

Multi-Objective Optimization Strategy Applied in Data Centers Electrical Subsystems

F. M. Sousa Sobrinho, G. R. A. Callou, W. S. Leonardo and B. Nogueira

Abstract—Data center infrastructures must have high availability, low cost, and high energy efficiency. However, these objectives are often conflicting. For instance, an additional Uninterruptible Power Supply (UPS) improves the system availability but may jeopardize the total cost and the energy consumption. This paper presents a strategy based on the Non-dominated Sorting Genetic Algorithm II (NSGA-II), a multi-objective evolutionary optimization algorithm, to improve the design of electrical data center architectures. To show the applicability of the proposed strategy, we present a comparative study between the brute force algorithm and proposed strategy based on NSGA-II, as it has shown promising results in multi-objective problems [1], [2] and [3]. When setting complex electrical infrastructure data center models, the results showed that applying the proposed strategy reduces the runtime 961 times and achieves Pareto optimal curvature with a difference of approximately 1%.

Index Term—Availability, Data Center, Pareto front, Modeling, Multi-objective Algorithm, Reliability Block Diagram.

I. INTRODUÇÃO

O aumento na necessidade de armazenamento de dados vem crescendo pela expansão de serviços tecnológicos e aplicações da internet como as mídias sociais e serviços de streaming [4]. Esses serviços trouxeram consigo o desafio para os projetistas desenvolverem Infraestruturas robustas de *data center*.

Um *data center* é composto por três pilares: (i) subsistema de energia, (ii) subsistema de resfriamento e (iii) subsistema de tecnologia da informação. O subsistema elétrico possui critérios que precisam ser atendidos por uma equipe capacitada tanto durante o projeto, quanto na manutenção [5]. Dentre estes critérios, uma infraestrutura elétrica dos *data centers* deve oferecer alta disponibilidade, com custo reduzido e eficiência energética maximizada. Muitas vezes estes critérios são conflitantes [6], por exemplo, considerando duas arquiteturas similares, mas com equipamentos diferentes, em que a primeira possua maior disponibilidade que a segunda. Em contrapartida, a segunda, mesmo com menor disponibilidade, pode ter um custo inferior em relação à primeira. Esses cenários com objetivos conflitantes acabam por motivar o estudo com foco em encontrar possíveis soluções para problemas multiobjetivos.

Este trabalho visa a proposição de uma estratégia para otimizar os critérios (disponibilidade, custo e eficiência energética) na construção do subsistema energético de *data center*. Como trata-se de um problema combinatório onde cada escolha de equipamentos para montagem da arquitetura de *data centers* impacta nos resultados, é necessário se valer de uma estratégia meta-heurística que busque otimizar de forma

integrada as seguintes funções objetivo: maximizar a disponibilidade, minimizar o custo e o desperdício energético (exergia). A fim de testar a aplicabilidade da estratégia baseada no NSGA-II, será feita a comparação de seus resultados com uma técnica de otimização global (força bruta). Tais resultados mostram que o conjunto das soluções da estratégia proposta se aproxima da curva de Pareto ótimo do algoritmo força bruta em uma diferença máxima de 1%. Além disso, a estratégia proposta teve o tempo de execução menor que o força bruta chegando a apresentar uma redução de 961 vezes no caso mais complexo. Este trabalho se diferencia dos demais da literatura, pois trata-se de uma estratégia de otimização multiobjetivo que retorna um conjunto de soluções ótimas não-dominadas.

O restante deste trabalho está organizado da seguinte forma. A Seção II apresenta os trabalhos relacionados. A Seção III apresenta a fundamentação teórica para auxiliar no entendimento do trabalho. A Seção IV explica a metodologia e os modelos propostos. A Seção V ilustra a aplicabilidade do algoritmo de otimização proposto através de um estudo de caso aplicando. Por fim, a Seção VI conclui o trabalho e apresenta possíveis direcionamentos da pesquisa.

II. TRABALHOS RELACIONADOS

Vários trabalhos vêm aplicando modelagens das infraestruturas de *data centers* a partir de técnicas de otimização, buscando encontrar soluções sustentáveis e com alta disponibilidade. Em [7] é proposta uma modelagem integrada para avaliar e otimizar a sustentabilidade, a disponibilidade e o custo de *data centers*. A semelhança com este trabalho consiste na utilização de três modelos formais *stochastic Petri nets* (SPN), *Reliability Block Diagrams* (RBD) e *Energy Flow Model* (EFM), sendo este último fundamental para a avaliação da métrica de exergia operacional. O algoritmo de otimização usado foi o *Greedy Randomized Adaptive Search Procedure* (GRASP). O trabalho, contudo, considera uma função objetivo mono-objetivo, e por isso não é capaz de encontrar a curva de Pareto.

Gregory et al. [8] apresentaram técnicas para maximizar a disponibilidade a partir da utilização de redundâncias como a *cold standby*. Os autores propuseram modelos em SPN para estimar o cálculo da disponibilidade. Apesar de utilizar a técnica *cold standby*, não foi foco dos autores a proposição de estratégias de otimização que possibilitasse a maximização da disponibilidade.

Em [9], os autores projetaram arquiteturas de *data center* que oferecem flexibilidade, automação, otimização e escalabilidade. De acordo com os resultados, a Programação Dinâmica (DP) foi a técnica de otimização que forneceu o mais alto nível de disponibilidade com o mais baixo custo.

Entretanto, este trabalho não explorou a eficiência energética das arquiteturas analisadas.

Em [10], os autores propuseram uma metodologia para estimar os dados de disponibilidade do *data center* com base nas informações de recursos de *hardware* fornecidas através da API Redfish. Os resultados da avaliação mostraram que a configuração de vPOD aumenta a disponibilidade quando comparada às estratégias convencionais. Contudo, não foi objetivo do trabalho a análise do custo.

Bennaceur et al. [11] apresentaram um modelo formal que analisa o impacto da variação da temperatura sobre a disponibilidade e confiabilidade nos subsistemas de *data centers*. Os autores utilizaram a modelagem *Production Trees* (PT), que permite a criação dos modelos levando em conta os fluxos de ar e elétrico do *data center*. Apesar da abrangência, não há nesse trabalho referência a uma técnica de otimização.

Em [12], algoritmos meta-heurísticos foram utilizados para otimizar arquiteturas de *data center*. A estratégia proposta visou comparar qual técnica resulta na maior maximização da disponibilidade e com o mínimo custo de aquisição dos componentes. Ao final do estudo, a técnica que utilizava o algoritmo *Differential Evolution* (DE) obteve os melhores resultados. Embora tenham empregado técnicas de otimização visando maximizar a disponibilidade, os autores não analisaram a eficiência energética das arquiteturas.

Os autores em [13] propuseram um algoritmo *electricity cost optimization* (ECO) baseado nos custos energético operacional do *data center*. Resultados experimentais demonstraram que o algoritmo proposto pode resultar em redução significativa do custo de energia, garantindo a qualidade do serviço e o tempo de resposta. No entanto, o trabalho se limita a analisar um equipamento (UPS).

A Tabela I apresenta um resumo comparativo entre os trabalhos relacionados, destacando os modelos aplicados, as métricas utilizadas e o tipo de otimização feita em cada estudo. É possível verificar que este trabalho difere dos demais, pois considera mais métricas que a literatura, além de adotar uma estratégia de otimização baseada em um algoritmo multiobjetivo, que é capaz de encontrar a curva de Pareto do problema, permitindo assim comparar diversos *trade-offs*.

TABELA I
COMPARAÇÃO ENTRE OS TRABALHOS RELACIONADOS

Trabalho	Modelos	Métrica	Otimização
[7]	SPN/RBD/EFM	Disponibilidade e custo	GRASP
[8]	SPN/RBD	Disponibilidade e custo	-
[9]	RBD	Disponibilidade e custo	DP
[10]	SPN/RBD	Disponibilidade	vPOD
[11]	SPN/PT	Disponibilidade e Eficiência	-
[12]	SPN	Disponibilidade	DE
[13]	SPN/RBD	Custo	ECO
Este trabalho	SPN/RBD/EFM	Disponibilidade, Eficiência e Custo	NSGA-II

III. FUNDAMENTAÇÃO TEÓRICA

Esta seção apresenta as informações para um melhor entendimento do trabalho desenvolvido.

A. Disponibilidade

A disponibilidade é a probabilidade de um sistema funcionar corretamente em um momento específico ou durante um período de tempo predefinido [14]. A disponibilidade de um sistema pode ser computada a partir dos tempos de falhas (MTTF - *Mean Time To Failure*) e reparo (MTTR - *Mean Time To Repair*) do sistema conforme mostrado na Equação 1:

$$D = \frac{MTTF}{MTTF + MTTR} \quad (1)$$

B. Custo

O custo é um dos objetivos a serem otimizados neste trabalho. O custo (C) de uma infraestrutura elétrica de *data center* leva em conta dois custos. O primeiro é o custo de aquisição (CA), e o outro é o custo operacional (CO), computado conforme mostrado na Equação 2, onde $E_{consumida}$ é o consumo energético em kW, E_{custo} é o preço do quilowatt/hora, T é o tempo, D é a disponibilidade.

$$CO = E_{consumida} * T * E_{custo} * D \quad (2)$$

C. Exergia

A exergia é uma métrica utilizada para quantificar a energia perdida pelo equipamento, ou seja, a energia que o sistema não conseguiu converter em trabalho útil [15]. Tal técnica permite avaliação do quão eficiente é um equipamento. Como uma infraestrutura de *data center* é composta por diversos equipamentos elétricos, para se ter o cálculo da exergia operacional é necessário o somatório das energias dissipadas (Ex_{op_i}) por cada um deles durante um intervalo (T), e durante o efetivo funcionamento (D) da arquitetura avaliada. A Equação (3) representa a exergia operacional:

$$Ex_{op} = \sum_{i=1}^n Ex_{op_i} * T * D \quad (3)$$

D. Diagrama de Bloco de Confiabilidade (RBD)

O RBD é usado para mostrar as interdependências entre todos os elementos de um sistema. Esta modelagem permite representar uma infraestrutura elétrica do *data center*, onde um equipamento tem suas interações e dependências com outros equipamentos [16]. Esta modelagem formal pode representar componentes conectados em série ou em paralelo, conforme mostrado na Figura 1. Quanto a composição é em série, se houver a falha de um equipamento, o sistema todo para de funcionar. Já na disposição em paralelo, o sistema para de funcionar se todos os equipamentos falharem.

A cada bloco são associados os valores médios dos tempos de falha e reparo. O sistema está funcional, se há pelo menos um caminho de componentes funcionando desde a entrada até a saída [17].

O cálculo da disponibilidade do RBD em série com n blocos é dado pela Equação 4:

$$D_{(s)} = \prod_{i=1}^n D_i \quad (4)$$

Já o cálculo da disponibilidade dos mesmos n blocos conectados em paralelo é realizado pela Equação 5.

$$D_{(p)} = \prod_{i=1}^n (1 - D_i) \quad (5)$$

onde: D_i representa a disponibilidade do equipamento i .

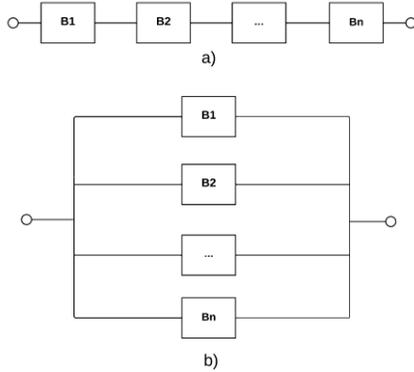


Fig. 1. Estrutura básica do RBD. a) Série. b) Paralelo.

E. *Redes de Petri Estocástica (SPN)*

A SPN é um formalismo usado para especificar sistemas dinâmicos de eventos discretos, mas também permite análise estocástica quanto a mudança de estado deste sistema. Nas redes de Petri estocásticas as transições podem disparar após um atraso probabilístico associado às transições temporizadas. A definição formal das SPNs pode ser encontrada em [18].

Os elementos básicos que compõem um SPN são: arco, lugar, *token* e transição. O arco pode ser padrão ou inibidor, já a transição pode ser do tipo estocástica (temporizada) ou imediata. Na Figura 2 é possível visualizar os principais elementos das SPNs.

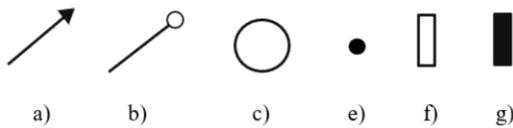


Fig. 2. Elementos das redes de Petri Estocásticas. a) arco; b) arco inibidor; c) lugar; d) ficha; e) transição estocástica; f) transição imediata.

A Figura 3 mostra um exemplo em SPN que possui dois lugares para representar os estados de um equipamento (UPS). Um *token* no lugar *UPS_ON* representa o estado em pleno funcionamento, no outro lugar (*UPS_OFF*) representa o estado quebrado. Esse modelo tem duas transições estocásticas, e que representam os períodos de falha e reparo do equipamento.

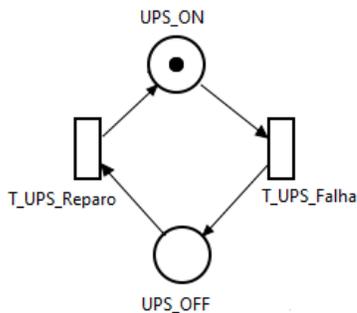


Fig. 3. Modelo SPN e UPS ligado.

F. *Otimização Multiobjetivo*

Problemas de otimização com duas ou mais funções objetivo são chamados de multiobjetivos. Este tipo de problema exige a otimização simultânea de vários interesses interdependentes e muitas vezes conflitantes e, por isso, não há uma solução única [19]. Em particular, o presente trabalho busca maximizar a disponibilidade (ver Equação 1), minimizar o custo (custo aquisição + Equação 2), e minimizar a exergia (Equação 3).

IV. METODOLOGIA

A Figura 4 ilustra a metodologia proposta neste trabalho, cujas as etapas para realizar a otimização dos modelos das arquiteturas elétricas de *data centers* são descritas ao longo desta seção.

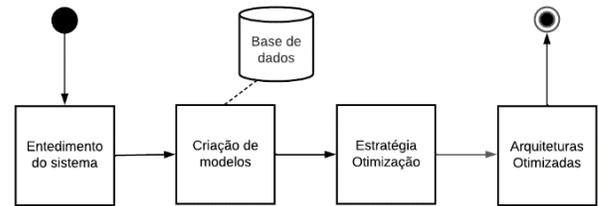


Fig. 4. Metodologia.

A. *Entendimento do Sistema*

A primeira etapa da metodologia deste trabalho é o entendimento do sistema. Esse estudo modela o sistema elétrico de *data center* [20] que é, na arquitetura base, composto por UPS (*Uninterruptible Power Supplies*), SDT (*Step Down Transformer*), *subpanel* e *powerstrip*.

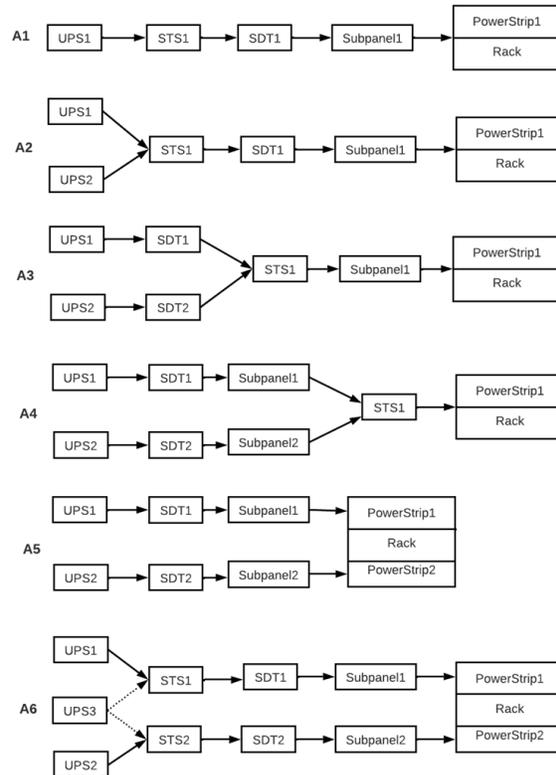


Fig. 5. Arquiteturas elétricas de *data center*.

A Figura 5 apresenta as arquiteturas que serão modeladas. Partindo da arquitetura A1, onde não há redundância, até a A5, em que existe um caminho todo tolerante a falhas. Além disso, a arquitetura A6 adiciona uma redundância *cold standby*.

Na falha de um equipamento em uma arquitetura sem redundância (A1), mesmo com os demais componentes ativos, o sistema ficará indisponível. Na arquitetura A2 existe redundância do UPS, e caso o UPS1 falhe, mas tenha o UPS2 e os demais componentes em pleno funcionamento, o sistema permanecerá disponível.

B. Criação de Modelos

A próxima etapa da metodologia proposta corresponde à criação dos modelos (ex., RBDs e SPNs). Esses modelos são criados através da ferramenta *Stars* [21], que auxilia na integração da estratégia de otimização com o ambiente de avaliação do Mercury [22].

Os parâmetros dos equipamentos como tempo médio de falha (MTTF), tempo médio de reparo (MTTR), custo e eficiência energética que compõem as arquiteturas que modeladas neste trabalho foram extraídos de [23] e [24].

Existem cenários em que apenas o RBD seria insuficiente para se avaliar a métrica da disponibilidade. É o caso de sistemas que adotam redundâncias em *cold standby*, por exemplo. Nesses casos, a metodologia adotada indica a utilização da modelagem em SPN.

Na arquitetura A6, o equipamento UPS3 só será ativado quando o UPS1 ou UPS2 estiverem em situação de falha. A Figura 6 apresenta o modelo SPN proposto para representar esse caso. Nesse modelo, a falha do UPS1 ou UPS2 faz com que um token seja gerado no lugar Controle, habilitando a transição T_{ativa} que é responsável pela ativação do UPS3. Após o tempo de ativação do UPS3, a ficha de Controle é consumida e o lugar $UPS3_ON$, inicialmente vazio, receberá uma ficha. A disponibilidade desse modelo é computada por: $P\{\#UPS1_ON=1\}OR\{\#UPS2_ON=1\}OR\{\#UPS3_ON=1\}$.

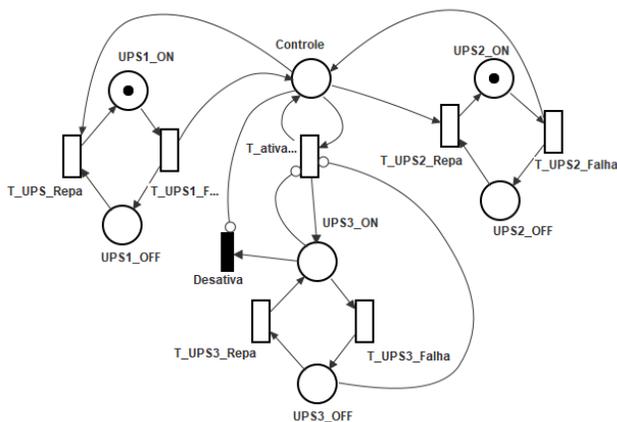


Fig. 6. Modelo SPN– ups1, up2 e ups3 (*Cold standby*).

A estratégia proposta faz uso da modelagem em conjunto de SPN e RBD. A Figura 7 apresenta a estratégia proposta, onde a parte em *cold* é modelada em SPN. O resultado da avaliação desse modelo é utilizado como parâmetro de entrada para um bloco RBD (B1), conforme mostrado na Figura 7 (c).

Uma vez que se tem o RBD completo, já com os blocos B1 e B2 parametrizados, pode-se avaliar todo o sistema e, assim, obter a disponibilidade de toda arquitetura analisada.

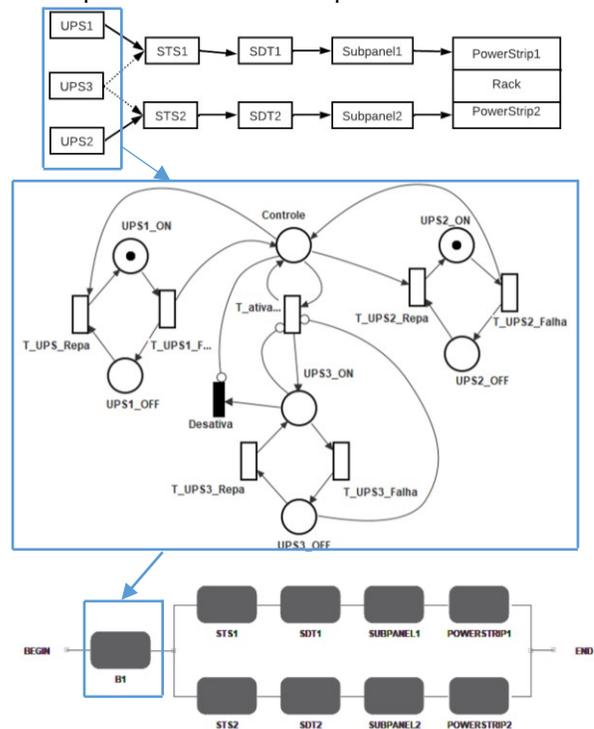


Fig. 7. Integração entre modelos. Modelagem *Cold standby* (UPS1, UPS2 e UPS3) em SPN convertida em um bloco RBD - B1.

C. Estratégia de Otimização Multiobjetivo

Esta subseção detalha o algoritmo adotado, baseado no NSGA-II [25], para realizar a otimização multiobjetivo. A estratégia adotada, que é baseada no NSGA-II, é apresentada no Algoritmo 1. O NSGA-II possui dois principais operadores: ordenação de indivíduos não dominados da população e uma métrica de promoção de diversidade da população, *Crowding Distance* (CD) [26].

O Algoritmo 1 tem como entradas o modelo de arquitetura (A1-A6), o tamanho da população, número de iterações do algoritmo e as funções alvo da otimização (disponibilidade, custo total e exergia operacional). A saída é um conjunto solução Pareto ótimo contendo arquiteturas otimizadas.

É importante destacar que as funções objetivo deste trabalho são computadas da seguinte forma: a disponibilidade é calculada usando simulação estocástica por meio da ferramenta *Mercury*, o custo (C) é obtido pela soma do custo de aquisição (CA) e o custo operacional CO (Equação 2), já a exergia operacional é calculada conforme apresentado na Equação 3. Cada arquitetura é composta por equipamentos, e a partir de uma lista de dispositivos disponíveis para cada equipamento, teremos os parâmetros que podem variar (MTTF, MTTR, custo de aquisição e eficiência energética).

Os indivíduos são constituídos tomando por base uma arquitetura pré-definida (A1-A6) no momento do uso algoritmo. Quando criado um indivíduo o mesmo recebe um conjunto de equipamentos. Cada um desses equipamentos possui apenas o seu tipo e conexão com os demais

equipamentos na arquitetura. A Figura 8 ilustra como é feita a construção de um indivíduo.

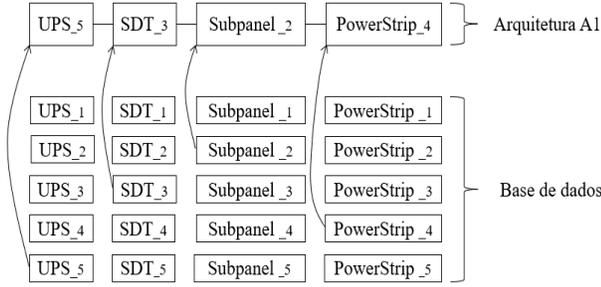


Fig. 8. Criação de um indivíduo da arquitetura A1.

A codificação do indivíduo utiliza uma base de dados que é mapeada num dicionário usando como chaves *Strings* que se referem aos tipos de componentes existentes. Então os equipamentos da arquitetura são percorridos um a um, e a partir desse mapeamento uma função de escolha aleatória e sem repetição seleciona um equipamento do mesmo tipo e insere seus atributos (MTTF, MTTR, custo de aquisição e eficiência energética) no equipamento do indivíduo que está sendo percorrido no momento. Esse processo se repete até o último equipamento do indivíduo.

Algoritmo 1 – Estratégia baseada no NSGA-II

Entrada: $N, g, f_k(X) \triangleright N$ tamanho da população, g gerações, $f_k(X)$ são as k funções objetivo.
01: arquitetura ← criarArquitetura \triangleright Indivíduo
02: *arquitecturas* ← gerarPopulaçãoAleatória (baseDados);
03: **para** $i \leftarrow 0$ até $N-1$ **faça**
04: **para** $j \leftarrow 0$ até $k-1$ **faça**
05: calcularObj $f_j(\text{arquitecturas}[i])$; \triangleright Comunicação Mercury
06: **fim para**;
07: **fim para**;
08: *arquitecturas* ← ordenar(ranquear(*arquitecturas*)) \triangleright non-dominated
09: **para** $i \leftarrow 0$ até $g-1$ **faça**
10: *paisEleitos* ← seleçãoTorneioBinario(*arquitecturas*)
11: *descendentes* ← operadorGenético(*paisEleitos*)
12: *arquitecturas* ← *arquitecturas*.adicionar(*descendentes*)
13: *arquitecturas* ← ordenar(ranquear(*arquitecturas*))
14: *arquitecturas* ← descartarPiores() \triangleright crowding distance
15: **fim para**
16: **retorna** *arquitecturas*

O Algoritmo 1 corresponde a implementação de uma estratégia baseada no NSGA-II que tem por objetivo otimizar a população inicial de entrada, considerando as suas funções objetivo (disponibilidade, custo e exergia). A população é composta por diversos indivíduos (lista de indivíduos). No caso deste trabalho, os modelos das arquiteturas dos subsistemas elétricos de *data center* representam a população. Estes modelos seguem o padrão adotado pelo trabalho realizado em [27].

A primeira linha deste algoritmo chama o método *criarArquitetura*, que define um modelo base de arquitetura. Na linha 2, ocorre a geração da população através da combinação de componentes extraída da base de dados que contém os atributos (MTTF, MTTR, custo de aquisição e eficiência energética) dos diversos componentes de cada tipo.

As linhas 3 a 7 realizam chamadas sucessivas ao método *calcularObj*, que se comunica com o ferramental *Mercury* [22] pelos métodos *metric* e *solve*, ambos necessários para avaliar as funções objetivo das arquiteturas. Esse método é invocado para se obter o valor de cada função, sendo expresso da seguinte forma: para cada função objetivo j das k funções objetivo é extraído o seu valor para cada arquitetura i (indivíduo) do conjunto de N arquiteturas (população).

A linha 8 realiza o ranqueamento das arquiteturas geradas, isto é, a população é classificada baseada no método o ranqueamento das soluções não dominadas usando o *crowding distance*. Esse algoritmo tem por base o algoritmo proposto originalmente em [25]. O laço principal da estratégia da estratégia inicia na linha 9, tendo como condição de parada o número de gerações. Para cada geração, os pais são escolhidos através da seleção torneio. Uma vez definido os pais, o algoritmo prossegue gerando os descendentes. Para isso, utiliza-se os operadores genéticos de cruzamento e mutação, onde através da seleção torneio um par de indivíduos são tomados e, a partir de um ponto de cruzamento, parte dos equipamentos de um indivíduo são trocados com o outro par, com 10% de chance de haver mutação, onde um equipamento de um indivíduo pode ser mudado por outro do mesmo tipo que esteja na base de dados.

A linha 12 é responsável por adicionar os descendentes ao conjunto de arquiteturas, novamente há o ranqueamento, por conta da inclusão de novos indivíduos na população. Feito isto, é necessário descartar as piores arquiteturas baseado no método *crowding distance* [26]. Ao final de g gerações, o Algoritmo 1 traz como resposta um conjunto de arquiteturas otimizadas por uma abordagem multiobjetivo, isto é, a fronteira aproximada de Pareto.

V. ESTUDO DE CASO

Essa seção apresenta um estudo de caso que busca otimizar as métricas de disponibilidade, custo e exergia nas infraestruturas elétricas de *data centers* (Figura 4). Além disso, essa seção também apresenta uma avaliação da performance do Algoritmo de otimização proposto em relação ao algoritmo de força bruta o qual gera todas as combinações dos indivíduos retornando de forma exaustiva todas as soluções não dominadas.

Para cada indivíduo gerado no força bruta foram avaliadas suas funções objetivo através de comunicação via *scripts* do Mercury [22]. Foram feitas normalizações dos resultados das funções conforme mostrado na Expressão 6.

$$f(x)_{normal} = \frac{f(x) - f(x)_{min}}{f(x)_{max} - f(x)_{min}} \quad (6)$$

A Tabela II apresenta os parâmetros de entrada (*MTTF*, *MTTR*, eficiência energética (EE) e custo de aquisição) adotados com base em [23] e [27]. Intervalos (min e máx.) foram definidos a partir dos valores bases dessa tabela para criação da base de equipamentos utilizados para a otimização. Sendo assim, para a eficiência energética, foi definida uma variação de 15%; para o MTTF foi adotada a variação de 50%; e considerando o custo dos equipamentos se variou 15%. Esse estudo adotou o MTTR de cada equipamento constante. Ao

todo foram gerados 10 equipamentos de cada tipo, que se combinados na arquitetura mais simples (A1) geram 10000 combinações possíveis. A estratégia multiobjetivo proposta e o algoritmo de força bruta foram aplicados nas 6 arquiteturas. O objetivo é maximizar a disponibilidade, diminuir a exergia operacional e o custo. O computador utilizado para rodar os modelos teve a seguinte configuração Win10, Intel i5 1.80 GHz, 8GB RAM, 1TB HD.

TABELA II
PARÂMETROS DE ENTRADA

Equip	EE (%)	MTTF (h)	MTTR (h)	CUSTO (R\$)
UPS	95,3	50000	8,0	15000
STS	95,5	58284	6,0	1000
SDT	98,5	282000	156	680
Subpanel	95,5	304000	2,4	250
Power Strip	97	274097	3,8	256

O algoritmo adotado teve como entrada uma população inicial de 100 indivíduos, 1000 gerações e foi executado 10 vezes, por se tratar de uma solução estocástica. Os resultados obtidos são apresentados no gráfico apresentado na Figura 9, onde é possível comparar a estratégia multiobjetivo proposta com o algoritmo da força bruta. Pelo gráfico é possível observar que as soluções geradas se aproximam bastante das soluções ótimas (fronteira de Pareto - soluções não dominadas).

- ▲ Resultados da Otimização Força Bruta
- Resultados da Otimização Baseada no NSGA-II

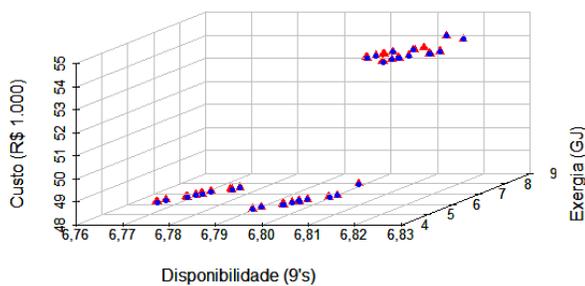


Fig. 9. Gráfico Comparativo: Alg. Baseado no NSGA-II x Força Bruta (arquitetura A6).

É importante destacar que foram computados os tempos de execução demandados para cada abordagem. Por exemplo, assumindo a arquitetura A1, o tempo com o algoritmo de otimização proposto demandou em média aproximadamente 200s em comparação com os 5600s necessários utilizando a força bruta. Já para a arquitetura A6, o tempo demandado médio foi de 521s para o algoritmo de otimização proposto versus 500.681s no algoritmo de busca exaustiva.

É considerável a diferença entre o tempo gasto pelos algoritmos, principalmente quando são arquiteturas mais complexas. Enquanto o algoritmo baseado no NSGA-II tem crescimento linear do tempo, o Força Bruta tem aumento de tempo de execução de forma exponencial.

De posse dos resultados é necessária a validação da estratégia proposta. Para tal, foi necessário comparar as métricas das arquiteturas otimizadas com o valor ótimo. Como o algoritmo de força bruta avalia todos os cenários, ele consegue obter a

arquitetura que tenha a máxima disponibilidade. Então, esta será a solução ótima que é utilizada para comparar com a disponibilidade obtida no algoritmo baseado no NSGA-II.

Não só a disponibilidade, após avaliar todos os cenários, e usando como critério as soluções não dominadas ranqueadas pelo *crowding distance*, é possível chegar à arquitetura que tenha o menor custo (independente das outras métricas). Similarmente, a mesma estratégia foi adotada para se obter os resultados para a exergia operacional. Obtendo assim o ótimo de cada métrica.

TABELA III
ALGORITMO PROPOSTO BASEADO NO NSGA-II X FORÇA BRUTA (A6)

	Valor Ótimo	Média	DP	IC
Força Bruta				
Disp. (9's)	6,818	6,8	0,27	(6,71 – 6,89)
Exergia (GJ)	4,23	6,20	1,57	(5,82 – 6,42)
Custo (R\$)	50185,1	50887	8235	(49272 – 52501)
Alg. Proposto				
Disp. (9's)	6,818	6,76	0,15	(6,65 – 6,87)
Exergia (GJ)	4,29	6,23	0,6	(6,10 – 6,43)
Custo (R\$)	50831,01	50996	3124	(49967 – 51232)

A Tabela III apresenta os resultados dos valores ótimos individuais de cada função objetivo, após o ranqueamento utilizando o operador *crowding distance* tanto do algoritmo de força bruta, quanto do algoritmo baseado no NSGA-II. O modelo da arquitetura utilizada para tal análise foi a A6. Além disso, nesta mesma tabela são apresentadas as médias de cada métrica analisada, os desvios padrões (DP) e o intervalo de confiança (IC). Sendo assim, pode-se perceber que os intervalos de confiança baseado nas médias das funções objetivo do Algoritmo Proposto baseado no NSGA-II e o Força Bruta se sobrepõem.

VI. CONCLUSÃO

Este trabalho propôs uma estratégia de otimização multiobjetivo baseada no NSGA-II aplicada a sistemas elétricos de *data centers*. É importante enfatizar que diferentemente de uma abordagem multiobjetivo clássica, como a de somas ponderadas, a estratégia proposta não atribui pesos as funções objetivo, entretanto usa os conceito de dominância de Pareto e distância de multidão. Além disso, foi utilizada uma abordagem híbrida dos modelos RBD e SPN, a fim de proporcionar dados robustos para um melhor entendimento.

Os resultados obtidos através do estudo de caso mostram que o conjunto dos modelos das arquiteturas elétricas de *data center* otimizadas pela estratégia adotada se aproxima da curvatura ótima de Pareto gerado pelo Força Bruta com uma diferença de aproximadamente 1%. Também é importante ressaltar que a estratégia apresentada conseguiu otimizar as métricas em estudo (disponibilidade, exergia operacional e custo) em todos os modelos abordados, tendo na modelagem mais complexa (A6) um tempo de execução 961 vezes menor, quando comparada ao algoritmo de força bruta. Assim, o tomador de decisão que utilizar a estratégia proposta neste trabalho tornará seu o processo decisório mais flexível, uma vez que terá, em tempo hábil, a possibilidade de escolha de arquiteturas elétricas

de *data center* com máxima disponibilidade e mínimos custo e exergia operacional.

Esse trabalho também apresentou uma análise estatística, que permite validar a estratégia proposta, onde gerações futuras das arquiteturas possuem indivíduos melhores, tomamos por base o desvio padrão de cada função da população gerada pelo algoritmo.

Como trabalho futuro sugere-se aplicar a estratégia proposta para outros subsistemas de *data center* (refrigeração e TI).

AGRADECIMENTOS

Os autores gostariam de agradecer ao CNPq, CAPES e à FACEPE pelo apoio para a realização desta pesquisa.

REFERÊNCIAS

- [1] Huo, J.; Liu, J.; Gao, H. "An NSGA-II Algorithm with Adaptive Local Search for a New Double-Row Model Solution to a Multi-Floor Hospital Facility Layout Problem". *Appl. Sci.* 2021.
- [2] Gaurav Srivastava, Alok Singh, Rammohan Mallipeddi, "NSGA-II with objective-specific variation operators for multiobjective vehicle routing problem with time windows," *Expert Systems with Applications*, vol. 176, 2021.
- [3] V. Manuel, R. César, K.I. Eduardo and I. P. Mario. "A multi-modal algorithm based on an NSGA-II scheme for phylogenetic tree inference," *Biosystems*, vol. 236, 2022.
- [4] M. Dayarathna, Y. Wen and R. Fan, "Data Center Energy Consumption Modeling: A Survey," in *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 732-794, Firstquarter 2016.
- [5] E. d. C. Sousa, "Os benefícios do investimento na capacitação profissional da equipe de manutenção e operação de data center." *Datacenter: projeto, operação e serviços-Unisul Virtual*, 2017
- [6] Y. Kumar. "Multiobjective, Multiconstraint Service Restoration of Electric Power Distribution System with Priority Customers," *IEEE Trans. on Power Delivery*, vol. 23, no. 1, (pp. 261–270). Jan. 2008
- [7] G. Callou, J. Ferreira, P. Maciel, D. Tutsch, and R. Souza, "An integrated modeling approach to evaluate and optimize data center sustainability, dependability and cost," *Energies*, vol. 7, no. 1, pp. 238–277, 2014.
- [8] L. Gregory, X. Liudong and D. Yuanshun, 2014. "Cold vs. hot standby mission operation cost minimization for 1-out-of-N system," *European Journal of Operational Research, Elsevier*, vol. 234(1), pp. 155-162.
- [9] G. L. Santos, P. T. Endo, D. Rosendo, D. M. Gomes, A. Moreira, J. Kelner, D. Sadok, G. E. Gonçalves, and M. Mahloo. "Maximizing the Availability of Composable Systems of Next-Generation Data Centers." *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*.
- [10] R. Daniel, G. Demis, S. Guto, L. Goncalves, L. G. F. Silva, E. P. Takako, S. D. F. Hadj, K. J. Mehta, M. Wildeman. "A methodology to assess the availability of next-generation data centers." *JOURNAL OF Supercomputing*, p. 1, 2019.
- [11] W. M. Bennaceur, "Electrical and Thermal System Impact on the Availability of a Data Center's System." *3rd Internacional Conference on System Reliability and Safety*, 2018.
- [12] G. Glauco, G. Demis, L. Guto, D. Rosendo, A. Moneira, J. Kelner, D. Sadok and P. Endo. "Optimizing the Cloud Data Center Availability Empowered by Surrogate Models." In: *Hawaii International Conference on System Sciences*, 2020.
- [13] Y. Huang, G. Li, P. Wang, F. Chang, J. Li, "Electricity Cost Optimization of Data Center Interactive Services with UPS." *15th International Computer Conference on Wavelet Active Media Technology and Information Processing*, 2018.
- [14] T. A. Nguyen, D. Min, E. Choi and T. D. Tran, "Reliability and Availability Evaluation for Cloud Data Center Networks Using Hierarchical Models," in *IEEE Access*, vol. 7, pp. 9273-9313, 2019.
- [15] T. V. Bezerra, W. S. Leonardo, G. Albuquerque Junior, G. Callou. "Dimensioning the relationship between availability and data center energy flow metrics." *Revista de Informática Teórica e Aplicada: RITA*, 2020.
- [16] W. Wang, J. M. Loman, R. G. Armo, P. Vassiliou, E. R. Furlong and D. Ogden, "Reliability block diagram simulation techniques applied to the IEEE Std. 493 standard network," in *IEEE Transactions on Industry Applications*, vol. 40, no. 3, pp. 887-895, May-June 2004.
- [17] J. M. Loman and W. Wang, "On reliability modeling and analysis of highly-reliable, large systems," in *Proc. 48th Annu. Reliability and Maintainability Symp.*, pp. 456–459, 2002.
- [18] M. A. Marsan. "Stochastic Petri Nets: An Elementary Introduction." In *Advances in Petri Nets* 1989, pp. 1-29: Springer, 1990.
- [19] S. Bechikh, A. Chaabani and L. Ben Said, "An Efficient Chemical Reaction Optimization Algorithm for Multiobjective Optimization," in *IEEE Transactions on Cybernetics*, vol. 45, Oct. 2015.
- [20] Association, T. I. "Telecommunications Infrastructure Standard for Data centers ANSI/TIA-942." [S.l.]: *Telecommunications Industry Association*, 2005.
- [21] W. S. Leonardo, G. R. A. Callou. "Stars: um ambiente integrado para avaliação de disponibilidade, custo e consumo de energia de sistemas." *Revista de Informática Teórica e Aplicada: RITA*, 2021.
- [22] D. M. Oliveira. "Advanced stochastic petri net modeling with the mercury scripting language." In: *11th EAI International Conference on Performance Evaluation Methodologies and Tools*. [S.l.: s.n.], 2017.
- [23] A. P. Guimarães, A. Pereira. "Análise de aspectos de dependabilidade em sistemas de data centers integrando as infraestruturas de comunicação, de potência e de refrigeração." *Revista Brasileira de Administração Científica*, 2020.
- [24] G. Callou, J. Ferreira, P. Maciel, D. Tutsch and R. Souza, "An integrated modeling approach to evaluate and optimize data center sustainability I dependability and cost", *Energies*, vol. 7, no. 1, 2014.
- [25] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan. "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, 2(6):182-197, 1999.
- [26] J. Sun, J. Deng, Y. Li. "Indicator & crowding distance-based evolutionary algorithm for combined heat and power economic emission dispatch," *Applied Soft Computing*, vol. 90, 2020.
- [27] F. F. L. Melo, E. Carneiro, G. R. A. Callou. "Optimization of Electrical Infrastructures at Data Centers through a DoE-based Approach." *Journal of Supercomputing*, 2021.



Francisco Monte Sousa Sobrinho is graduated in computer engineering at University of Pernambuco (UPE) and is a Msc. candidate at Federal Rural University of Pernambuco (UFRPE). He has interest in optimization, data center and artificial intelligence: evolutionary metaheuristics.



Gustavo Rau de Almeida Callou is an associate professor at the Federal Rural University of Pernambuco (UFRPE), Brazil. He has a Ph.D. in Computer Science from the Federal University of Pernambuco (UFPE), Brazil, with a split-site doctoral program at Bergische Universität Wuppertal, Germany, in the area of Performance Evaluation. Callou is the leader of the System Modeling and Optimization Research Group at UFRPE, and his key research interests include Petri nets, Reliability Analysis, Fault-Tolerant Computing, Performance Engineering, Sustainability, Computer Networking, Cloud Computing, and Data Centers



Wenderson de Souza Leonardo is graduated in Bachelor of Information Systems at Federal Rural University of Pernambuco (UFRPE) and is a Msc. candidate at Federal Rural University of Pernambuco (UFRPE).



Bruno Nogueira is an assistant professor at the Institute of Computing at Universidade Federal de Alagoas, Brazil. He obtained his BSc (2009), MSc (2010), PhD (2015) in computer science from Federal University of Pernambuco. His research interests focused in Optimization, Performance and dependability evaluation, and High-performance computing.