado na Figura 4.7 e calculado em (4.24).

Figura 4.7: Interpolação linear para uma penalidade de desempenho especificada.

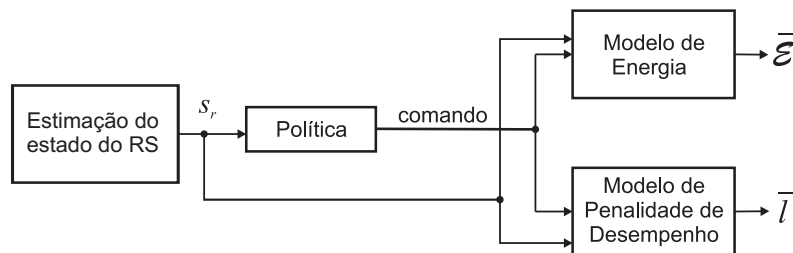


$$T(\hat{q}, L) = \frac{(\hat{q} - q_k) \cdot [T(q_{k-1}, L) - T(q_k, L)]}{q_{k-1} - q_k} + T(q_k, L) \quad (4.24)$$

### 4.3.5 Simulação da política de energia

Para verificar os resultados da política proposta e mesmo comparar com outras políticas, é utilizado o esquema apresentado no diagrama de blocos na Figura 4.8. A partir do log de uma carga de trabalho (taxa de *download* + taxa de *upload*), é estimado o estado do requisitante de serviços no bloco *Estimacão do estado do RS*, i.e., identifica-se se o dispositivo está ativo ou ocioso. O estado do RS é agora entrada para o bloco *Política*, que é o bloco responsável pela tomada de decisão para cada intervalo ocioso, ou seja, é onde é definido se haverá ou não desligamento para aquele intervalo ocioso, e qual o tempo ótimo de desligamento.

Figura 4.8: Diagrama de blocos para simulação de uma política de energia.



Nos blocos *Modelo de Energia* e *Modelo de Penalidade de Desempenho*, são computados

o consumo médio da energia,  $\bar{\mathcal{E}}$ , e a penalidade de desempenho,  $\bar{l}$ , respectivamente. Esses blocos precisam receber como entrada o comando emitido pela política e também o estado real do RS. Dessa forma é possível computar corretamente o consumo de energia e se houve ou não uma penalidade de desempenho para aquele intervalo ocioso.

O pseudo-código que representa o funcionamento da política proposta é apresentado no Algoritmo 1 onde  $E(I)$ ,  $T$ ,  $q$  e *intervaloOcioso* são dados,  $s_r$  é entrada e *comando* é a saída.

---

**Algoritmo 1:** Algoritmo da Política Proposta
 

---

```

Dados:  $E(I)$ ,  $T(\hat{q}, L)$ ,  $\hat{q}$  e  $i$ 
Entrada:  $s_r$ 
Saída: comando
se  $s_r == 0$  então
  |  $i = i + 1$ ; /* requisitante de serviços está ocioso */
  |  $\hat{q} = \text{JMC}(s_r)$ ; /* técnica (JMC) apresentada na Seção 3.2.2 */
  |  $T(\hat{q}, L) = \frac{(\hat{q}-q_k) \cdot [T(q_{k-1}, L) - T(q_k, L)]}{q_{k-1} - q_k} + T(q_k, L)$ ; /* Tempo ótimo de desligamento */
  |  $E(I) = \frac{1}{\hat{q}}$ ; /* valor médio do intervalo ocioso */
  | se  $E(I) \geq T(\hat{q}, L)$  então
  | | comando = desligar;
  | | se  $i \geq T(\hat{q}, L)$  então
  | | | comando = ligar;
  | | fim
  | fim
fim
senão
  |  $i = 0$ ; /* requisitante de serviços está ativo */
fim

```

---

## 4.4 Projeto de experimento

Os resultados apresentados neste trabalho, tais como, os comprimentos das cargas de trabalho utilizados em simulação, para determinar que a política proposta obteve melhor desempenho que outra em termos do consumo médio de energia e da penalidade de desempenho, foram definidos a partir do conceito de intervalo de confiança [15], considerando  $\rho = 95\%$  (nível de confiança) e portanto  $\alpha = 5\%$  (nível de significância, ou erro).

#### 4.4.1 Determinando o intervalo de confiança para $\rho = 95\%$

O princípio do intervalo de confiança é baseado no fato de que é impossível determinar uma média exata  $\mu$  para uma população de infinitas amostras  $N$ , considerando um número finito  $n$  de amostras  $\{x_1, \dots, x_n\}$ . Porém, em termos probabilísticos é possível determinar um intervalo ao qual  $\mu$  pertence com probabilidade igual a  $\rho$  e ao qual  $\mu$  não pertence com probabilidade igual a  $\alpha$ .

Para determinar o extremo esquerdo  $c_1$  e extremo direito  $c_2$  deste intervalo, chamado de intervalo de confiança, considera-se uma probabilidade  $1 - \alpha$ , tal que o valor  $\mu$  pertença a este intervalo de confiança, para  $n$  repetições de um determinado experimento realizado. Assim, temos a seguinte relação:

$$Pr\{c_1 < \mu < c_2\} = 1 - \alpha \quad (4.25)$$

onde,

- $(c_1, c_2)$  é o intervalo de confiança;
- $\alpha$  é o nível de significância, expresso como uma fração e tipicamente perto de zero, por exemplo, 0,05 ou 0,1;
- $(1 - \alpha)$  é o coeficiente de confiança; e
- $\rho = 100 \cdot (1 - \alpha)$ , é o nível de confiança, tradicionalmente expresso como porcentagem e tipicamente perto de 100%, por exemplo, 90% ou 95%.

Com o teorema do limite central [15] é possível determinar o intervalo de confiança para poucas amostras. Este teorema afirma que, se as observações em uma amostra  $\{x_1; \dots; x_n\}$  são independentes e vêm da mesma população que tem uma média  $\mu$  e um desvio padrão  $\sigma$ , então a média da amostra para amostras grandes é aproximadamente normalmente distribuída com média  $\mu$  e desvio padrão  $\sigma/\sqrt{n}$ .

$$\bar{x} \simeq N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad (4.26)$$

Então, considerando (4.25) e o Teorema do Limite Central (4.26), obtêm-se o intervalo de confiança  $(c_1, c_2)$  para  $\rho = 95\%$  e  $\alpha = 5\%$  da seguinte forma:



$$\left(\bar{x} - z_{1-\alpha/2} \times \frac{s}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2} \times \frac{s}{\sqrt{n}}\right) \quad (4.27)$$

onde,

- $\bar{x}$  é a média para  $n$  repetições;
- $z_{1-\alpha/2}$  é igual a 1,96. Esse valor determina 95% para o nível de confiança, como definido na Tabela A.2, do Apêndice A da referência [15];
- $n$  é igual ao número de repetições; e
- $s$  é o desvio padrão das médias para as  $n$  repetições.

Com relação ao valor 1,96 para o termo  $z_{1-\alpha/2}$ , também chamado de quantil, este é baseado no Teorema do Limite Central e por ser frequentemente utilizado, encontra-se na tabela de Quantis da Unidade de Distribuição Normal. Esta tabela pode ser encontrada no Apêndice A, Tabela A.2, da referência [15]. Para determinar este valor, temos:

$$1 - \alpha/2 = (1 - 0,05/2) = 0,975 \quad (4.28)$$

#### 4.4.2 Determinando o valor de $n$ para obter $\rho = 95\%$

O nível de confiança depende da quantidade  $n$  de amostras coletadas para um dado experimento. Assim, quanto maior o valor de  $n$ , maior será o nível de confiança. Entretanto, obter uma quantidade grande de amostras exige mais esforço e tempo. Portanto, é importante definir o valor de  $n$  de tal forma que consiga-se poupar esforço e tempo, porém mantendo o nível de confiança desejado, ou seja,  $\rho = 95\%$ .

Para um nível de confiança  $\rho = 95\%$  e  $\alpha = 5\%$ , o intervalo de confiança é:

$$(\mu(1 - 0,05), \mu(1 + 0,05)) \quad (4.29)$$

Então, igualando o intervalo de confiança (4.29) ao intervalo de confiança (4.27), obtemos (4.30).

$$\left(\mu - z_{1-\alpha/2} \times \frac{s}{\sqrt{n}}, \mu + z_{1-\alpha/2} \times \frac{s}{\sqrt{n}}\right) = (\mu(1 - 0,05), \mu(1 + 0,05)) \quad (4.30)$$

Portanto, organizando a expressão para isolar a variável  $n$ , cada experimento deverá ser repetido  $n$  vezes, indicado em (4.31), para um nível de confiança  $\rho = 95\%$ , o que implica em  $z = 1,96$  a partir de (4.28).

$$n = \left(\frac{1,96 \times s}{0,05 \times \mu}\right)^2 \quad (4.31)$$

Para definir a quantidade de amostras, vamos definir analiticamente o desvio padrão para a energia,  $\mathcal{E}(T)$  e para a penalidade,  $l(T)$ , respectivamente em (4.32) e (4.33).

$$s_{\mathcal{E}(T)} = \sqrt{Var[\mathcal{E}(T)]} \quad (4.32)$$

$$s_{l(T)} = \sqrt{Var[l(T)]} \quad (4.33)$$

A variância para a penalidade de desempenho é dada em (4.34).

$$\begin{aligned} Var[l(T)] &= E[l(T) - E[l(T)]]^2 \\ &= E[l(T)^2] - E[l(T)]^2 \end{aligned} \quad (4.34)$$

De (4.21) obtêm-se o valor de  $E[l(T)]^2$ . Já o valor de  $E[l(T)^2]$ , é dado em (4.35):

$$\begin{aligned} E[l(T)^2] &= \sum_{i=1}^{\infty} Pr[I = i] \cdot l(T, i)^2 \\ &= \sum_{i=1}^{T-1} Pr[I = i] \cdot 1^2 + \sum_{i=T}^{\infty} Pr[I = i] \cdot 0^2 \\ &= 1 - (1 - q)^{T-1} \end{aligned} \quad (4.35)$$

A variância para a energia é dada em (4.36):

$$\begin{aligned} Var[\mathcal{E}(T)] &= E[(\mathcal{E}(T) - E[\mathcal{E}(T)])^2] \\ &= E[\mathcal{E}(T)^2] - E[\mathcal{E}(T)]^2 \end{aligned} \quad (4.36)$$

De (4.20) obtêm-se o valor de  $E[\mathcal{E}(T)]^2$ . Já o valor de  $E[\mathcal{E}(T)^2]$  é dado em (4.37).

$$\begin{aligned} E[\mathcal{E}(T)^2] &= \sum_{i=1}^{\infty} Pr[I = i] \cdot \mathcal{E}(T, i)^2 \\ &= Pr[I = 1] \cdot \mathcal{E}(T, 1)^2 + \sum_{i=2}^{T-1} Pr[I = i] \cdot \mathcal{E}(T, i)^2 + \sum_{i=T}^{\infty} Pr[I = i] \cdot \mathcal{E}(T, i)^2 \\ &= Pr[I = 1] \cdot (P_{on}h + E_s)^2 + \sum_{i=2}^{T-1} Pr[I = i] \cdot [P_{on}h + P_{off}(i - h - T_{sd}) + E_s]^2 \\ &\quad + \sum_{i=T}^{\infty} Pr[I = i] \cdot [P_{on}h + P_{off}(T - h - T_s) + P_{on} \cdot (i - T) + E_s]^2 \\ &= (P_{on}h + E_s)^2 \cdot Pr[I = 1] + [P_{on}h - P_{off}(h + T_{sd}) + E_s]^2 \cdot \sum_{i=2}^{T-1} Pr[I = i] + \\ &\quad 2 \cdot [P_{on}h - P_{off}(h + T_{sd}) + E_s] \cdot \sum_{i=2}^{T-1} Pr[I = i] \cdot i + P_{off}^2 \cdot \sum_{i=2}^{T-1} Pr[I = i] \cdot i^2 + \\ &\quad [P_{on}(h - T) + P_{off}(T - h - T_s) + E_s]^2 \cdot \sum_{i=T}^{\infty} Pr[I = i] + \\ &\quad + 2 \cdot [P_{on}(h - T) + P_{off}(T - h - T_s) + E_s] \cdot \sum_{i=T}^{\infty} Pr[I = i] \cdot i + \\ &\quad + P_{on}^2 \cdot \sum_{i=T}^{\infty} Pr[I = i] \cdot i^2 \end{aligned} \quad (4.37)$$

Utilizando (4.3), algumas simplificações podem ser feitas em (4.37). Além das apresentadas em (4.15), (4.16), (4.17), (4.18) e (4.19), verifica-se também:

$$\begin{aligned}
 \sum_{i=1}^{\infty} Pr[I = i] \cdot i^2 &= \sum_{i=1}^{\infty} i \cdot (i + 1) \cdot (1 - q)^{i-1} \cdot q - \sum_{i=1}^{\infty} i \cdot (1 - q)^{i-1} \cdot q \\
 &= \sum_{i=1}^{\infty} i \cdot (i + 1) \cdot (1 - q)^{i-1} \cdot q - \frac{1}{q} \\
 &= q \frac{2}{q^3} - \frac{1}{q} \\
 &= \frac{2}{q^2} - \frac{1}{q}
 \end{aligned} \tag{4.38}$$

$$\begin{aligned}
 \sum_{i=1}^{\infty} Pr[I = i] \cdot i^2 &= \sum_{i=1}^{\infty} (1 - q)^{i-1} \cdot q \cdot i^2 - \sum_{i=1}^{T-1} (1 - q)^{i-1} \cdot q \cdot i^2 \\
 &= \frac{2}{q^2} - \frac{1}{q} - \sum_{i=1}^{T-1} (1 - q)^{i-1} \cdot q \cdot i^2
 \end{aligned} \tag{4.39}$$

Substituindo (4.15), (4.16), (4.17), (4.18), (4.19), (4.38) e (4.39), a variância da energia é dada em (4.40).

$$\begin{aligned}
 Var[\mathcal{E}(T)] = & (P_{on}h + E_s)^2 \cdot q + [P_{on}h - P_{off}(h + T_{sd}) + E_s]^2 \cdot \sum_{i=2}^{T-1} (1-q)^{i-1} \cdot q + \\
 & + 2[P_{on}h - P_{off}(h + T_{sd}) + E_s] \cdot P_{off} \cdot \sum_{i=2}^{T-1} (1-q)^{i-1} \cdot i + \\
 & + P_{off}^2 \cdot \sum_{i=2}^{T-1} (1-q)^{i-1} \cdot i^2 + \\
 & + [P_{on}(T-h) + P_{off}(T-h-T_s) + E_s]^2 \cdot \sum_{i=T}^{\infty} (1-q)^{T-1} + \\
 & + 2[P_{on}(T-h) + P_{off}(T-h-T_s) + E_s] \cdot P_{on} \cdot \left[ \frac{1}{q} - \sum_{i=1}^{T-1} (1-q)^{i-1} q \cdot i \right] + \\
 & + P_{on}^2 \cdot \left[ \frac{2}{q^2} - \frac{1}{q} - \sum_{i=1}^{T-1} (1-q)^{i-1} \cdot q \cdot i^2 \right] \tag{4.40}
 \end{aligned}$$

Para cada valor de  $T$ , tem-se um valor esperado de energia, variância da energia, valor esperado da penalidade de desempenho e variância da penalidade de desempenho. Dessa forma, utilizando (4.31), é possível definir o valor  $n$  de repetições para cada  $T$ , em cada tipo de carga de trabalho, necessárias para 95% de nível de confiança. Na Figura 4.9 estão apresentados os valores de  $n$  para a carga de trabalho ( $q = 0,1$ ). Nas Figuras 4.10 e 4.11 são apresentados os resultados de (4.31) para as cargas de trabalho com ( $q = 0,3$ ) e ( $q = 0,6$ ), respectivamente.

Figura 4.9: Valor  $n$  de repetições para cada  $T$ , para a carga de trabalho com  $(q = 0, 1)$ .

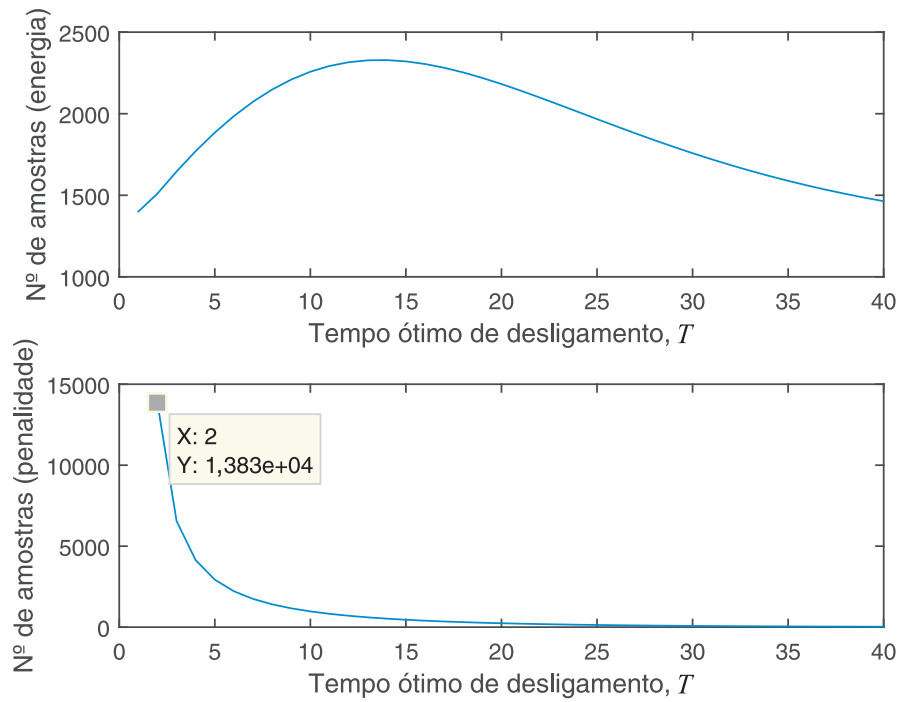


Figura 4.10: Valor  $n$  de repetições para cada  $T$ , para a carga de trabalho com  $(q = 0, 3)$ .

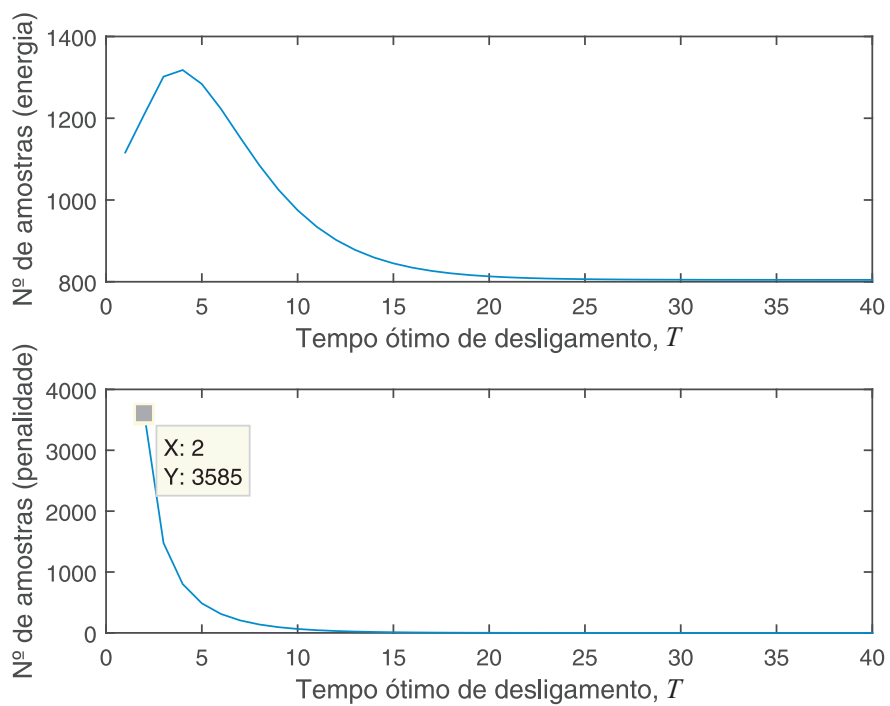
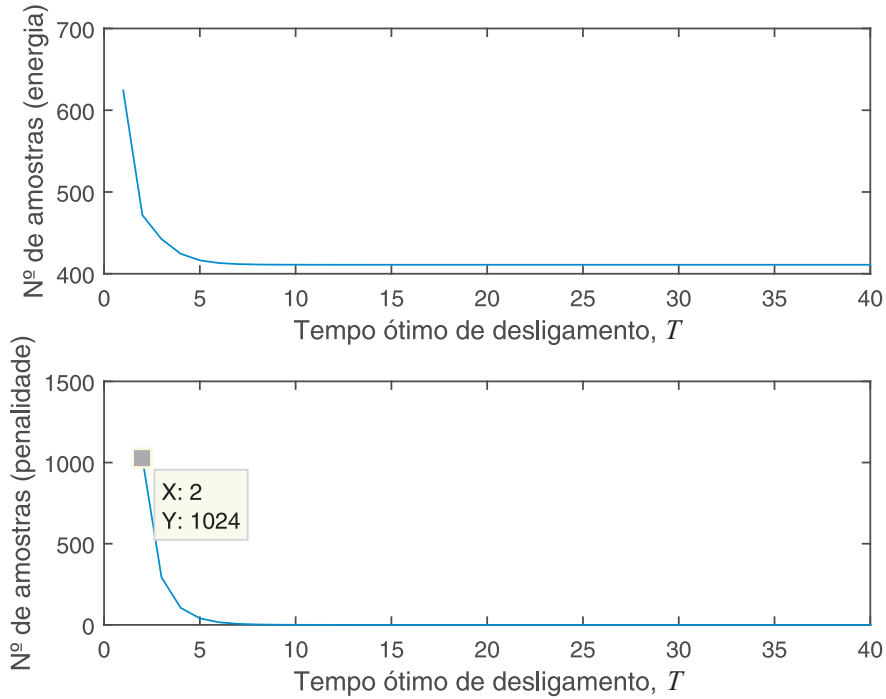


Figura 4.11: Valor  $n$  de repetições para cada  $T$ , para a carga de trabalho com  $(q = 0,6)$ .



Para definir um único valor de  $n$  para cada tipo de carga de trabalho e assim realizar as simulações, foi escolhido para cada caso o maior valor de  $n$ . Na Figura 4.12 estão apresentados esses valores. Utilizando um ajuste de curvas, é possível aproximar essa curva por uma soma de duas exponenciais. O valor de  $n$  para qualquer carga de trabalho é obtido em (4.41).

$$n = 40550 \cdot e^{-19,71 \cdot q} + 12510 \cdot e^{-4,24 \cdot q} \tag{4.41}$$

Durante um intervalo de tempo  $N$ , uma carga de trabalho apresenta a composição de intervalos ociosos e intervalos ativos apresentada na Figura 4.13. Onde o valor esperado do intervalo ocioso é  $\frac{1}{q}$ , dado em (4.4), e de maneira semelhante, o valor esperado para o intervalo ativo é  $\frac{1}{p}$ .

Figura 4.12: Valor  $n$  de repetições *versus* cargas de trabalho.

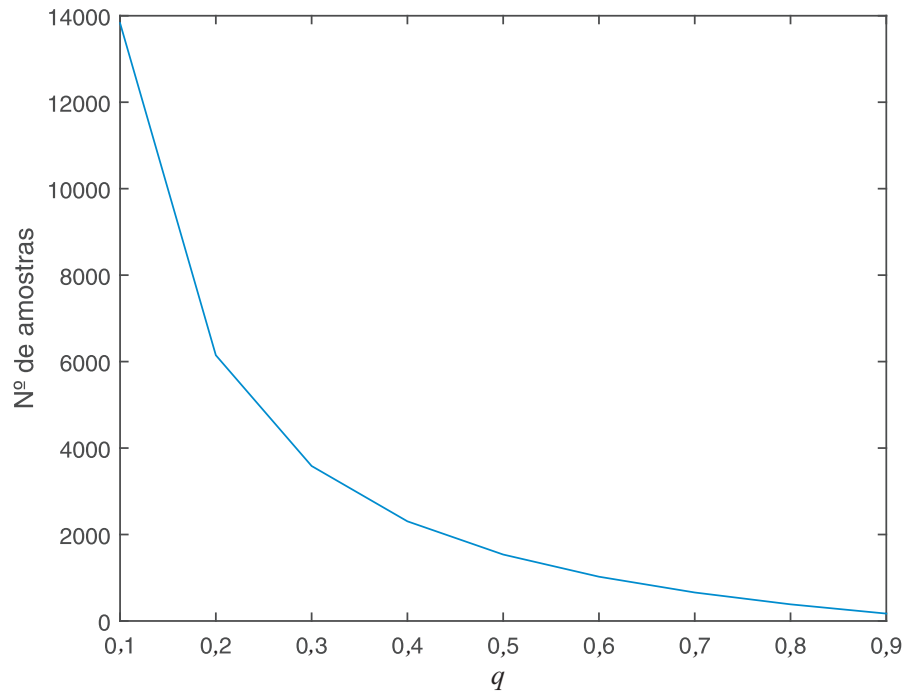
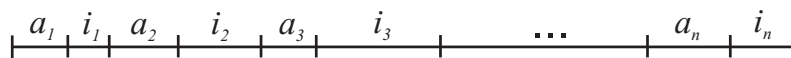


Figura 4.13: Ilustração da composição dos intervalos ociosos e intervalos ativos ao longo do intervalo de tempo  $N$ .



A média de todos os intervalos ociosos ao longo do intervalo de tempo  $N$ , é dado em (4.42):

$$\frac{\sum_{k=1}^n i_k}{n} \approx \frac{1}{q} \tag{4.42}$$

A média de todos os intervalos ativos ao longo do intervalo de tempo  $N$ , é dado em (4.43):

$$\frac{\sum_{k=1}^n a_k}{n} \approx \frac{1}{p} \tag{4.43}$$

A partir de (4.42) e (4.43) é possível definir o comprimento  $N$  da amostra em (4.44):



$$\begin{aligned}
N &= \sum_{k=1}^n i_k + \sum_{k=1}^n a_k \\
&\cong \frac{1}{q} + n\frac{1}{p} \\
&\cong n\left(\frac{1}{q} + \frac{1}{p}\right)
\end{aligned} \tag{4.44}$$

## 4.5 Considerações finais

Nesse capítulo foi apresentado todo o processo utilizado para o desenvolvimento da política proposta para o gerenciamento dinâmico de energia, denominada Política do Tempo Ótimo de Desligamento, TOD. A carga de trabalho foi modelada como uma cadeia de Markov de apenas dois estados, *ativo* e *ocioso*, e a partir das probabilidades de transição,  $p_{0,1} = q$ , é possível determinar o valor esperado para os intervalos ociosos,  $E[I] = \frac{1}{q}$ , ao longo do tempo. Após definir um modelo acurado para a carga de trabalho, foram definidos estados de consumo de energia para um provedor de serviços que realiza transições entre os estados *on* e *off*, considerando todos os casos possíveis de desligamentos.

Considera-se nesse trabalho que exista um *tempo ótimo* para o qual o dispositivo com GDE deverá permanecer desligado ou em um estado de baixo consumo. Entretanto, esse tempo não é necessariamente todo o intervalo ocioso, pois os comprimentos dos intervalos ociosos variam ao longo do tempo. Para cada carga de trabalho conhecida a priori, é possível determinar o tempo ótimo,  $T$ , sobre métricas de economia de energia e penalidade de desempenho. Todo o processo de construção da política é apresentado em detalhes. São desenvolvidos os modelos analíticos para penalidade de desempenho e para consumo de energia. Um problema de otimização é formulado e a tabela de políticas é definida.

Para simular a política proposta e obter resultados para variadas cargas de trabalho, é apresentado o diagrama de blocos e o Algoritmo utilizado para realizar as simulações. Em situações práticas, amostras das cargas de trabalho vão variar no comprimento, e, quando muito pequenas, apresentam variância grande. Para definir comprimentos de amostras que sejam suficientes para realizar as simulações e obter resultados confiáveis, é utilizado o Teorema do Limite Central, e para tanto define-se analiticamente o desvio padrão para

o consumo de energia e para a penalidade de desempenho. De forma que, ao fim desse capítulo, todo o processo para a construção da política e simulação foram definidos.

No próximo capítulo será apresentado um estudo de caso, onde a política proposta será aplicada para um dispositivo de rede Wi-Fi, e comparada com outras duas políticas frequentemente utilizadas.

# Capítulo 5

## Estudo de Caso

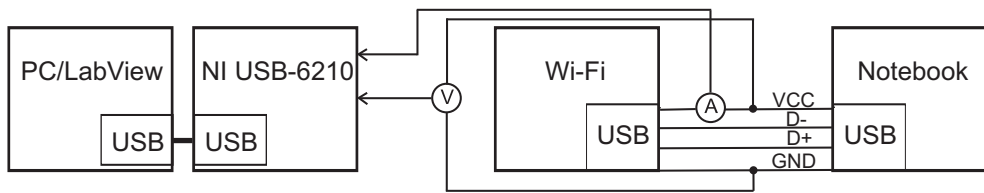
Nesse capítulo é apresentado um estudo de caso, onde a política proposta foi aplicada para um dispositivo de rede, Wi-Fi. As cargas de trabalho para interfaces de rede variam no tempo, de acordo com a utilização do usuário e também com processos que estejam sendo executados em background. É possível solicitar do sistema operacional informações sobre a carga de trabalho, como as taxas de (*download* e *upload*), ao longo do tempo. Utilizando um adaptador Wireless USB, foi possível medir a corrente e a tensão demandados pelo dispositivo e assim obter informações sobre o consumo de energia necessárias para compor o modelo analítico de energia. O problema de otimização foi formulado e resolvido *off-line*, uma tabela de políticas foi definida com os *tempos ótimos de desligamento*,  $T$ , para cada carga de trabalho. Através de simulações em variados cenários de cargas de trabalho, o comportamento da política proposta foi validado e comparado com outras duas políticas: *timeout* e preditiva, para o dispositivo de rede.

### 5.1 Modelo de energia para a interface de rede

Para medir a tensão e corrente que são fornecidas à interface de rede, um módulo de aquisição de dados NI USB-6210 é conectado a um computador pessoal (PC) via *Universal Serial Bus* (USB), como ilustrado na Figura 5.1. No PC é utilizado o programa LabVIEW no sistema operacional Windows. Dessa forma é possível medir o consumo pela interface de rede [41].

Como estudo de caso, foi escolhido como provedor de serviços o adaptador USB - Wi-Fi

Figura 5.1: Diagrama de blocos para medição da tensão e corrente demandados pelo adaptador Wireless USB.



(EduP), padrão 802.11n e 150 Mbps, representado na Figura 5.2. Foram realizadas medições do consumo dessa interface de rede, e não apenas estimações a partir do uso total do sistema computacional como feito anteriormente por Vilar et al. [42].

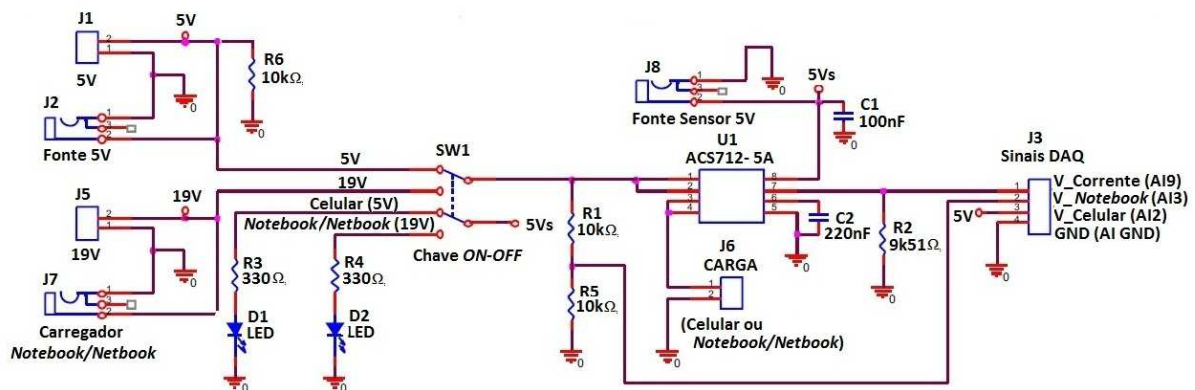
Figura 5.2: Micro Adaptador USB - Wi-Fi 150Mbps



Fonte: acesso online<sup>1</sup>

Na Figura 5.3 é apresentado o diagrama elétrico do circuito proposto por Melo et al. [43], utilizado para monitoramento da tensão e corrente demandados pelo adaptador de rede Wi-Fi.

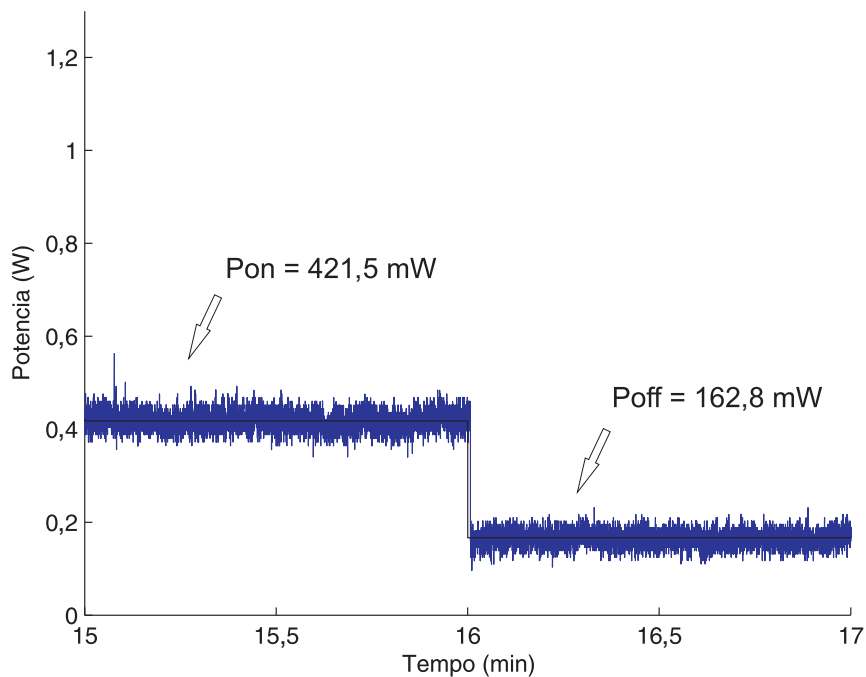
Figura 5.3: Diagrama elétrico do circuito elétrico para monitoramento da potência consumida.



Fonte: Melo et al. [43].

Foram realizadas medições do consumo de energia da interface de rede enquanto ativa e ociosa,  $P_{on}$ , durante a transição para o estado de baixo consumo, e no estado de baixo consumo,  $P_{off}$ . De forma semelhante foram realizadas medições do consumo de energia enquanto a interface de rede estava no estado de baixo consumo,  $P_{off}$ , durante a transição para o estado ativo, e no estado ativo,  $P_{on}$ . Nas Figuras 5.4 e 5.5, estão representados o consumo de potência para a interface de rede durante os estados ativo porém ocioso,  $P_{on}$ , e desligado,  $P_{off}$  bem como os momentos de transição entre esses estados.

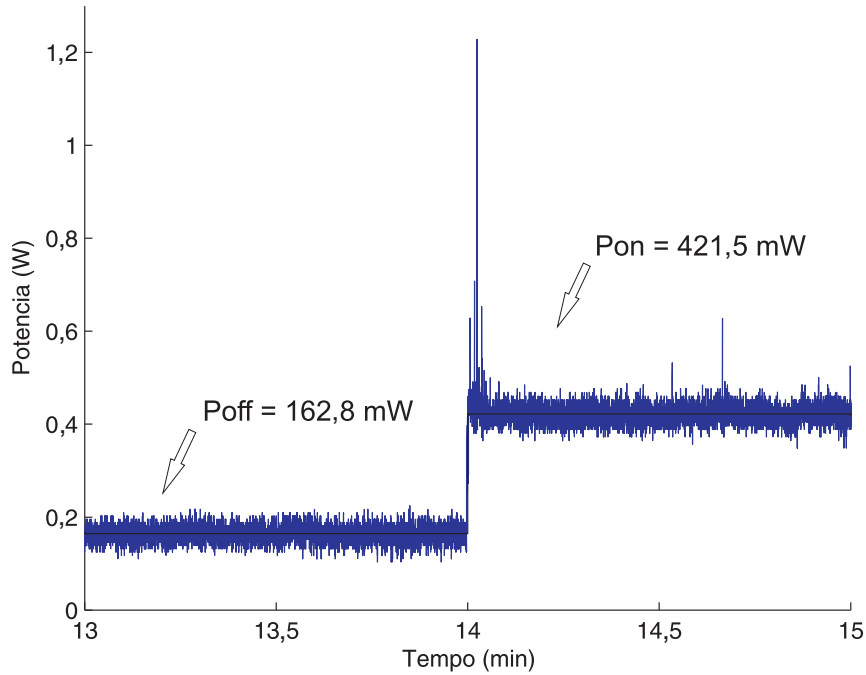
Figura 5.4: Potência média consumida pela interface ativa,  $P_{on}$ , e desligada,  $P_{off}$ .



A partir dessas medições foi possível obter os valores para  $P_{off}$ ,  $P_{off}$ ,  $T_{sd}$ ,  $T_{wu}$  e  $E_s$ . Utilizando o intervalo de confiança foi possível determinar a quantidade de repetições necessárias para se obter os valores médios apresentados na Tabela 5.1.

Tabela 5.1: Consumo médio da interface de rede.

Consumo da interface ativa e ociosa, $P_{on}$	421,5 (mW)
Consumo da interface desligada, $P_{off}$	162,8 (mW)
Energia do chaveamento, $E_s$	2,4 (J)
Tempo para desligar, $T_{sd}$	1 (s)
Tempo para religar, $T_{wu}$	5 (s)

Figura 5.5: Potência média consumida pela interface desligada,  $P_{off}$ , e ativa,  $P_{on}$ .

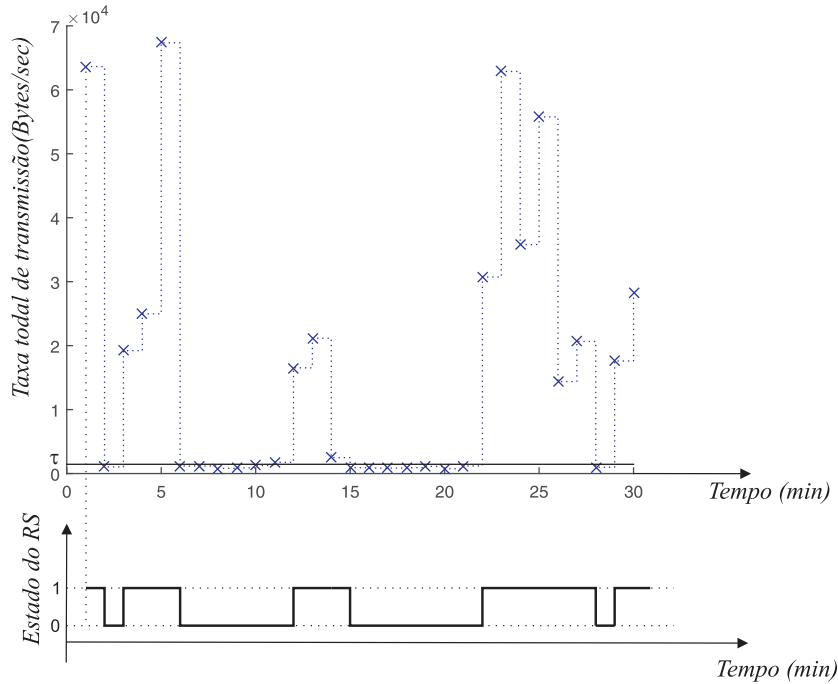
## 5.2 Carga de trabalho para interfaces de rede

Na Figura 5.6 está ilustrado um exemplo de *log* da carga de trabalho de uma interface de rede, variando de acordo com a utilização. A partir da aquisição das taxas de *download* e *upload*, é possível definir um limiar que permita identificar os intervalos em que o RS está ativo ou ocioso.

O limiar da interface de rede é definido a partir de experimentos controlados onde se tem conhecimento da aplicação que está sendo executada ou se um intervalo ocioso está ocorrendo. A partir da aquisição desses dados, é calculada uma função custo (5.1), para definir o limiar que minimize o erro entre o estado estimado da interface de rede,  $\hat{a}(t)$ , e o estado real informado pelo usuário,  $a(t)$ . Em tempo de execução a taxa média (*download* + *upload*) é comparada com o limiar pré-estabelecido e dessa forma é possível inferir o estado da interface de rede [23], como ilustrado na Figura 5.6.

$$J_{t_0, t_f}(\tau) = \sum_{t_0=1}^{t_f} \varepsilon^2(t) = \sum_{t_0=1}^{t_f} (a(t) - \hat{a}(t|\tau))^2 \quad (5.1)$$

Figura 5.6: Intervalo de tempo de uma carga de trabalho gerada pelo uso de uma interface de rede.



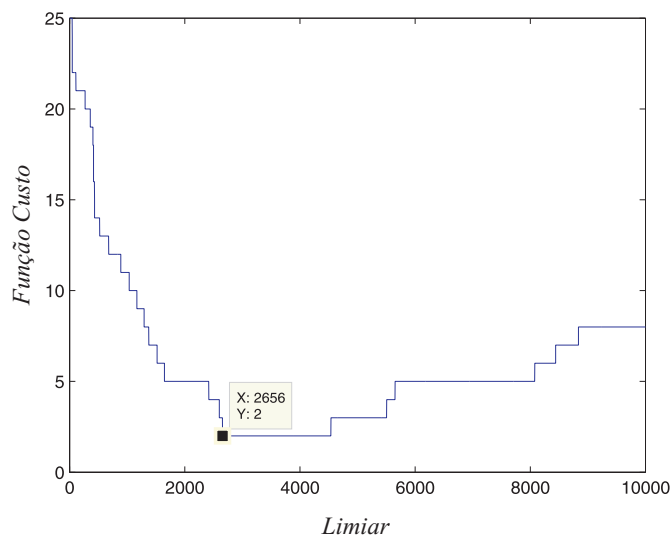
Seguindo o procedimento utilizado por Luiz et al. [23] escolhe-se o menor valor do limiar que minimize a função custo. Esse procedimento pode ser observado na Figura 5.7.

### 5.3 Obtenção da política do tempo ótimo de desligamento

Para resolver o *problema de otimização* (PO) proposto em (4.22) é preciso encontrar os tempos ótimos de desligamentos  $T$  que minimizam o consumo de energia, mantendo o consumo de energia inferior a  $\mathcal{E}_{on}$  para uma dada restrição de penalidade de desempenho, onde  $L$  é a penalidade de desempenho máxima aceitável.

Os valores de  $P_{on}$ ,  $P_{off}$ ,  $E_s$ ,  $T_{sd}$  e  $T_{vu}$  que serão utilizados para obtenção da política de energia são os mesmos fornecidos na Seção 5.1. Como discutido na Seção 4.3.4, é preciso construir a tabela de políticas (*look-up table*) [7], para que o gerenciador de energia tome decisões rápidas em tempo de execução com gasto de energia desprezível. Para construir a tabela, o problema de otimização foi resolvido *off-line* para diversas cargas de trabalho com o parâmetro  $q = (0, 1, 0, 2, 0, 3, 0, 4, 0, 5, 0, 6, 0, 7, 0, 8, 0, 9)$ .

Figura 5.7: Função custo para a seleção do limiar de ociosidade da interface de rede.



Nas Figuras 5.8, 5.9 e 5.10 são apresentados resultados analíticos e resultados de simulação com a política proposta para cargas de trabalho estacionárias. Para obter os resultados analíticos foram utilizadas as expressões de valor esperado de energia (4.20) e valor esperado de penalidade de desempenho (4.21). Para os resultados de simulação foi utilizado o esquema apresentado no diagrama de blocos da Figura 4.8, onde as durações de cada carga de trabalho foram definidas como discutido no Projeto de Experimento na Seção 4.4.

Na Figura 5.8, o problema de otimização é resolvido para uma carga de trabalho leve,  $q = 0, 1$ . À medida que o  $T$  cresce, o valor esperado da energia diminui e a penalidade de desempenho aumenta. Isso acontece porque quanto maior o  $T$ , maior o tempo em que o dispositivo permanecerá desligado. Consequentemente, maior a penalidade, pois quando a estimativa extrapola o valor do intervalo ocioso, o religamento preditivo não ocorre.

Na Figura 5.9 o problema de otimização é resolvido para a carga de trabalho intermediária, ( $q = 0, 4$ ). Os resultados para essa carga são significativamente diferentes que para uma carga de trabalho leve como a ( $q = 0, 1$ ). Nessa carga de trabalho os intervalos ociosos são mais curtos, e consequentemente, haverá poucos valores de  $T$  que vão minimizar o consumo de energia.

Na Figura 5.10, o problema de otimização é resolvido para uma carga de trabalho intensa, ( $q = 0, 6$ ). Observa-se que poucos valores de  $T$  conseguem reduzir o consumo de energia. Isso acontece porque essa carga de trabalho é mais intensa que as anteriores, logo, os intervalos



Figura 5.8: Resultados de energia e penalidade de desempenho para a carga de trabalho leve, ( $q = 0,1$ ).

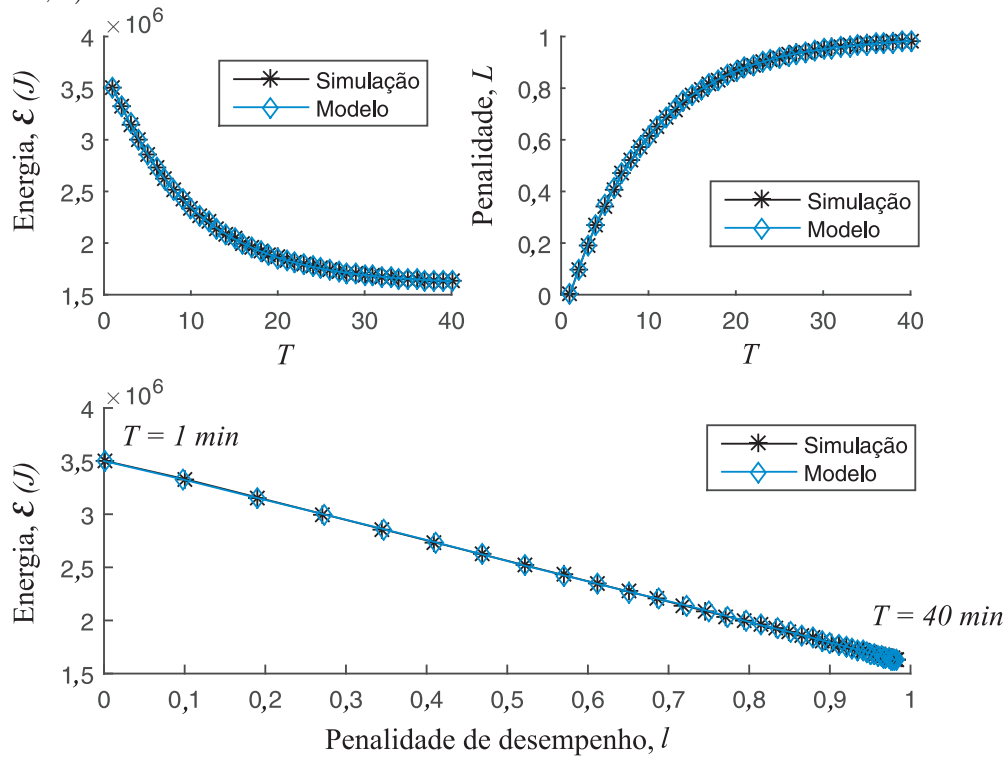
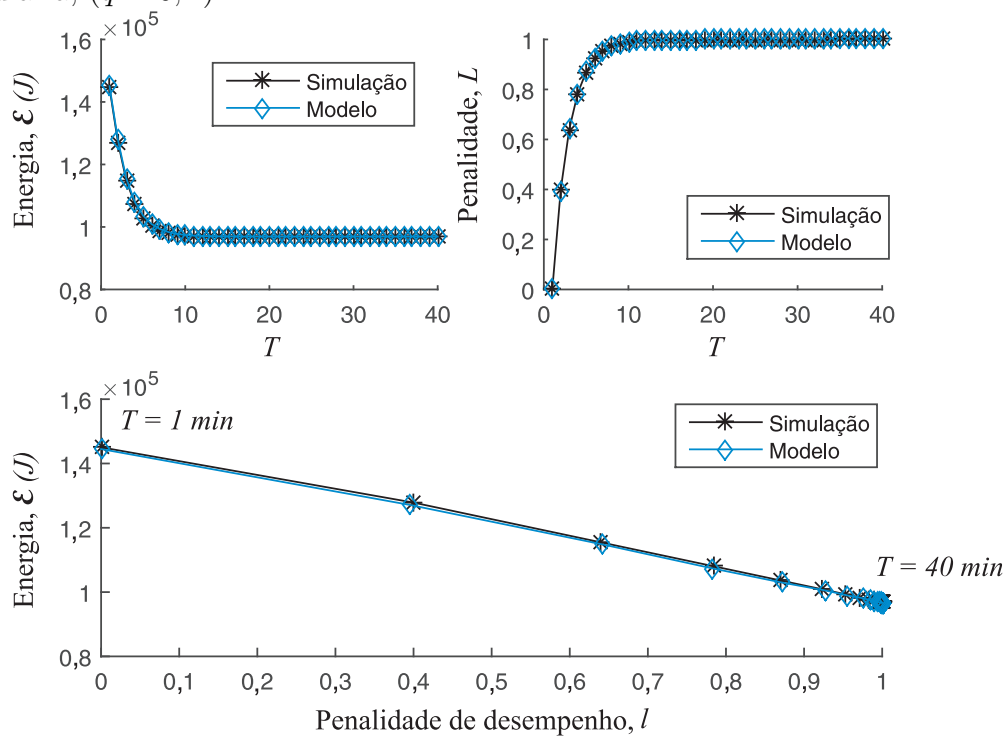


Figura 5.9: Resultados de energia e penalidade de desempenho para a carga de trabalho intermediária, ( $q = 0,4$ ).





Em tempo de execução, num dado período de amostragem, uma estimativa da carga de trabalho,  $\hat{q}$ , é obtida, e de acordo com a restrição de penalidade adotada, uma técnica de interpolação linear é aplicada para que um  $T$  ótimo seja escolhido, como apresentado na Seção 4.3.4. Simulações foram realizadas para comparar o desempenho da política proposta, *política do tempo ótimo de desligamento*, TOD, com relação à política de *timeout* [11] e a política preditiva recursiva [27].

## 5.4 Comparação de políticas

O comportamento das políticas foi simulado para cargas de trabalho leves, intermediárias e de uso intensivo. Para determinar o modelo da carga de trabalho ao longo do tempo, foi utilizada a técnica de janelas deslizantes de múltiplos comprimentos, JMC, apresentada na Seção 3.2.2. Através de simulações também é avaliado o impacto da estimação da carga de trabalho para a política proposta.

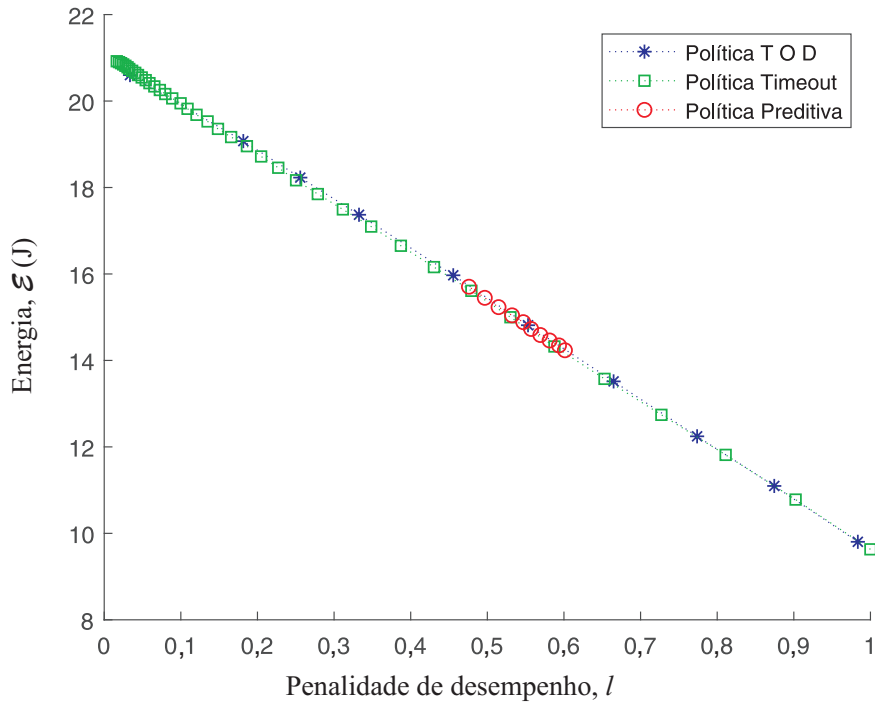
### 5.4.1 Comparações no cenário estacionário

A duração das cargas de trabalho utilizadas nas simulações apresentadas nessa seção seguiu o procedimento descrito na Seção 4.4. A política de *timeout* foi simulada com  $\delta = (1, 2, 3, 4, \dots, 40)$ , e a política preditiva com o parâmetro  $a = (0, 1, 0, 2, 0, 3, \dots, 1)$ . Para identificar o modelo de cada carga de trabalho foi utilizada a técnica de janelas deslizantes de múltiplos comprimentos, JMC, abordada na Seção 3.2.2.

Na Figura 5.11, estão apresentados os resultados de simulação para uma carga de trabalho estacionária ( $q = 0, 1$ ). Essa carga é considerada *leve*, caracterizada por longos intervalos ociosos. Para esse tipo de carga de trabalho, as três políticas apresentaram valores de consumo de energia muito semelhantes. A política preditiva apresenta um intervalo pequeno de variação  $(\mathcal{E}(a), l)$  quando comparada às demais políticas. De forma que, com essa política não seria possível alcançar menores valores no consumo de energia mesmo que a penalidade de desempenho não fosse uma restrição. A política proposta e a política de *timeout* apresentaram comportamento muito semelhante para essa carga de trabalho. Esse comportamento é coerente, pois quando verifica-se em uma carga de trabalho intervalos ociosos muito lon-

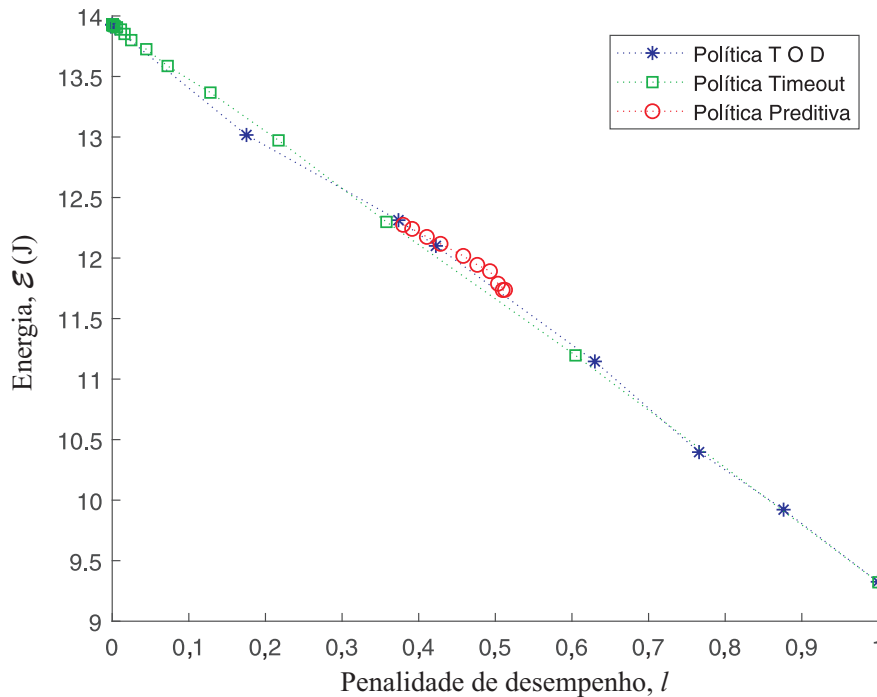
gos, as penalidades de desempenho ficam mais espaçadas no tempo, e analisando do ponto de vista do incômodo do usuário, essas penalidades mais distantes no tempo causam um incômodo menor [5].

Figura 5.11: Curva da Energia *versus* Penalidade de desempenho para a carga de trabalho ( $q = 0,1$ ).



Na Figura 5.12, estão apresentados os resultados de simulação para uma carga de trabalho estacionária ( $q = 0,4$ ). Essa carga já não pode ser considerada tão *leve* quanto a ( $q = 0,1$ ). Comparando os resultados dessa simulação com os apresentados na Figura 5.11, percebe-se que, embora os valores para economia de energia continuem “alinhados” entre a política de *timeout* e a proposta, é possível obter economia de energia para diferentes penalidades de desempenho utilizando a política proposta. Quanto mais intensa é a carga de trabalho, mais variados são os pontos de atuação da política proposta quando comparada à política de *timeout*. Essa observação ficará mais clara com as próximas simulações. A política preditiva apresenta uma área mais restrita de atuação.

Figura 5.12: Curva da Energia *versus* Penalidade de desempenho para carga de trabalho ( $q = 0,4$ ).

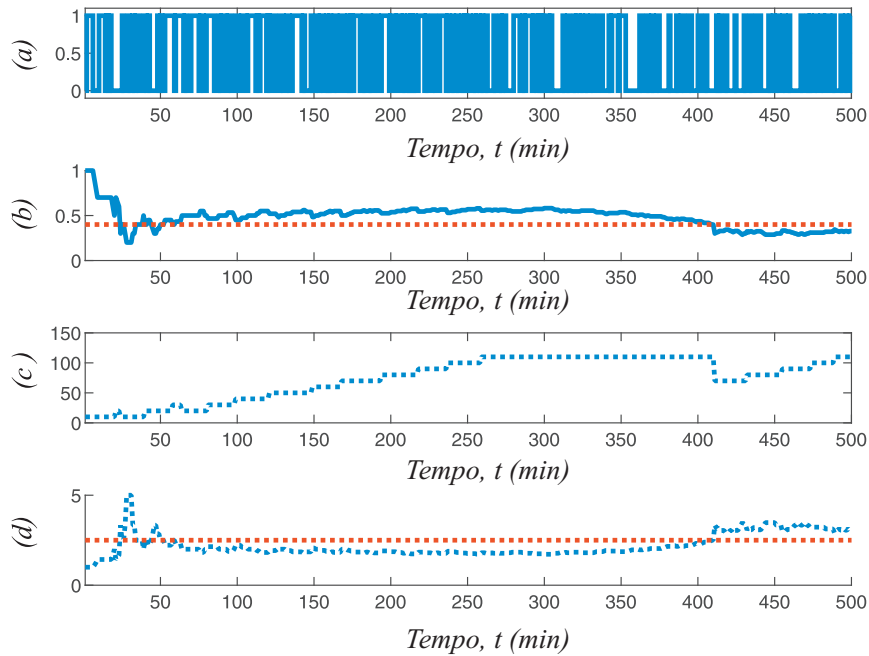


Para a simulação cujos resultados estão apresentados na Figura 5.12, a carga de trabalho é estimada utilizando a técnica JMC. Na Figura 5.13 (a) está representada a carga de trabalho com ( $q = 0,4$ ), na Figura 5.13 (b) é possível observar as estimativas para essa carga ao longo do tempo. Na Figura 5.13 (c) estão representados os comprimentos de janelas utilizados e por fim, na Figura 5.13 (d), os valores estimados para o intervalo ocioso,  $1/\hat{q}$ .

Na Figura 5.14, estão apresentados os resultados de simulação para uma carga de trabalho estacionária ( $q = 0,6$ ). Como dito anteriormente, quanto mais intensa é a carga de trabalho mais “limitada” é a atuação da política de *timeout*. Para essa carga de trabalho, para restrições de penalidade de desempenho ( $0,5 < l \leq 1$ ), a política proposta apresenta quatro valores possíveis de economia de energia. Em contrapartida, com a política de *timeout* só seria possível obter redução no consumo aceitando-se o caso de penalidade de desempenho máxima.

À medida que os intervalos ociosos ficam mais curtos, melhor é o desempenho de uma política que utiliza do conhecimento do “modelo” da carga de trabalho e realiza religamen-

Figura 5.13: Estimação da carga de trabalho: (a) carga de trabalho com ( $q = 0,4$ ); (b)  $\hat{q}$  estimado; (c) comprimentos de janelas utilizados; (d) valor estimado do intervalo ocioso  $1/\hat{q}$ .



tos preditivos. Em contrapartida, a política *preditiva* utilizada para fins de comparação se mantém em uma área mais restrita de atuação, e não apresenta vantagens do ponto de vista de consumo de energia quando em comparação às demais. Uma possível justificativa é que essa política não consegue modelar tão bem a carga de trabalho, além de não realizar os desligamentos possíveis, durante os intervalos ociosos grandes o suficiente para reduzir o consumo. A política preditiva eventualmente realiza desligamentos em intervalos excessivamente curtos.

Para a simulação cujos resultados estão apresentados na Figura 5.14, a carga de trabalho é estimada utilizando a técnica JMC. Na Figura 5.15 (a) está representada a carga de trabalho com ( $q = 0,6$ ). Na Figura 5.15 (b) é possível observar as estimativas para essa carga ao longo do tempo. Na Figura 5.15 (c) estão representados os comprimentos de janelas utilizados e por fim, na Figura 5.15 (d), os valores estimados para o intervalo ocioso,  $1/\hat{q}$ .

Quanto mais intensa a carga de trabalho é, melhor é o desempenho da política proposta. O que acontece nesses casos são intervalos ociosos cada vez mais curtos. Dessa forma, co-

Figura 5.14: Curva da Energia *versus* Penalidade de desempenho para carga de trabalho ( $q = 0,6$ ).

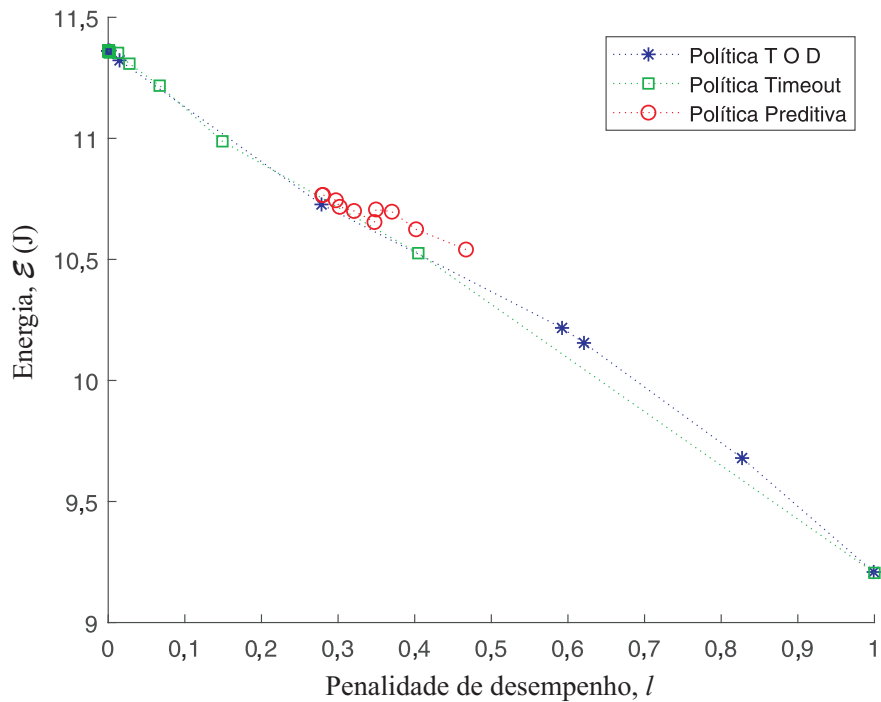
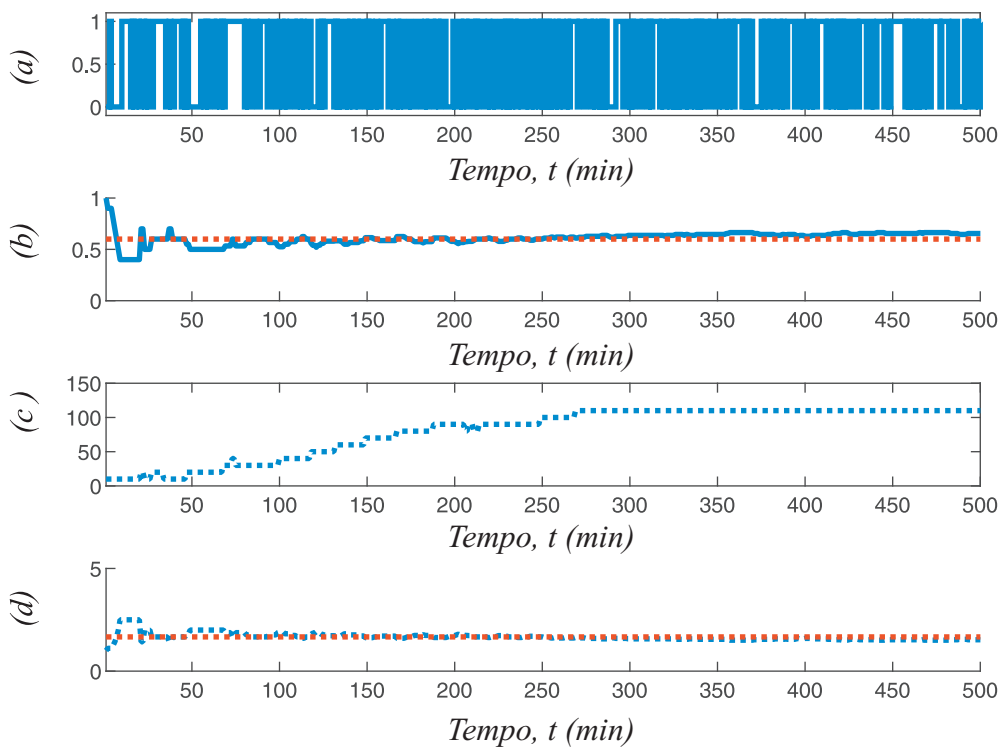


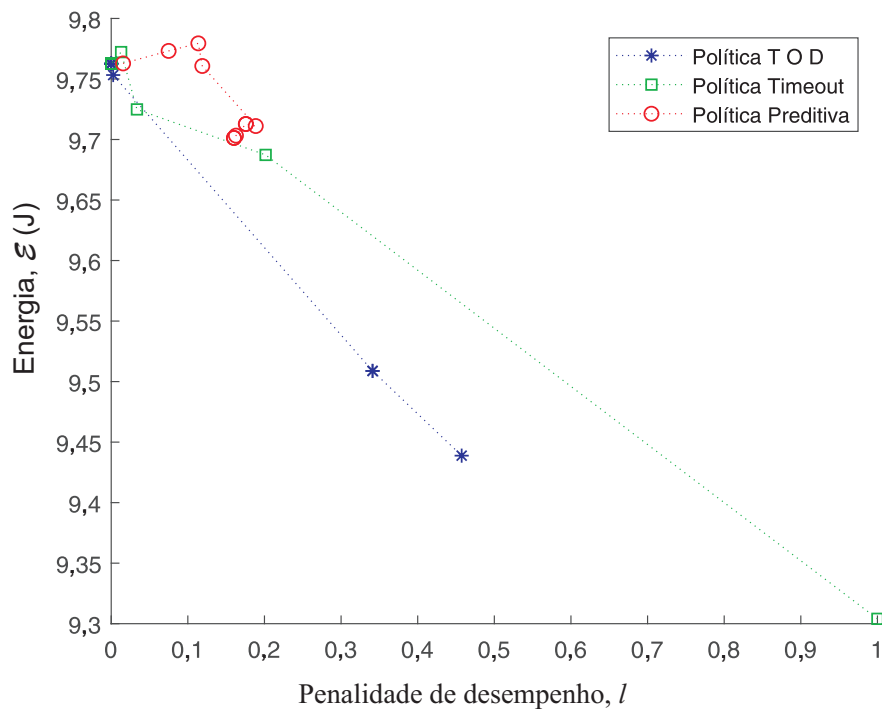
Figura 5.15: Estimação da carga de trabalho: (a) carga de trabalho com ( $q = 0,6$ ); (b)  $\hat{q}$  estimado; (c) comprimentos de janelas utilizados; (d) valor estimado do intervalo ocioso  $1/\hat{q}$ .



nhecer o comportamento da carga de trabalho permite que ocorram mais religamentos de fato preditivos, isto é, religamentos que acontecem antes do término do intervalo ocioso, e também evita que ocorram desligamentos desnecessários, ou seja, desligamentos que acontecem durante intervalos ociosos muito curtos, principalmente quando esses intervalos tem a mesma duração do intervalo de amostragem, como discutido na Seção 4.2.1.

Na Figura 5.16, estão apresentados os resultados de simulação para uma carga de trabalho estacionária ( $\hat{q} = 0,8$ ). Essa é uma carga de trabalho de uso intensivo. Como a política proposta utiliza um modelo acurado da carga de trabalho, na construção da *tabela de políticas* ficou definido que para esse modelo de carga, ( $q = 0,8$ ), poucos valores do tempo ótimo de desligamento,  $T$ , poderão ser utilizados para alcançar redução no consumo de energia. Já a política de *timeout* realiza desligamentos em todos os intervalos ociosos, ( $\delta = 1$ ) e com isso gera uma penalidade de desempenho em todo desligamento. A política preditiva utilizada para fins de comparação chega a consumir mais que  $\mathcal{E}_{on}$  e se distancia da economia de energia para cargas de trabalho de uso intensivo.

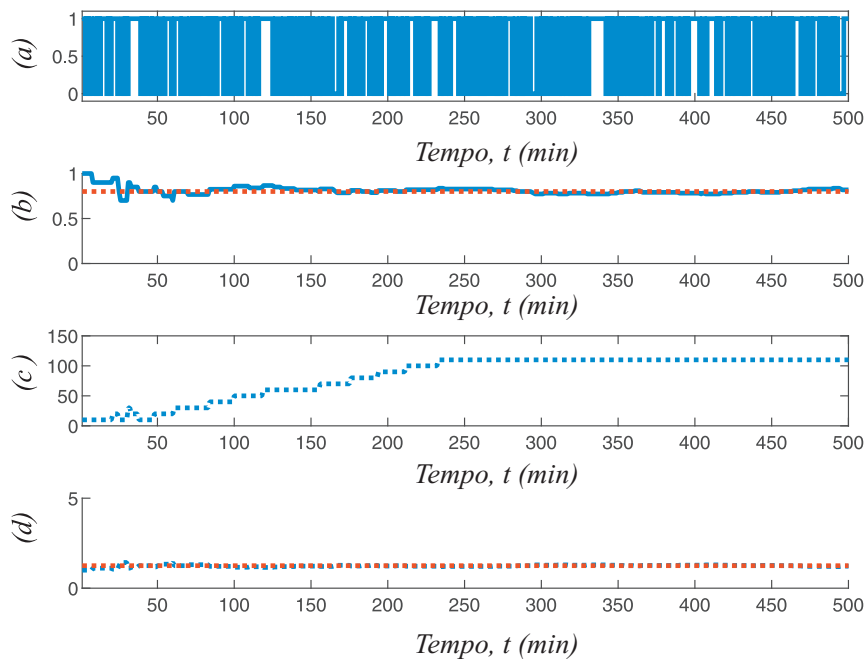
Figura 5.16: Curva da Energia *versus* Penalidade de desempenho para carga de trabalho ( $q = 0,8$ ).





Para a simulação cujos resultados estão apresentados na Figura 5.16, a carga de trabalho é estimada utilizando a técnica JMC. Na Figura 5.17 (a) está representada a carga de trabalho com ( $q = 0,8$ ), na Figura 5.17 (b) é possível observar as estimativas para essa carga ao longo do tempo. Na Figura 5.17 (c) estão representados os comprimentos de janelas utilizados e por fim, na Figura 5.17 (d), os valores estimados para o intervalo ocioso,  $1/\hat{q}$ .

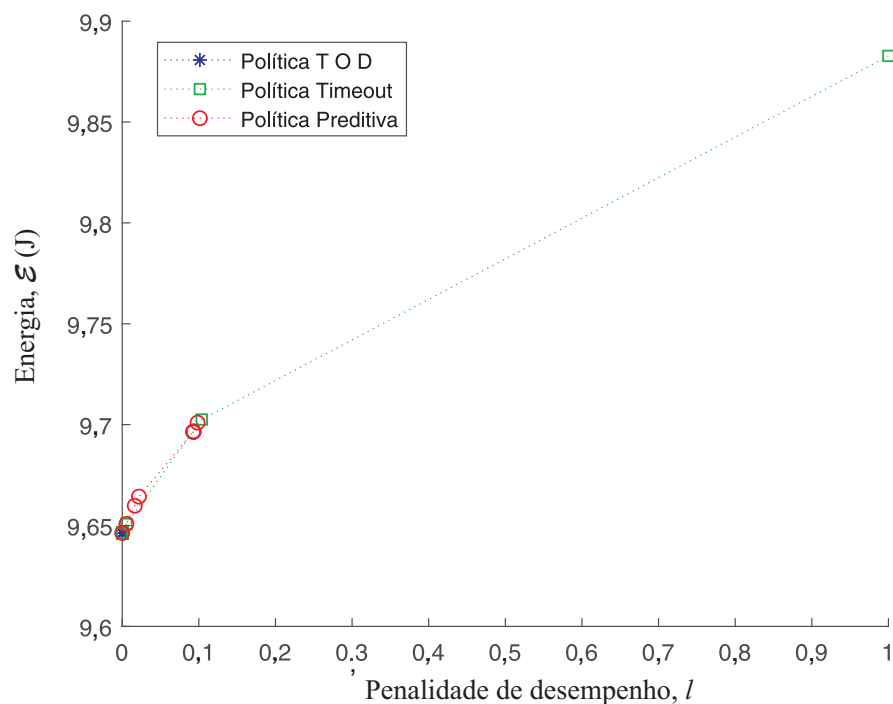
Figura 5.17: Estimação da carga de trabalho: (a) carga de trabalho com ( $q = 0,8$ ); (b)  $\hat{q}$  estimado; (c) comprimentos de janelas utilizados; (d) valor estimado do intervalo ocioso  $1/\hat{q}$ .



Na Figura 5.18 estão apresentados os resultados de simulação para o caso mais crítico para se obter redução no consumo de energia, dentre as cargas de trabalho abordadas nessa seção. A carga de trabalho ( $\hat{q} = 0,9$ ) é uma carga de uso muito intensivo. O valor esperado para o intervalo ocioso é muito próximo da duração do intervalo de amostragem. De forma que, para uma política que atua sem o conhecimento do comportamento da carga de trabalho, *timeout*, ou uma política que não utiliza um modelo tão acurado da mesma, *predictiva*, serão realizados muitos desligamentos incorretos que, ao invés de contribuir para redução no consumo de energia, irão aumentar o consumo. Para uma carga de trabalho tão

intensa como essa, utilizando a política *predictiva* e a política de *timeout* o consumo pode ser maior do que se não houvesse nenhuma política atuando. Para esse tipo de carga de trabalho, a política proposta oferece um grande diferencial, pois, ao identificar o modelo da carga de trabalho, nenhum desligamento deverá ocorrer.

Figura 5.18: Curva da Energia *versus* Penalidade de desempenho para carga de trabalho ( $q = 0,9$ ).

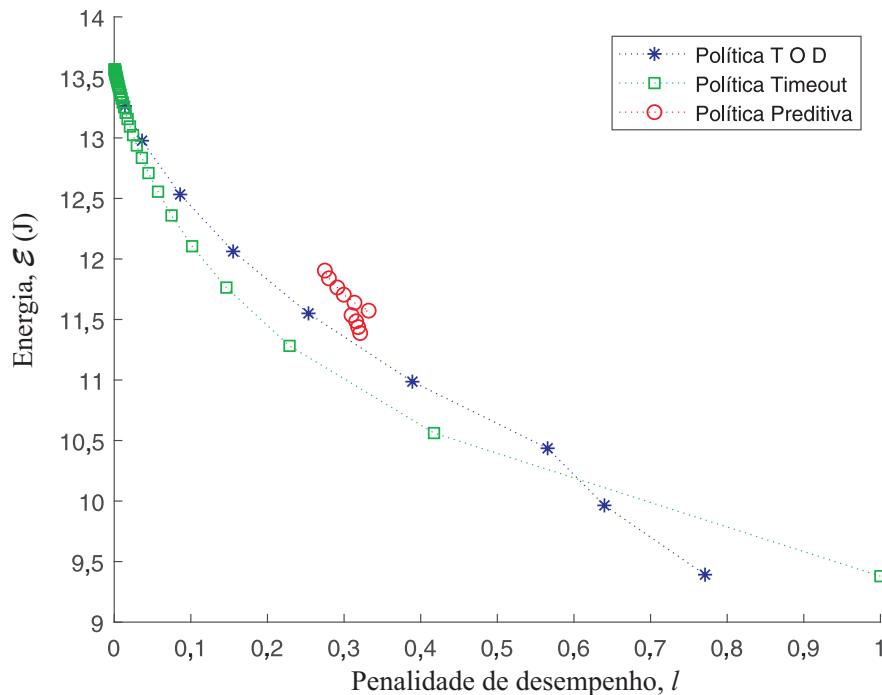


### 5.4.2 Comparações no cenário não-estacionário

Para verificar o funcionamento da política proposta no cenário não-estacionário, cargas de trabalho estacionárias foram concatenadas para gerar cargas de trabalho não-estacionárias, como descrito na Seção 3.2. Algumas combinações foram feitas, para verificar o desempenho da técnica proposta quanto ao *consumo de energia* e *penalidade de desempenho*, quando comparada às outras duas políticas: política de *timeout* e política *predictiva*. Para fins de comparação, optou-se por utilizar amostras de mesmo tamanho, isto é, utilizando (4.44) determinou-se a duração da amostra para garantir que os resultados estejam dentro do intervalo de confiança. Cada carga de trabalho tem um  $N$  específico. Para generalizar as durações das amostras, optou-se por utilizar o maior  $N$  para todas as cargas de trabalho.

Na Figura 5.19 estão apresentados os resultados quando as políticas foram comparadas para a carga de trabalho (5.2). O diferencial dessa composição de carga de trabalho é que ela apresenta uma grande variação de utilização, iniciando com uma carga de uso leve, ( $q = 0,1$ ), e terminando com uma carga de uso intensivo, ( $q = 0,9$ ). Essa simulação levanta uma suspeita que ficará mais clara na continuação dessa seção. Cargas de trabalho consideradas de uso *leve* ( $q = 0,1$ ) apresentam longos intervalos ociosos, e por isso, políticas com religamento preditivo podem não apresentar um ganho significativo na economia de energia. Uma justificativa para isso é que em intervalos muito longos, eventualmente ocorrem religamentos excessivamente preditivos, e em políticas como a de *timeout* os religamentos são realizados pelo usuários, todos esses apresentados na Seção 4.2.1. Como nesse tipo de carga de trabalho as penalidades de desempenho ficam mais espaçadas no tempo, políticas preditivas acabam não apresentando seu melhor desempenho.

Figura 5.19: Curva da Energia *versus* Penalidade de desempenho para a carga de trabalho não-estacionária (5.2).



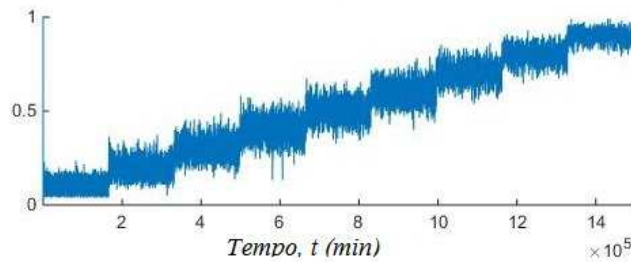
A política proposta não apresenta menores valores de consumo de energia para penalidades menores que ( $l \leq 0,5$ ). Entretanto, com a política proposta é possível obter a economia máxima de energia com penalidade ( $0,5 < l \leq 0,8$ ), enquanto que com a política de *timeout*

só seria possível obter esse mesmo consumo para uma penalidade ( $l = 1$ ).

$$q = \begin{cases} 0,1, & \text{se } 0 \leq t \leq Nmin \\ 0,2, & \text{se } (N + 1)min < t \leq (2N)min \\ 0,3, & \text{se } (2N + 1)min < t \leq (3N)min \\ 0,4, & \text{se } (3N + 1)min < t \leq (4N)min \\ 0,5, & \text{se } (4N + 1)min < t \leq (5N)min \\ 0,6, & \text{se } (5N + 1)min < t \leq (6N)min \\ 0,7, & \text{se } (6N + 1)min < t \leq (7N)min \\ 0,8, & \text{se } (7N + 1)min < t \leq (8N)min \\ 0,9, & \text{se } (8N + 1)min < t \leq (9N)min \end{cases} \quad (5.2)$$

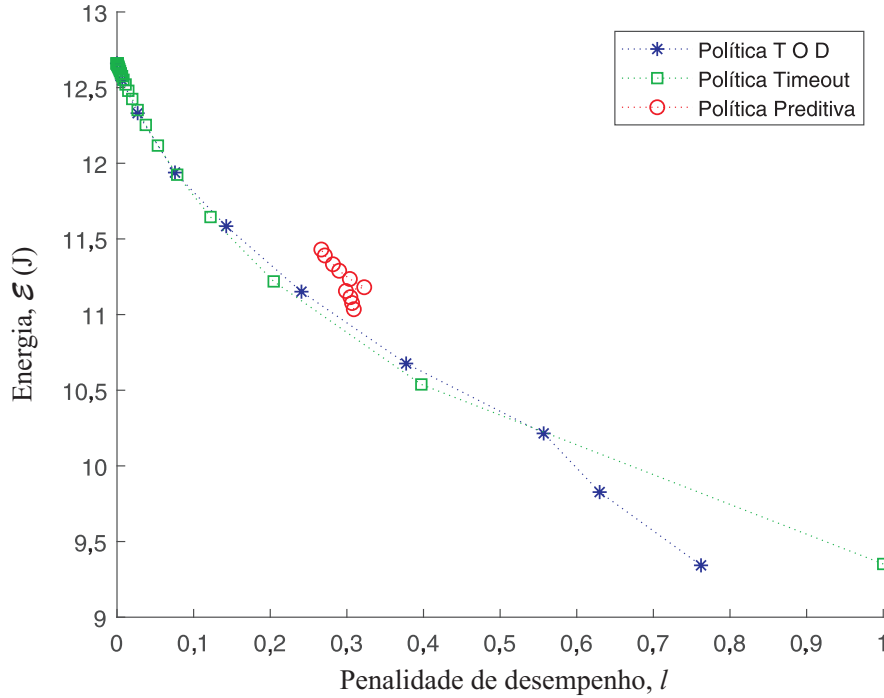
Na Figura 5.20 estão apresentados as estimativas para a carga de trabalho (5.2) ao longo do tempo.

Figura 5.20: Estimativas para a carga de trabalho (5.2) utilizando a técnica JMC.



Na Figura 5.21 estão apresentados os resultados quando as políticas são comparadas para a carga de trabalho (5.3). Essa é uma carga muito semelhante à (5.2), a principal diferença é a retirada da carga de uso leve, ( $q = 0,1$ ). Com essa alteração, observa-se que para penalidades menores que 0,5, a política proposta apresenta resultados mais próximos da política de timeout. Para penalidade maiores que 0,5, é possível obter a economia máxima de energia com uma penalidade menor. Além disso, verifica-se mais pontos de atuação utilizando a política proposta, e que não seriam possíveis utilizando a política de *timeout*. A política preditiva embora apresente penalidades baixas, não consegue reduzir tanto o consumo quanto as demais e também não oferece uma variação ao longo do eixo da penalidade. De forma que com essa política não é possível obter menores valores para o consumo de energia mesmo que uma penalidade maior fosse aceita.

Figura 5.21: Curva da Energia *versus* Penalidade de desempenho para a carga de trabalho não-estacionária (5.3).

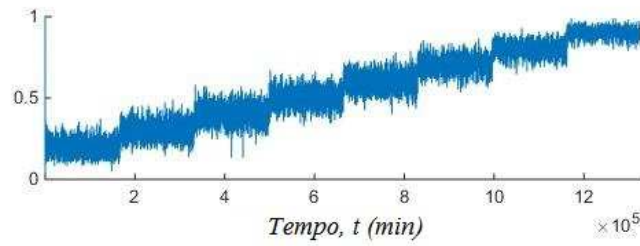


$$q = \begin{cases} 0,2, & \text{se } 0 \leq t \leq Nmin \\ 0,3, & \text{se } (N + 1)min < t \leq (2N)min \\ 0,4, & \text{se } (2N + 1)min < t \leq (3N)min \\ 0,5, & \text{se } (3N + 1)min < t \leq (4N)min \\ 0,6, & \text{se } (4N + 1)min < t \leq (5N)min \\ 0,7, & \text{se } (5N + 1)min < t \leq (6N)min \\ 0,8, & \text{se } (6N + 1)min < t \leq (7N)min \\ 0,9, & \text{se } (7N + 1)min < t \leq (8N)min \end{cases} \quad (5.3)$$

Na Figura 5.22 estão apresentados as estimativas para a carga de trabalho (5.3) ao longo do tempo.

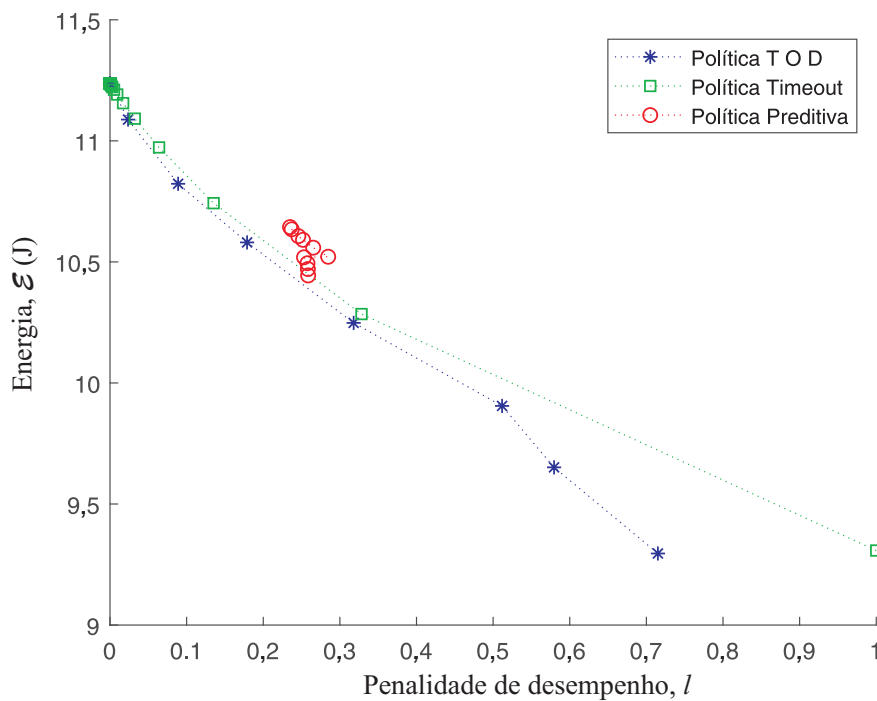
Na Figura 5.23 estão apresentados os resultados quando as políticas são comparadas para a carga de trabalho (5.4). A diferença entre essa carga de trabalho e a (5.2) foi a retirada das cargas ( $q = 0,1$ ), ( $q = 0,2$ ) e ( $q = 0,3$ ), que são cargas mais leves. Nessa simulação fica mais evidente que para cargas de trabalho de medianas para uso intensivo,

Figura 5.22: Estimativas para a carga de trabalho (5.3) utilizando a técnica JMC.



utilizando a política proposta é possível obter menores valores de consumo de energia com penalidades menores, quando comparada com as políticas de *timeout* e *predictiva*.

Figura 5.23: Curva da Energia *versus* Penalidade de desempenho para a carga de trabalho não-estacionária (5.4).

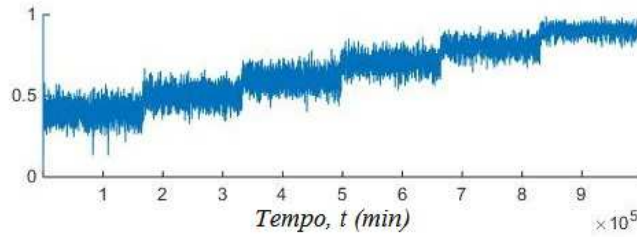


Observa-se também que, com a política proposta, o menor valor de consumo de energia é alcançado para uma penalidade bem menor, ( $l \approx 0,7$ ), enquanto que com a política de *timeout* essa penalidade seria máxima, ( $l = 1$ ).

$$q = \begin{cases} 0,4, & \text{se } 0 \leq t \leq Nmin \\ 0,5, & \text{se } (N + 1)min < t \leq (2N)min \\ 0,6, & \text{se } (2N + 1)min < t \leq (3N)min \\ 0,7, & \text{se } (3N + 1)min < t \leq (4N)min \\ 0,8, & \text{se } (4N + 1)min < t \leq (5N)min \\ 0,9, & \text{se } (5N + 1)min < t \leq (6N)min \end{cases} \quad (5.4)$$

Na Figura 5.24 estão apresentados as estimativas para a carga de trabalho (5.4) ao longo do tempo.

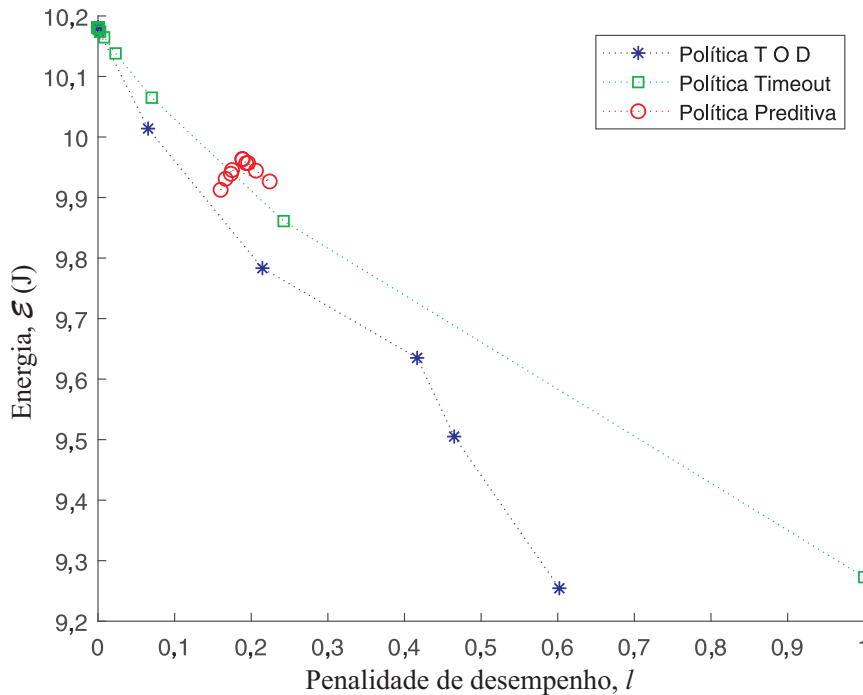
Figura 5.24: Estimativas para a carga de trabalho (5.4) utilizando a técnica JMC.



Na Figura 5.25 estão apresentados os resultados quando as políticas são comparadas para a carga de trabalho (5.5). Essa é uma carga mais intensa que as apresentadas anteriormente. Para essa carga de trabalho, a política proposta apresenta ainda mais vantagens. Do ponto de vista da economia de energia, a política proposta alcança menores valores de consumo de energia em todas as situações. Quanto à penalidade de desempenho, além de causar menos penalidades, observa-se também que com a política proposta o menor valor de consumo de energia é alcançado para uma penalidade bem menor, ( $l \approx 0,6$ ), enquanto que com a política de *timeout* essa penalidade seria máxima, ( $l = 1$ ).

$$q = \begin{cases} 0,6, & \text{se } 0 \leq t \leq Nmin \\ 0,7, & \text{se } (N + 1)min < t \leq (2N)min \\ 0,8, & \text{se } (2N + 1)min < t \leq (3N)min \\ 0,9, & \text{se } (3N + 1)min < t \leq (4N)min \end{cases} \quad (5.5)$$

Figura 5.25: Curva da Energia *versus* Penalidade de desempenho para a carga de trabalho não-estacionária (5.5).

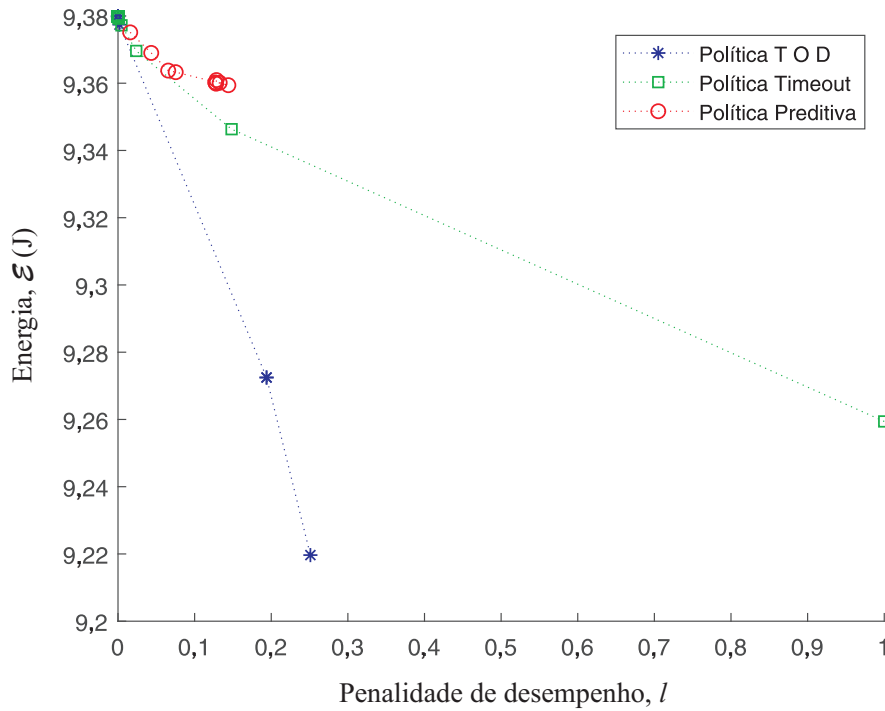


Na Figura 5.26 estão apresentados os resultados quando as políticas são comparadas para a carga de trabalho (5.6). Essa é uma carga de trabalho não-estacionária de uso intensivo. Os intervalos ociosos são muito curtos, de forma que, para a carga de trabalho ( $q = 0,9$ ), esses intervalos se aproximam da duração do intervalo de amostragem,  $h$ . De forma que é comum que, quando o gerenciador de energia identifica um intervalo ocioso, o requisitante de serviços já esteja tendo seu estado alterado para intervalo ativo, e o gerenciador de energia, com uma política de *timeout*, acabe por desligar erroneamente o dispositivo. Por isso, observa-se nos resultados dessa simulação que a política de *timeout* não alcança o consumo mínimo de energia mesmo para uma penalidade ( $l = 1$ ).

A política proposta apresenta um grande diferencial para cargas de trabalho de uso intensivo. Essa política é construída *off-line*, e quando o problema de otimização é resolvido para carga de trabalho ( $q = 0,9$ ), observa-se que, para nenhum valor do tempo ótimo de desligamento,  $T$ , o consumo de energia será menor que  $\mathcal{E}_{on}$ . Por isso, verifica-se na simulação apresentada na Figura 5.26, que a política proposta alcança menores valores de consumo de energia e com uma penalidade significativamente inferior, ( $l \leq 0,3$ ), quando comparada



Figura 5.26: Curva da Energia *versus* Penalidade de desempenho para a carga de trabalho não-estacionária (5.6).

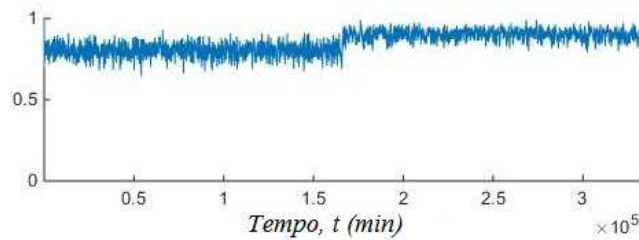


com a política de *timeout*, ( $l = 1$ ).

$$q = \begin{cases} 0,8, & \text{se } 0 \leq t \leq Nmin \\ 0,9, & \text{se } (N + 1)min < t \leq (2N)min \end{cases} \quad (5.6)$$

Na Figura 5.27 estão apresentados as estimativas para a carga de trabalho (5.6) ao longo do tempo.

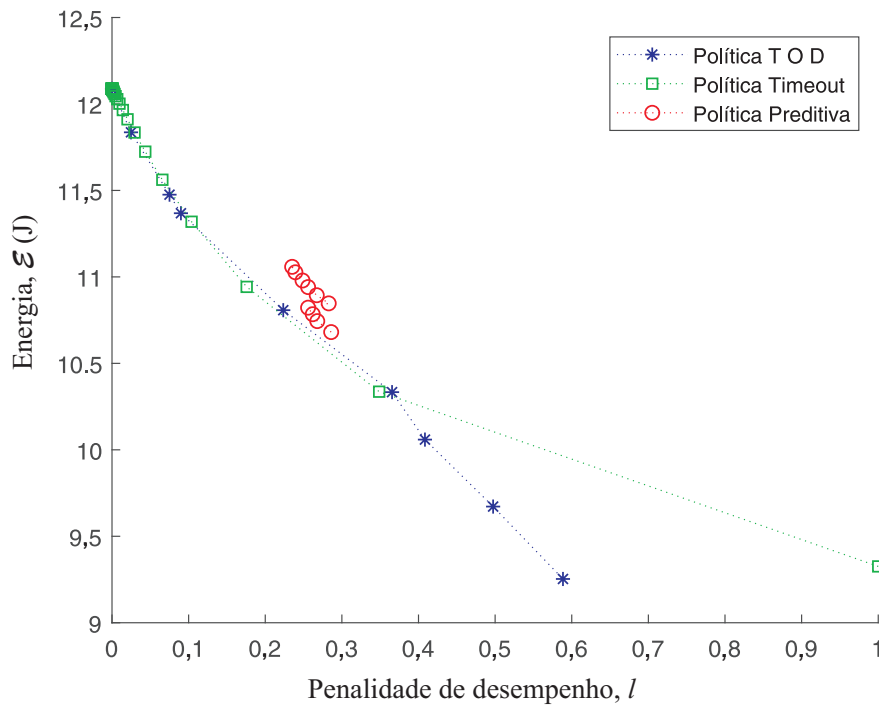
Figura 5.27: Estimativas para a carga de trabalho (5.6) utilizando a técnica JMC.



Foram realizadas simulações com variações bruscas de cargas de trabalho, (5.7), para verificar o comportamento da política proposta diante de uma carga de trabalho não-

estacionária que inicia com uma carga mediana, passa por uma carga de uso leve, e altera rapidamente para uma carga de uso intensivo. Os resultados apresentados na Figura 5.28 mostram que, para  $(l \leq 0,4)$  as políticas de *timeout* e a política proposta apresentam um comportamento semelhante, o que é justificável pela influência da carga de trabalho leve. Para  $(l \geq 0,4)$  a política proposta apresenta um melhor desempenho que a política de *timeout*. Mais uma vez a política proposta alcança menores valores de consumo de energia e com uma penalidade significativamente inferior,  $(l \approx 0,6)$ , quando comparada com a política de *timeout*,  $(l = 1)$ .

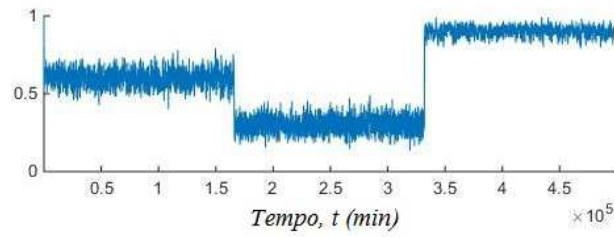
Figura 5.28: Curva da Energia *versus* Penalidade de desempenho para a carga de trabalho não-estacionária (5.7).



$$q = \begin{cases} 0,6, & \text{se } 0 \leq t \leq Nmin \\ 0,3, & \text{se } (N + 1)min < t \leq (2N)min \\ 0,9, & \text{se } (2N + 1)min < t \leq (3N)min \end{cases} \quad (5.7)$$

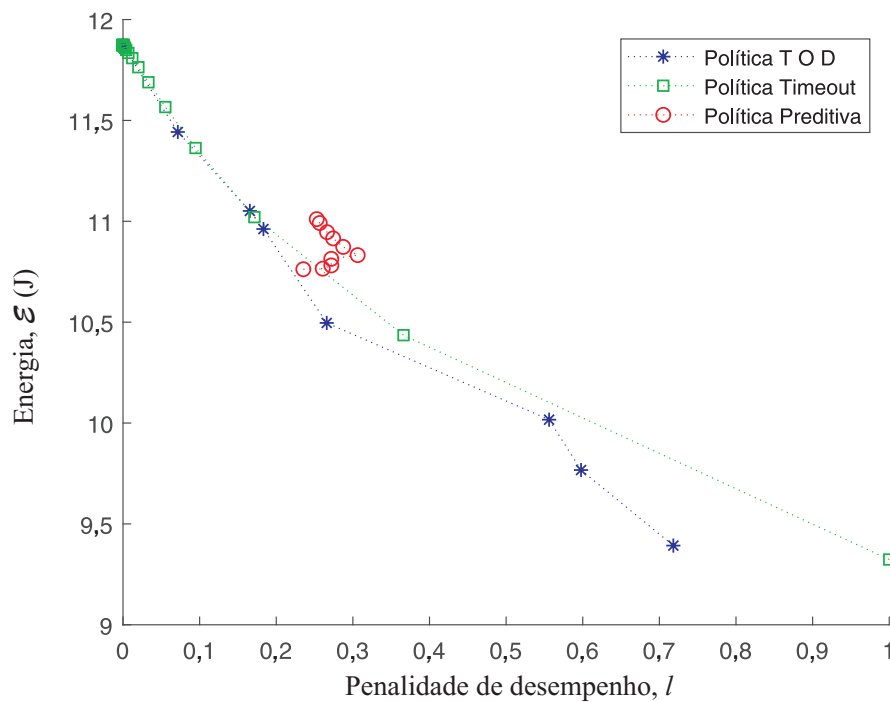
Na Figura 5.29 estão apresentadas as estimativas para a carga de trabalho (5.7) ao longo do tempo.

Figura 5.29: Estimativas para a carga de trabalho (5.7) utilizando a técnica JMC.



Na Figura 5.30 estão apresentados os resultados para uma outra simulação com cargas de trabalho não-estacionárias com variações abruptas. A carga de trabalho utilizada é descrita em (5.8). Para  $(l > 0,2)$  a política proposta apresenta menores valores para o consumo de energia que a política de *timeout*. Além disso, é possível alcançar o consumo mínimo com uma penalidade de desempenho inferior a  $l < 0,8$ , enquanto que utilizando a política de *timeout* essa economia só seria alcançada aceitando uma penalidade de  $l = 1$ .

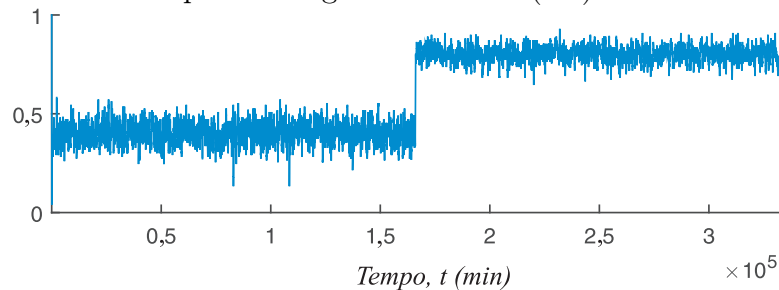
Figura 5.30: Curva da Energia *versus* Penalidade de desempenho para a carga de trabalho não-estacionária (5.8).



$$q = \begin{cases} 0,4, & \text{se } 0 \leq t \leq Nmin \\ 0,8, & \text{se } (N + 1)min < t \leq (2N)min \end{cases} \quad (5.8)$$

Na Figura 5.31 estão apresentados as estimativas para a carga de trabalho (5.8) ao longo do tempo.

Figura 5.31: Estimativas para a carga de trabalho (5.8) utilizando a técnica JMC.



### 5.4.3 Cargas de trabalho experimentais

As cargas de trabalho obtidas a partir da observação de utilização do usuário, ditas cargas “reais”, variam de acordo com a utilização do usuário e também com processos que estejam executando em *background*. Foram realizadas coletas no laboratório Embedded, ao longo de dias distintos, onde cada carga foi coletada por um intervalo de tempo de 60 minutos.

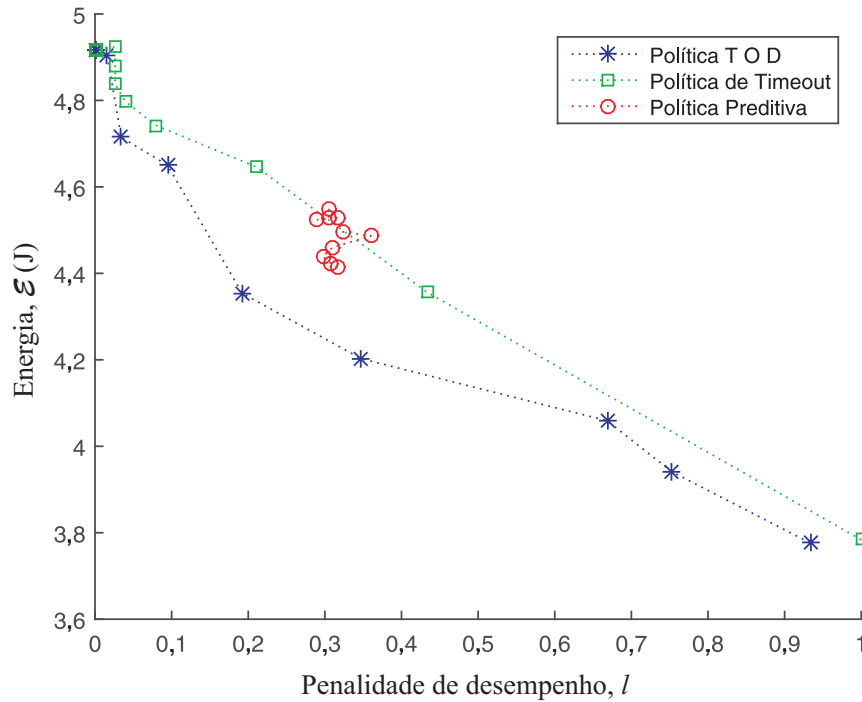
Devido à necessidade de utilizar uma técnica para estimar o comportamento da carga de trabalho, onde nesse trabalho a técnica é a JMC, é preciso que as cargas de trabalho tenham um duração mínima para que as janelas sejam preenchidas e as estimativas se aproximem do valor real da carga de trabalho. Por isso, cada carga de trabalho foi replicada por 5 vezes, totalizando 5 horas de aquisição para cada carga de trabalho. Esse seria o tempo médio de utilização que o brasileiro passa por dia no computador pessoal<sup>2,3</sup>.

Na Figura 5.32 estão apresentados os resultados de simulação para a carga de trabalho (5.9). Pode-se observar que, com a política proposta é possível obter menores valores de consumo de energia e penalidade de desempenho em comparação com a política de *timeout*. Assim como, em todas as outras simulações realizadas nesse trabalho, a política preditiva se concentra em uma área mais restrita.

<sup>2</sup><https://canaltech.com.br/produtos/brasileiro-passa-em-media-53-horas-por-dia-no-computador-pessoal-diz-pesquisa-45737/>

<sup>3</sup><https://noticias.r7.com/tecnologia-e-ciencia/estudo-revela-que-brasileiro-passa-mais-de-nove-horas-por-dia-na-internet-23012015>

Figura 5.32: Curva da Energia *versus* Penalidade de desempenho para carga de trabalho (5.9)

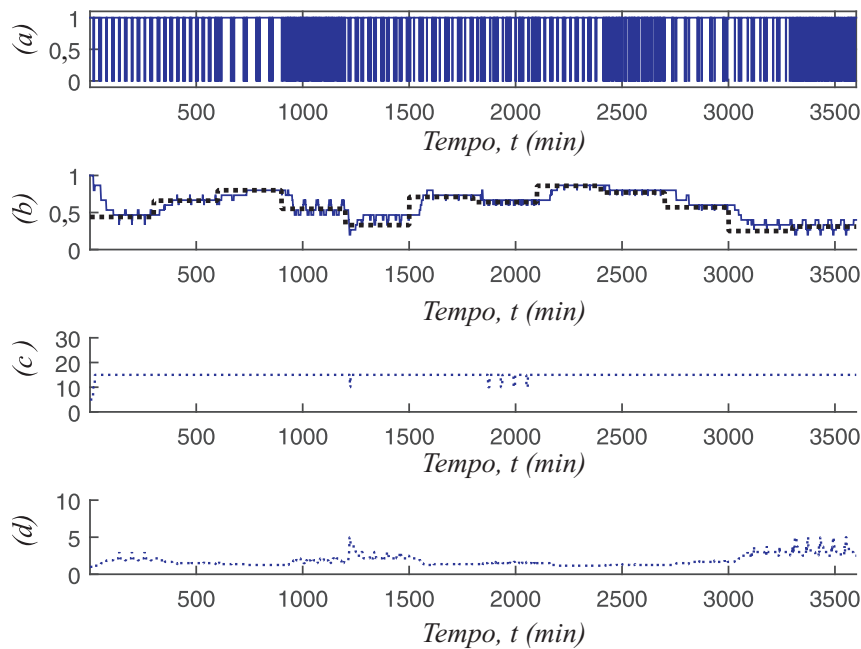


$$q = \begin{cases} 0,44, & \text{se } 0 \leq t \leq 300min \\ 0,66, & \text{se } (300)min < t \leq (600)min \\ 0,8, & \text{se } (600)min < t \leq (900)min \\ 0,55, & \text{se } (900)min < t \leq (1200)min \\ 0,33, & \text{se } (1200)min < t \leq (1500)min \\ 0,71, & \text{se } (1500)min < t \leq (1800)min \\ 0,64, & \text{se } (1800)min < t \leq (2100)min \\ 0,86, & \text{se } (2100)min < t \leq (2400)min \\ 0,77, & \text{se } (2400)min < t \leq (2700)min \\ 0,57, & \text{se } (2700)min < t \leq (3000)min \\ 0,25, & \text{se } (3000)min < t \leq (3300)min \\ 0,31, & \text{se } (3300)min < t \leq (3600)min \end{cases} \quad (5.9)$$

Na Figura 5.33, são apresentados os resultados obtidos utilizando a técnica JMC para identificar o modelo da carga de trabalho, processo necessário para as simulações apresen-

tadas na Figura 5.32.

Figura 5.33: Estimação da carga de trabalho: (a) carga de trabalho não-estacionária; (b)  $q$  estimado; (c) comprimentos de janelas utilizados; (d) valor estimado do intervalo ocioso  $1/\hat{q}$ .



## 5.5 Implementação da política

Nesta seção será apresentado como a técnica proposta pode ser implementada em um sistema computacional. O núcleo do sistema operacional Linux atualiza em tempo de execução arquivos de sistema que são de interesse para o gerenciamento de energia, disponibilizando informações sobre vários parâmetros relacionados à interface de rede em utilização. Por exemplo, no Linux Ubuntu versão 14.10 é possível acessar informações sobre a quantidade de dados recebidos e enviados pelo dispositivo de rede escolhido:

- Quantidade de bytes enviados: `/sys/class/net/wlan0/statistics/tx_bytes`
- Quantidade de bytes recebidos: `/sys/class/net/wlan0/statistics/rx_bytes`

A política de energia apresentada da Seção 5.3 pode ser implementada como uma aplicação em espaço de usuário, i.e., o gerenciador de energia obtém as informações sobre a carga

de trabalho (*taxa de download + taxa de upload*) e a partir das políticas otimizadas pré-calculadas, decide qual comando enviar para o *driver* de Wi-Fi que está sob gerenciamento, como mostrado no diagrama de blocos na Figura 5.34, e no fluxograma da Figura 5.35.

Figura 5.34: Arquitetura do *software* proposto para o gerenciamento dinâmico de energia.

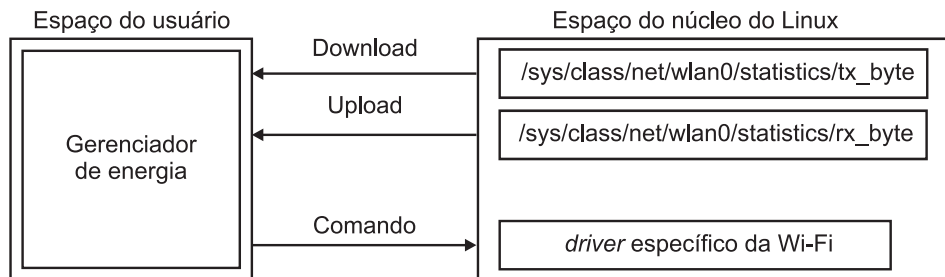
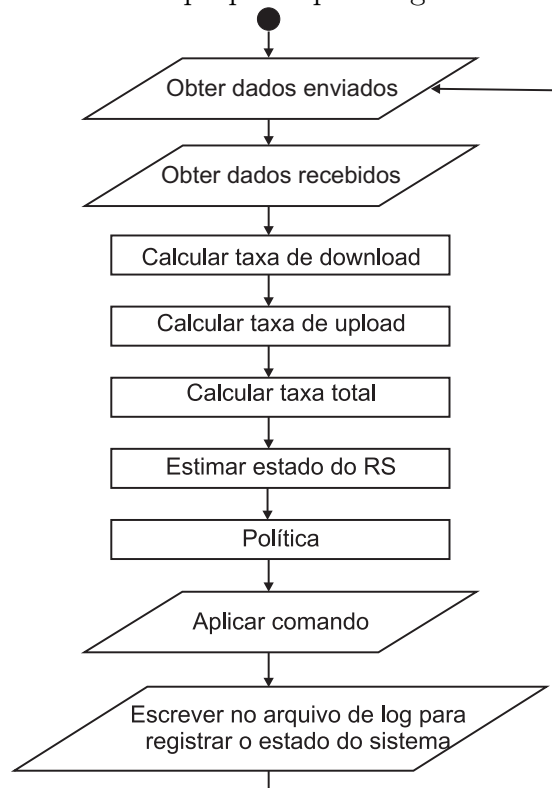


Figura 5.35: Fluxograma da técnica proposta para o gerenciamento dinâmico de energia.



## 5.6 Considerações finais

Nesse Capítulo, a técnica proposta para o gerenciamento dinâmico de energia apresentada no Capítulo 4, denominada Política do Tempo Ótimo de Desligamento, TOD, foi aqui aplicada para um dispositivo de rede, Wi-Fi. Utilizando uma plataforma experimental foi possível medir os valores da corrente e tensão demandados pelo dispositivo, durante intervalos ativos e ociosos, bem como a energia de transição entre estados de consumo, *on* e *off*. Essas informações foram utilizadas para compor o modelo analítico de energia para o dispositivo. O problema de otimização foi formulado e resolvido para cada carga de trabalho e simulações foram feitas para validar a política proposta e também compará-la com outras duas políticas: a política de *timeout* e a *preditiva*.

Foram utilizadas variadas cargas de trabalho. Desde cargas de uso leve até as de uso intensivo. A duração de cada uma dessas cargas foi definida a partir do conceito de intervalo de confiança, apresentado na Seção 4.4. Vários cenários foram compostos para validar a política proposta e definir bem a sua aplicação. Em resultados de simulação, verifica-se que a política proposta apresenta melhores resultados quanto mais intensa é a carga de trabalho. Entretanto, em qualquer cenário criado é possível obter uma economia maior de energia para valores específicos de penalidade de desempenho. Isso já era esperado, visto que a política proposta é baseada na resolução de um problema de otimização, possui um modelo acurado de energia e penalidade de desempenho do sistema e é capaz de identificar variações na carga de trabalho, que para sistemas computacionais de propósito geral, costumam ser inicialmente desconhecidas e variantes no tempo.

Foram obtidas cargas de trabalho a partir da observação de utilização do usuário e concatenadas para compor uma carga de trabalho não-estacionária e típica para o caso de estudo. Nas simulações realizadas com essas cargas de trabalho, a política proposta apresentou melhor desempenho que as políticas de *timeout* e *preditiva*. Embora essa seja uma carga de trabalho experimental, esse resultado reforça a teoria e os resultados apresentados ao longo desse capítulo.



# Capítulo 6

## Conclusão

O gerenciamento dinâmico de energia permite reduzir o consumo de energia a partir de alterações no estado de consumo de dispositivos que encontram-se ociosos ou sub-utilizados. Variadas abordagens são comumente utilizadas e classificadas em três grandes categorias: baseadas em *timeout*, preditivas e estocásticas. Dentro desse contexto, é proposta neste trabalho de tese, uma política para o gerenciamento dinâmico de energia, denominada Política do Tempo Ótimo de Desligamento, TOD. Considera-se que exista um tempo ótimo para o qual o dispositivo deverá permanecer desligado ou num estado de baixo consumo, e que esse tempo não é necessariamente todo o intervalo ocioso. Além disso, espera-se que o religamento do dispositivo seja feito de forma preditiva. Assim, é formulado e resolvido um problema de otimização, que pretende minimizar o consumo de energia para a variável *tempo ótimo de desligamento*,  $T$ , atendendo especificações de restrição de penalidade de desempenho e consumo inferior a  $\mathcal{E}_{on}$ , que é o consumo máximo aceitável. Para determinar o consumo de energia para cada situação possível de desligamento e religamento, e computar a penalidade envolvida em cada situação, são desenvolvidos modelos analíticos para o consumo de energia e para a penalidade de desempenho. É construída uma tabela de políticas que consiste em uma matriz, onde para cada carga de trabalho são definidos os valores ótimos de desligamento,  $T$ , que minimizam o consumo de energia atendendo uma dada restrição de penalidade de desempenho. Em tempo de execução, são obtidas estimativas da carga de trabalho, utilizando a técnica das janelas deslizantes de múltiplos comprimentos, JMC. Como as estimativas podem ser diferentes das cargas de trabalho utilizadas na

construção da tabela de políticas, é feita uma interpolação para escolher dinamicamente a política mais adequada para tal carga de trabalho.

A técnica de gerenciamento dinâmico de energia desenvolvida neste trabalho e apresentada no Capítulo 4 foi aplicada para um dispositivo de rede Wi-Fi no Capítulo 5. Para fins de comparação, a política proposta foi simulada e comparada com uma política de *timeout* e uma *preditiva*. Vários cenários foram simulados e pode-se definir em quais deles a política proposta apresenta melhor desempenho em comparação com as demais. Primeiramente foram utilizadas cargas de trabalho estacionárias cujos comprimentos foram definidos a partir do teorema do limite central e cada caso foi analisado individualmente. Verifica-se que para cargas de trabalho leves, a política proposta apresenta um comportamento semelhante à política de *timeout*, quanto ao consumo de energia e a penalidade de desempenho. Entretanto, quanto mais intensa é a carga de trabalho, mais significativa é a redução de consumo de energia utilizando a política proposta. Para ampliar os cenários, cargas de trabalho estacionárias foram concatenadas para formar cargas não-estacionárias e assim comparar o desempenho da política proposta com as outras duas políticas. No cenário não-estacionário, verifica-se novamente que, quando na composição da carga de trabalho estão presentes cargas de trabalho muito leves, ( $q = 0, 1$ ), a política proposta não apresenta seu melhor desempenho. Entretanto, mesmo nesse cenário, ainda é possível obter menores valores de consumo de energia utilizando a política proposta, definindo penalidades de desempenho mais altas, ( $l \geq 0, 5$ ). Com a retirada das cargas de trabalho leves, ( $q = 0, 1; 0, 2; 0, 3$ ), vários outros cenários foram criados e em todos eles os resultados de simulação apontam que, com a política proposta é possível obter menores valores de consumo de energia e penalidade de desempenho.

A técnica proposta neste trabalho foi aplicada para interfaces de rede, Wi-Fi, entretanto, esse foi apenas um caso de aplicação. O desenvolvimento dessa política permite sua aplicação para qualquer dispositivo em um sistema computacional que permita gerenciamento de energia, e para o qual seja possível verificar variações na carga de trabalho ao longo do tempo.

## 6.1 Sugestões para Trabalhos Futuros

Na Seção 3.2.2 foi apresentado o método de janelas deslizantes de múltiplos comprimentos, JMC, que foi utilizado para estimação da carga de trabalho. A acurácia desse método está diretamente associada aos comprimentos de janelas utilizados. Janelas pequenas diminuem o atraso de identificação de uma carga estacionária para outra, em contrapartida, janelas grandes diminuem o erro de amostragem. De toda forma, as cargas de trabalho necessitam ter uma duração mínima para que seu modelo seja identificado. De forma que, nesse trabalho as durações das amostras de cargas de trabalho foram definidas a partir do teorema do limite central, considerando um intervalo de confiança e um nível de significância. Uma proposta de continuação desse trabalho é realizar aquisição de cargas de trabalho para interfaces de rede, Wi-Fi, preferencialmente em *notebooks* e computadores de mesa, por tempos longos, por exemplo: meses de aquisição. Tal aquisição deve ser realizada para diferentes usuários, de forma que seja possível verificar as variações de padrão das cargas de trabalho ao longo do tempo.

Na Seção 5.4.2 foi levantada a questão que, à medida que a carga de trabalho torna-se mais intensa, o desempenho da técnica proposta torna-se ainda melhor quando comparado a uma política de *timeout*. De fato, se a carga de trabalho é muito leve, i.e., constituída de intervalos ociosos muito longos, a penalidade de desempenho do ponto de vista do usuário pode ser “esquecida” como discutido no Capítulo 2, e dessa forma, para uma política que não utilize o religamento preditivo, o dispositivo poderá permanecer mais tempo desligado/estado de baixo consumo durante os intervalos ociosos e assim alcançará uma economia de energia mais significativa que com uma política de religamento preditivo. Dessa forma, pretende-se avaliar a viabilidade de implementar um seletor de políticas que atue de acordo com o modelo da carga de trabalho.

No Capítulo 5 a política proposta neste trabalho foi aplicada para um dispositivo de rede, Wi-Fi. A validação foi feita através de simulações nas quais foram considerados dados reais dos estados de consumo para esse dispositivo, como consumo de energia para o estado *on* e *off*, e a energia consumida nas transições de estados. Uma sugestão de trabalhos futuros é implementar essa e as outras duas políticas utilizadas neste trabalho, em um sistema computacional para validação experimental.

# Referências bibliográficas

- 1 LI, D.; GUO, B.; SHEN, Y.; LI, J.; HUANG, Y. Make image more energy efficient for mobile oled displays. In: *2016 13th International Conference on Embedded Software and Systems (ICCESS)*. [S.l.: s.n.], 2016. p. 143–147.
- 2 SHIH, Hung Cheng; WANG, Kuochen. An adaptive hybrid dynamic power management method for handheld devices. In: *IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC'06)*. [S.l.: s.n.], 2006. v. 1, p. 6 pp.–.
- 3 OKAMURA, H.; DOHI, T. Dynamic power management with optimal time-out policies. *IEEE Systems Journal*, v. 11, n. 2, p. 962–972, June 2017. ISSN 1932-8184.
- 4 SCHERRER, T.; KIM, S. Y.; YI, C. Low complexity, real-time adjusted power management policy using golden section search. In: *2013 International SoC Design Conference (ISOCC)*. [S.l.: s.n.], 2013. p. 229–232.
- 5 LU, Yung-Hsiang; MICHELI, G. De. Comparing system level power management policies. *IEEE Design & Test of Computers*, IEEE Computer Society Press, Los Alamitos, CA, USA, v. 18, n. 2, p. 10–19, 2001. ISSN 0740-7475.
- 6 JIANG, Q.; XI, H.-S.; YIN, B.-Q. Adaptive optimisation of timeout policy for dynamic power management based on semi-markov control processes. *Control Theory Applications, IET*, v. 4, n. 10, p. 1945–1958, Out 2010. ISSN 1751-8644.
- 7 CHUNG, Eui-Young; BENINI, L.; BOGLIOLO, A.; LU, Yung-Hsiang; MICHELI, G. De. Dynamic power management for nonstationary service requests. *IEEE Trans. Comput.*, IEEE Computer Society, Washington, DC, USA, v. 51, n. 11, p. 1345–1361, 2002. ISSN 0018-9340.
- 8 BENINI, L.; BOGLIOLO, A.; MICHELI, G. De. A survey of design techniques for system-level dynamic power management. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, v. 8, n. 3, p. 299–316, Jun 2000. ISSN 1063-8210.
- 9 WU, Kaiqiang; LIU, Yi; ZHANG, Haiwen; QIAN, Depei. Adaptive power management with fine-grained delay constraints. In: *2010 3rd International Conference on Computer Science and Information Technology*. [S.l.: s.n.], 2010. v. 2, p. 633–637.
- 10 DHIMAN, G.; ROSING, T. S. Dynamic power management using machine learning. In: *IEEE/ACM International Conference on Computer Aided Design*. [S.l.: s.n.], 2006. p. 747–754. ISSN 1092-3152.

- 11 VILAR, B. M. J. C.; SILVA, F. d. M.; LUIZ, S. O. D.; SILVA, J. J.; ALMEIDA, H.; PERKUSICH, A. Dynamic timeout power policy for network interfaces. In: *2014 IEEE Fourth International Conference on Consumer Electronics Berlin (ICCE-Berlin)*. [S.l.: s.n.], 2014. p. 279–283. ISSN 2166-6814.
- 12 BENINI, L.; BOGLIOLO, A.; PALEOLOGO, G. A.; MICHELI, G. De. Policy optimization for dynamic power management. *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, v. 18, n. 6, p. 813–833, Jul 1999. ISSN 0278-0070.
- 13 REN, Z.; KROGH, B.H.; MARCULESCU, R. Hierarchical adaptive dynamic power management. *IEEE Transactions on Computers*, v. 54, n. 4, p. 409–420, Abr 2005. ISSN 0018-9340.
- 14 LUIZ, S. O. D.; PERKUSICH, A.; LIMA, A. M. N. Multisize sliding window in workload estimation for dynamic power management. *IEEE Transactions on Computers*, v. 59, n. 12, p. 1625–1639, Dez 2010. ISSN 0018-9340.
- 15 The art of computer systems performance analysis: Techniques for experimental design, measurement, simulation and modeling, by raj jain. *SIGSIM Simul. Dig.*, ACM, New York, NY, USA, v. 21, n. 2, p. 53–, dez. 1990. ISSN 0163-6103. Disponível em: <<http://doi.acm.org/10.1145/382264.1108821>>.
- 16 LUIZ, S. O. D. *Gerenciamento Dinâmico de Energia em Processadores com Cargas de Trabalho Variantes no Tempo*. Tese (Doutorado) — Universidade Federal de Campina Grande, Campina Grande, PB, BRASIL, 2012.
- 17 SUH, Jinwoo; KANG, Dong-In; CRAGO, S.P. Dynamic power management of multiprocessor systems. In: *Conf. Rec. WPDRTS '02*. Ft. Lauderdale, FL, USA: IEEE, 2002.
- 18 HUANG, Shao-Min; HU, Chen; QI, Long-Ning. Predictive dynamic power management policy with consideration of outlying data. In: *8th International Conference on Solid-State and Integrated Circuit Technology, 2006. ICSICT '06*. Shanghai: IEEE, 2006. p. 2064–2066.
- 19 WENG, Li-Chuan; WANG, XiaoJun; LIU, Bin. A survey of dynamic power optimization techniques. In: *System-on-Chip for Real-Time Applications, 2003. Proceedings. The 3rd IEEE International Workshop on*. NJ, USA: IEEE, 2003. p. 48–52.
- 20 OKAMURA, H.; MIYATA, S.; DOHI, T. A markov decision process approach to dynamic power management in a cluster system. *IEEE Access*, v. 3, p. 3039–3047, 2015. ISSN 2169-3536.
- 21 KVETON, Branislav; GANDHI, Prashant; THEOCHAROUS, Georgios; MANNOR, Shie; ROSARIO, Barbara; SHAH, Nilesh. Adaptive timeout policies for fast fine-grained power management. In: *Proceedings of the 19th National Conference on Innovative Applications of Artificial Intelligence - Volume 2*. [S.l.]: AAAI Press, 2007. (IAAI'07), p. 1795–1800. ISBN 978-1-57735-323-2.

- 22 GREENAWALT, P.M. Modeling power management for hard disks. In: *Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, 1994., MASCOTS '94., Proceedings of the Second International Workshop on.* Durham, NC: IEEE, 1994. p. 62–66.
- 23 LUIZ, S. O. D.; PERKUSICH, A.; CRUZ, B. M. J.; NEVES, B. H. M.; ARAUJO, G. M. da S. Optimization of timeout-based power management policies for network interfaces. *IEEE Transactions on Consumer Electronics*, v. 59, n. 1, p. 101–106, Fev 2013. ISSN 0098-3063.
- 24 MADANI, N.M.; MASOUMI, M. A new optimization method for ctmdp system-level power management techniques. In: *The 16th International Conference on Microelectronics, 2004. ICM 2004 Proceedings.* Piscataway: IEEE, 2004. p. 215–218.
- 25 LU, Yung-Hsiang; MICHELI, G. De. Adaptive hard disk power management on personal computers. In: *Ninth Great Lakes Symposium on VLSI, 1999. Proceedings.* Ypsilanti, MI: IEEE, 1999. p. 50–53. ISSN 1066-1395.
- 26 CHUNG, Eui-Young; BENINI, L.; MICHELI, G. De. Dynamic power management using adaptive learning tree. In: *Conf. Rec. ICCAD'99.* Piscataway, NJ, USA: IEEE Press, 1999. p. 274–279. ISBN 0-7803-5832-5.
- 27 HWANG, Chi-Hong; WU, Allen C.-H. A predictive system shutdown method for energy saving of event-driven computation. *ACM Trans. Des. Autom. Electron. Syst.*, ACM, New York, NY, USA, v. 5, n. 2, p. 226–241, 2000. ISSN 1084-4309.
- 28 SRIVASTAVA, M. B.; CHANDRAKASAN, A. P.; BRODERSEN, R. W. Predictive system shutdown and other architectural techniques for energy efficient programmable computation. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, v. 4, n. 1, p. 42–55, Mar 1996. ISSN 1063-8210.
- 29 ZHAI, J.; LI, Y.; CHEN, H. An online optimization for dynamic power management. In: *2016 IEEE International Conference on Industrial Technology (ICIT).* [S.l.: s.n.], 2016. p. 1533–1538.
- 30 SIMUNIC, T.; BENINI, L.; GLYNN, P.; MICHELI, G. De. Event-driven power management. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, v. 20, n. 7, p. 840–857, Jul 2001. ISSN 0278-0070.
- 31 QIU, Qinru; PEDRAM, M. Dynamic power management based on continuous-time markov decision processes. In: *Proceedings 1999 Design Automation Conference (Cat. No. 99CH36361).* [S.l.: s.n.], 1999. p. 555–561.
- 32 QIU, Q.; QU, Q.; PEDRAM, M. Stochastic modeling of a power-managed system-construction and optimization. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, v. 20, n. 10, p. 1200–1217, Out 2001. ISSN 0278-0070.
- 33 SIMUNIC, T.; BENINI, L.; GLYNN, P.; MICHELI, G. De. Dynamic power management of laptop hard disk. In: *Proceedings Design, Automation and Test in Europe Conference and Exhibition 2000 (Cat. No. PR00537).* [S.l.: s.n.], 2000. p. 736.

- 34 LU, Yung-Hsiang; CHUNG, Eui-Young; SIMUNIC, T.; BENINI, T.; MICHELI, G. de. Quantitative comparison of power management algorithms. In: *Proceedings Design, Automation and Test in Europe Conference and Exhibition 2000 (Cat. No. PR00537)*. [S.l.: s.n.], 2000. p. 20–26.
- 35 TAN, Ying; QIU, Qinru. A framework of stochastic power management using hidden markov model. In: *Conf. Rec. DATE'08*. Munich, Germany: EDAA, 2008. p. 92–97.
- 36 QIU, Qinru; TAN, Ying; WU, Qing. Stochastic modeling and optimization for robust power management in a partially observable system. In: *Conf. Rec. DATE'07*. Nice Acropolis, France: EDAA, 2007. p. 1–6.
- 37 LUIZ, S. O. D.; PERKUSICH, A.; LIMA, A. M. N. Stochastic learning-based weak estimation for dynamic power management. In: *Conf. Rec. CBA'10*. Bonito, MS, Brasil: Anais do CBA 2010, 2010.
- 38 LUIZ, S. O. D.; LIMA, A. M. N.; PERKUSICH, A.; SILVA, J. J.; BUBLITZ, F. M.; ALMEIDA, H. Feasibility of a timeout power management policy. In: *Third International Conference on Consumer Electronics Berlin (ICCE-Berlin)*. [S.l.: s.n.], 2013. p. 181–184. ISSN 2166-6814.
- 39 MEYER, Paul L. *Probabilidade: aplicações à estatística*. 2. ed. Rio de Janeiro: LTC - Livros Tecnicos e Cientificos Editora S.A., 1983. ISBN 85-216-0294-4.
- 40 NEVES, B. H. M.; VILAR, B. M. J. C.; MENDES, R. R.; BUBLITZ, F. M.; SILVA, J. J.; LUIZ, S. O. D.; PERKUSICH, A.; ALMEIDA, H. Computer systems power model estimation. In: *Third International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*. Berlin: IEEE, 2013. p. 173–176.
- 41 VILAR, B. M. J. C.; LUIZ, S. O. D.; PERKUSICH, A. Comparison of dynamic power management policies for wi-fi network interfaces. In: *2016 12th IEEE International Conference on Industry Applications (INDUSCON)*. [S.l.: s.n.], 2016. p. 1–6.
- 42 VILAR, B. M. J. C.; LUIZ, S. O. D.; PERKUSICH, A.; SANTOS, D. R. Dynamic power management for network interfaces. In: *Conf. Rec. of ICCE-Berlin 2015*. [S.l.: s.n.], 2015. p. 383–387.
- 43 MELO, T. R.; LUIZ, S. O. D.; SILVA, J. J.; NEVES, B. H. M. Experimental platform for power measurements of computer systems. In: *Fourth International Conference on Consumer Electronics Berlin (ICCE-Berlin)*. [S.l.: s.n.], 2014. p. 424–428. ISSN 2166-6814.