UNIVERSIDADE
FEDERAL RURAL
DE PERNAMBUCO

Ewerton Cleyton Silva de Queiroz

Analisando o Backup-as-a-Service como uma Estratégia de Recuperação de Desastres

Recife

2021

Ewerton Cleyton Silva de Queiroz

Analisando o Backup-as-a-Service como uma Estratégia de Recuperação de Desastres

Artigo apresentado ao Curso de Licenciatura em Computação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Licenciado em Computação.

Universidade Federal Rural de Pernambuco – UFRPE

Departamento de Computação

Curso de Licenciatura em Computação

Orientador: Ermeson Carneiro de Andrade

Coorientador: Júlio Rodrigues de Mendonça Neto

Recife

2021

Resumo

Nos ambientes modernos, falhas dos sistemas de Tecnologia da Informação e Comunicação (TIC) podem ter consequências graves para os negócios como perda de dados, insatisfação do cliente e perda de receita. Soluções de recuperação de desastres (DR), tais como Backup como Serviço (BaaS), vêm sendo adotadas por empresas como forma de evitar esses problemas e garantir a continuidade dos negócios. No entanto, existem diversas variáveis a serem consideradas na adoção de uma solução de DR. Portanto, neste artigo, apresentamos uma abordagem integrada utilizando experimentos e modelagem para avaliar um ambiente de BaaS para fins de DR. Em nossa análise, consideramos importantes métricas de DR como disponibilidade, *downtime*, RTO (*Recovery Time Objective*) e RPO (*Recovery Point Objective*). Os resultados mostraram que quando o BaaS é adotado, a disponibilidade do ambiente pode variar de acordo com a quantidade de dados de backup ou restauração. Além disso, uma análise de sensibilidade realizada apontou que o RTO e o RPO foram influenciados principalmente pelo tempo médio para restaurar o centro de dados (DC) e pelo intervalo de backup, respectivamente. A abordagem proposta pode ajudar empresas ou indivíduos interessados em adquirir soluções de DR no processo de tomada de decisão.

Palavras-chave: Análise de sensibilidade, Backup-as-a-Service, Recuperação de desastres, Redes de Petri, Tolerância a falhas.

Abstract

In modern environments, failures in information and communication technology (ICT) systems can have several consequences for a business, like data and revenue loss and customers dissatisfaction. Disaster recovery (DR) solutions, as Backup-as-a-Service (BaaS), has been adopted by companies as a way to avoid these problems and assure business continuity. Nevertheless, there are plenty of variables to consider during the adoption of a DR solution. Then, in this work, we present an integrated approach using experiments and models to evaluate a BaaS environment designed for DR. In our analysis, we consider relevant DR metrics like availability, downtime, RTO (Recovery Time Objective), and RPO (Recovery Point Objective). The results demonstrate that once BaaS is applied, the environment availability can vary according to the amount of data needed to be backed up or restored. Furthermore, sensitivity analysis indicates that the time needed to recover the data center and the backup interval are the most important parameter values for metrics like RTO and RPO. The proposed approach can help companies or individuals involved in the decision-making process for purchasing a DR solution.

Keywords: Sensitivity analysis, Backup-as-a-Service, Disaster recovery, Petri networks, Fault tolerance.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

TIC	Tecnologia da Informação e Comunicação
DC	Centro de Dados
VM	Máquina Virtual
KVM	Máquina Virtual baseada em Kernel
ICMP	Protocolo de Mensagem de Controle da Internet
RAM	Memória de Acesso Aleatório
HDD	Disco Rígido
vCPU	Unidade Central Virtual de Processamento
DR	Recuperação de Desastre
BD	Banco de Dados
BaaS	Backup-as-a-Service
RTO	Recovery Time Objective
RPO	Recovery Point Objective
DSPN	Rede de Petri Estocástica e Determinística
MTTF	Tempo Médio para ocorrência de uma Falha
MTTR	Tempo Médio de Reparo
SLA	Acordo de Nível de Serviço
GB	Gigabyte
TB	Terabyte
MB	Megabyte
IC	Intervalo de Confiança
I.S.	Índice de Sensibilidade

Sumário

1 Introdução

Sistemas de TIC são responsáveis por apoiar atividades crucias nos negócios modernos e, na maioria dos casos, precisam funcionar 24 horas por dia, 7 dias por semana (ROONEY; MCBRIDE; HANIF, 2008). No entanto, falhas de hardware ou software, desastres naturais (ex.: enchentes e furacões) ou ações humanas (ex.: incêndios e cyber ataques) podem ocorrer a qualquer momento, em qualquer empresa e sem aviso prévio (ANDRADE et al., 2017). Segundo uma pesquisa da Zetta (ZETTA, 2016), dentre as empresas pesquisadas, 67% poderiam ter prejuízo acima de US$ 20.000,00 por dia de *downtime*.

Com o objetivo de diminuir o impacto gerado pelas falhas inesperadas, empresas têm adotado estratégias de DR (ZETTA, 2016). No entanto, as estratégias de DR não são baratas. Especialmente para as empresas de pequeno e médio porte, as soluções de DR podem se tornar caras pois envolvem a aquisição de recursos computacionais que podem ser subutilizados devido a raridade dos desastres. Além disso, existem no mercado diversas soluções de DR que podem ser adotadas (ex.: migração de dados e replicação de ambientes) a depender das necessidades a serem supridas (ex.: disponibilidade e custo).

Nesse sentido, um relatório da Unitrends (UNITRENDS, 2018) apontou que médias e pequenas empresas ainda utilizam o backup como principal estratégia para proteção e recuperação de dados. O relatório ainda mostra que está havendo uma migração do backup em mídias físicas (ex.: fitas e discos) para utilização do backup em nuvem, o BaaS. A alta adoção do backup se dá principalmente pelo baixo custo envolvido na sua implantação. Contudo, mesmo que a utilização do BaaS exija um baixo investimento inicial quando comparado a outras estratégias de DR, é preciso considerar os prós e contras dessa estratégia para fins de DR.

Alguns trabalhos foram desenvolvidos com foco na análise de sistemas de backup. Yin *et al.* (YIN et al., 2012) apresentaram comparações de diferentes técnicas de backup. O estudo concluiu que um elevado número de rotinas de backup completo impacta negativamente na disponibilidade dos dados, contudo também apresenta menores índices de perda de dados. Xia *et al.* (XIA et al., 2014) apresentaram uma abordagem de modelagem analítica para garantir a proteção de dados dos sistemas considerando o menor impacto possível na disponibilidade e no desempenho. As análises mostraram que a execução de backup *offline* completo pode reduzir a disponibilidade dos dados e aumentar a rejeição de requisições. Nguyen *et al.* (NGUYEN; KIM; PARK, 2016) apresentaram um modelo para DCs tolerantes a falhas. Utilizando o modelo

proposto, os autores analisaram o custo do *downtime* e a disponibilidade do cenário adotado.

Diferentemente dos trabalhos citados anteriormente, neste analisaremos um ambiente de BaaS utilizando uma abordagem integrada de experimentos e modelos. Nós propomos modelos de redes de Petri estocásticas e determinísticas (DSPNs) (MARSAN; CHIOLA, 1987) para avaliar métricas-chave de recuperação de desastres, como disponibilidade, *downtime*, RPO e RTO. Os resultados revelam que a disponibilidade de um ambiente BaaS pode variar de acordo com a quantidade de dados dos backups ou restaurações que precisam ser realizadas pelo sistema. Além disso, através de uma análise de sensibilidade, identificamos os parâmetros que mais influenciam o RPO, RTO e disponibilidade em ambientes BaaS.

Este artigo está organizado da seguinte forma: no Capítulo 2 são apresentados conceitos para um melhor entendimento do trabalho. O Capítulo 3 discute a solução de BaaS adotada para este estudo. O Capítulo 4 detalha a configuração do ambiente de testes criado e a condução dos experimentos. O Capítulo 5 apresenta os modelos DSPN propostos para avaliar o ambiente BaaS adotado. O Capítulo 6 discute tanto os resultados obtidos pelos experimentos quanto pelos modelos DSPN. Por fim, a Capítulo 7 apresenta as conclusões e introduz os trabalhos futuros.

2 Fundamentos

2.1 Recuperação de Desastres e Backup-as-a-Service

A recuperação de desastres é a prática de tornar sistemas capazes de sobreviver a falhas inesperadas ou extraordinárias (REESE, 2009). Na análise de soluções de DR, duas métricas são primordiais, o RPO e o RTO. O RPO aponta a quantidade máxima de dados que pode ser perdida desde a realização do último backup até a ocorrência de uma falha, enquanto o RTO representa o tempo máximo necessário para a recuperação do serviço logo após uma interrupção inesperada. Assim, conhecer os valores dessas métricas pode auxiliar na escolha das melhores estratégias de DR para indivíduos ou empresas.

Apesar da crescente adoção de tecnologias de computação em nuvem na última década, várias empresas ainda adotam o backup como a principal estratégia de DR, principalmente devido ao baixo custo associado à sua implementação (UNITRENDS, 2018). Nesse cenário, o BaaS emerge com o objetivo facilitar o gerenciamento, a segurança e a proteção de dados provendo armazenamento em grandes DCs. O BaaS trata os procedimentos de backup e restauração de arquivos através de serviços de rede. Dessa forma, se faz necessária conexões de Internet capazes de prover uma baixa latência. O BaaS é comumente oferecido por nuvens públicas (ex.: Amazon Web services e Microsoft Azure), mas soluções em nuvens privadas também são possíveis de serem implementadas.

2.2 Modelagem e Avaliação utilizando DSPNs

Neste trabalho, utilizamos uma extensão das redes de Petri denominada *redes de Petri estocásticas e determinísticas* para modelar e avaliar um ambiente de BaaS. As DSPNs são um formalismo de modelagem usado para representar e analisar diferentes características de sistemas, como desempenho, disponibilidade, segurança e tolerância a falhas. Para análise de características de DR, a disponibilidade é uma métrica essencial. A disponibilidade de um sistema é dada pela probabilidade deste sistema estar disponível ao longo do tempo e pode ser calculada usando a Equação 2.1 (CASSANDRAS; LAFORTUNE, 2010):

$$A = \frac{MTTF}{MTTF + MTTR} \tag{2.1}$$

onde A é a disponibilidade resultante, o $MTTF$ (*Mean Time to Failure*) corresponde ao tempo médio para a ocorrência de falhas no sistema e o $MTTR$ (*Mean Time to Re-*

pair) corresponde ao tempo médio decorrido para reparo do sistema (CASSANDRAS; LAFORTUNE, 2010).

Outra métrica bastante utilizada em contratos de SLA é o tempo de inatividade ou *downtime*. O *downtime* representa quanto tempo um sistema ficou indisponível dentro de um determinado intervalo de tempo. O downtime (*D*) pode ser calculado em função do tempo de observação desejado (*T*) e da disponibilidade do sistema (*A*): $D = (1 - A) \times T$

3 Solução de BaaS adotada

Nesse capítulo, apresentamos o ambiente de BaaS adotado nesta pesquisa. Escolhemos avaliar uma solução que é bastante adotada e fácil de ser configurada por pequenas e médias empresas (UNITRENDS, 2018). A Figura 1 mostra a solução de DR adotada utilizando um ambiente BaaS. O ambiente é composto por um DC primário e um DC BaaS. O DC primário representa uma infraestrutura computacional já existente em uma empresa e é composto por quatro máquinas virtuais (VMs). Onde, das quatro VMs, duas hospedam uma aplicação de servidor web que utiliza um servidor de banco de dados localizado em outra VM (servidor BD). Por último, um balanceador de carga é responsável por distribuir as requisições dos usuários entre os dois servidores web, a fim de evitar a sobrecarga de um único servidor. O DC BaaS contém os servidores que oferecem serviços de backup e restauração de dados. Em caso de um desastre no DC primário, os dados das VMs podem ser perdidos. Nesse tipo de situação, os servidores de backup do DC BaaS executam uma restauração dos dados do último backup das VMs assim que o DC primário retornar ao estado operacional. Vale a pena ser ressaltado que neste trabalho consideramos apenas a política de backup completo.

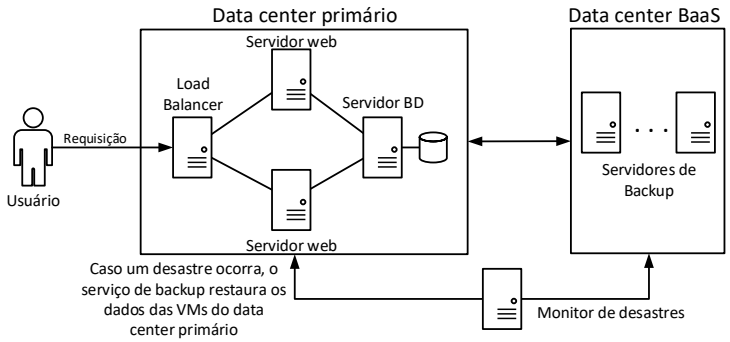

Figura 1 – Solução de BaaS adotada

Externamente ao DC primário e ao DC BaaS, um monitor de desastres verifica a conectividade entre ambos. O monitor de desastres é o responsável por ativar a operação de restauração de dados em caso de desastres no DC primário. Portanto, se o monitor de desastres detectar um evento deste tipo, que coloque todo o DC primário *offli*ne, ele ativa o serviço de restauração. Dessa forma, os dados são restaurados pelo DC BaaS depois que a infraestrutura do DC primário retornar ao funcionamento.

4 Arquitetura Experimental

Esse capítulo apresenta a configuração do ambiente de testes montado e detalha a execução dos experimentos no ambiente BaaS. O ambiente configurado representa a solução exibida na Figura 1. Na montagem do ambiente foram utilizados três *hosts* físicos. O primeiro *host*, representando o DC primário, possuía processador Intel Core i7-3770, 8 GB de RAM, 1 TB de HDD. Nele eram hospedadas as VMs referentes ao balanceador de carga, servidores web e de banco de dados, cada uma possuindo 1 v-CPU, 512 MB de RAM e 20 GB de HDD. O segundo *host*, representando o DC BaaS, foi configurado com processador Intel Core i7-3770, 4 GB de RAM e 1 TB de HDD. Ele hospedava as VMs dos servidores de backup (*Director* e *Storage*), configuradas com o software *Bacula Community* na versão 9.0.7 (Bacula Systems, 2018) e que possuíam 1 v-CPU e 512 MB de RAM. Porém, 5 GB de HDD para o *Director* e 600 GB de HDD para o *Storage*. O último *host*, representando o monitor de desastres, foi utilizado exclusivamente para monitorar e injetar falhas artificias no ambiente e possuía processador Intel Core i5-650, 4 GB de RAM e 500 GB de HDD. Todas as máquinas físicas utilizaram o *Debian* na versão 9.3 como sistema operacional e o KVM na versão 1.4.0 como software de virtualização.

Utilizando o ambiente descrito anteriormente, executamos os experimentos em duas fases. A primeira fase (1) tinha como objetivo observar o tempo médio necessário para se completar tarefas de backup e restauração de dados. Além disso, os resultados obtidos a partir desses experimentos foram utilizados como parâmetros de entrada para os modelos propostos. Já na segunda fase (2), executamos experimentos de injeção de falhas no ambiente a fim de comparar os resultados obtidos através dos experimentos com os resultados obtidos através dos modelos DSPN.

Na *fase 1*, foram executadas inicialmente 40 operações de backup completo e 40 operações de restauração de dados usando cargas de trabalho de 5 GB, 10 GB e 15 GB. Nós configuramos o conjunto de arquivos com tamanhos distintos para simular um ambiente real. Considerando a quantidade total de dados (5 GB, 10 GB e 15 GB), usamos: 20% para arquivos de 1 MB, 20% para arquivos de 3 MB, 20% para arquivos de 10 MB, 20% para arquivos de 50 MB e 20% para arquivos de 500 MB. Os arquivos utilizados no experimento foram gerados de forma aleatória por meio da ferramenta *data duplicator* (*dd*) (Unix Tutorial, 2019).

Na *fase 2*, utilizamos o *Eucabomber* (SOUZA et al., 2013) para gerar falhas e reparos artificiais na infraestrutura. Configuramos o experimento para funcionar por 480 horas (20 dias), gerando eventos de falhas e reparos com intervalos de tempo dis-

tribuídos exponencialmente. O tempo total do experimento representa um valor de 5,4 anos de simulação do sistema, pois foi utilizado um fator de redução de 100 nos valores dos tempos de falha e reparo dos componentes. A utilização do fator de redução visou adequar a duração do experimento em um tempo viável. A ferramenta *Eucabomber* foi instalada no *host* do monitor de desastres. Além disso, o monitor de desastres também era responsável por verificar a acessibilidade de cada VM do ambiente a cada cinco segundos usando o protocolo ICMP. Adicionalmente, se um desastre era detectado no DC primário, o monitor de desastre solicitava ao DC BaaS que restaurasse todos os dados das VMs no momento em que a infraestrutura do DC primário tivesse retomado as operações.

5 Modelos para a solução BaaS

Esse capítulo apresenta os modelos DSPN propostos para avaliação da solução de BaaS adotada. A Figura 2 exibe os modelos DSPN que representam o DC primário e o DC Baas. Na parte superior da Figura 2 são exibidos os modelos para o DC primário e seus componentes. A Figura 2 (a) apresenta o modelo DSPN para o DC primário. Nas Figuras 2 (b), (c), (d) e (e) são mostrados os modelos DSPN para a rede do DC primário, o balanceador de carga, os servidores web e o servidor de banco de dados, respectivamente.

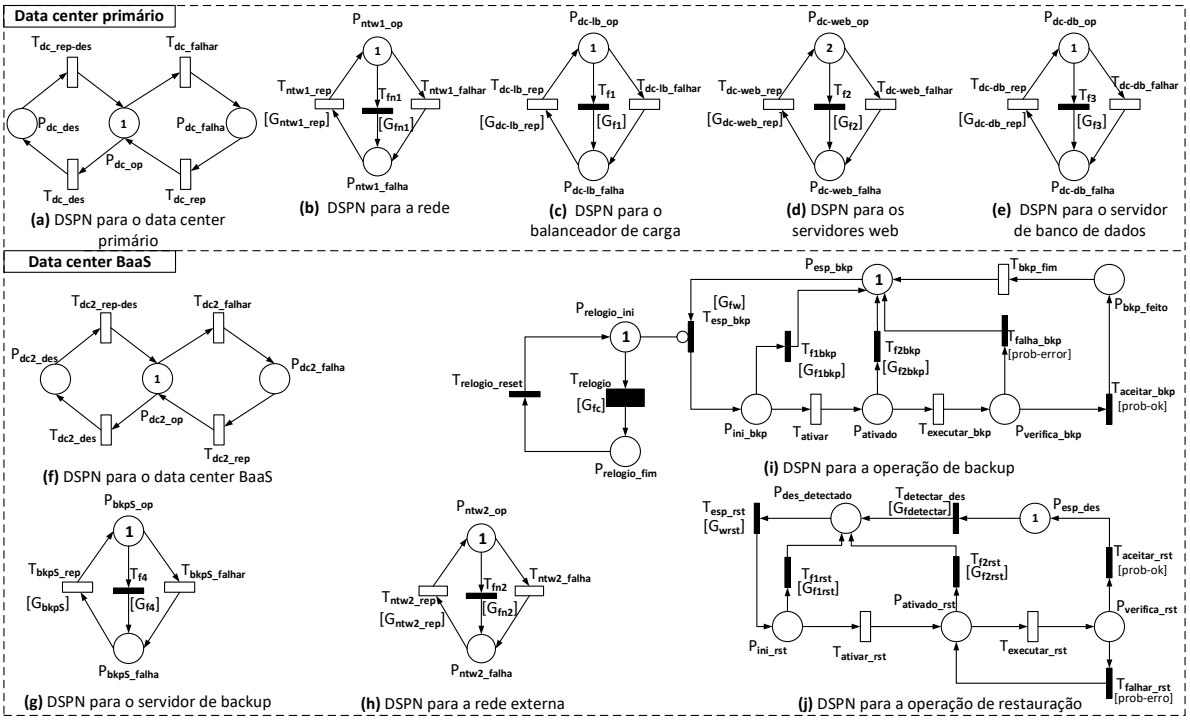

Figura 2 – Modelos DSPN para a solução de BaaS adotada

Para focar na avaliação de métricas relativas a recuperação de desastres, os modelos DSPN dos componentes reproduzem comportamentos de falha e reparo. Por exemplo, a Figura 2 (c) apresenta o comportamento de falha e reparo para o balanceador de carga. Neste modelo, um *token* presente no lugar $P_{dc\text{-}lb\_op}$ significa que o balanceador de carga está operacional. O caso contrário é indicado pela presença de um *token* no lugar $P_{dc\text{-}lb\_falha}$. As transições $T_{dc\text{-}lb\_falhar}$ e $T_{dc\text{-}lb\_rep}$ são responsáveis pela mudança desses estados. Além disso, a redundância de componentes é representada pelos números de *tokens* no modelo DSPN (ver Figura 2 (d)).

Diferentemente dos demais componentes, o modelo DSPN para o DC primário (Figura 2 (a)) possui duas transições representando falhas: uma representando falhas transientes (T_{dc_falhar}) e outra modelando eventos de desastre (T_{dc_des}). As falhas transi-

entes são falhas facilmente recuperáveis e exigem pouco tempo para serem corrigidas, enquanto que os eventos de desastre exigem mais tempo para serem reparados. Os modelos também consideram a dependência entre o DC primário e seus componentes. Portanto, quando o DC primário falha, as VMs que estão hospedadas neste DC também ficam indisponíveis. Nos modelos DSPN, essa dependência é modelada pelas funções de guarda G_{fn1}, G_{f1}, G_{f2} e G_{f3} presentes nas transições imediatas T_{fn1}, T_{f1}, T_{f2} e T_{f3}, respectivamente. No entanto, para o modelo DSPN da rede (ver Figura 2 *(b)*), a função de guarda G_{fn1} só é satisfeita em casos de eventos de desastres. A Tabela 1 exibe todas as funções de guarda dos modelos DSPN da Figura 2.

A parte inferior da Figura 2 mostra os modelos de DSPN para o DC BaaS e seus componentes. A Figura 2 (f) apresenta o modelo DSPN para o DC BaaS. As Figuras 2 (g) e (h) exibem os modelos DSPN, representando o servidor de backup e a rede externa, respectivamente. Além disso, as Figuras 2 (i) e (j) apresentam as DSPNs para as operações de backup e restauração, respectivamente. Esses modelos DSPN também consideram a dependência entre o DC BaaS e seus componentes. Consequentemente, quando o DC BaaS não estiver operacional, o servidor de backup também ficará indisponível. Além disso, para executar operações de backup e restauração, é necessário que o servidor de backup esteja operacional. Esse comportamento é modelado pelas funções de guarda G_{fw} e G_{wrst}, respectivamente (ver Tabela 1).

O modelo da operação de backup (Figura 2 (i)) inicia com um *token* no lugar P_{esp_bkp}. A transição T_{esp_bkp} é disparada quando a função de guarda G_{fw} é satisfeita

Tabela 1 – Funções de guarda dos modelos DSPN da Figura 2

Guarda	**Função de habilitação**
G_{f1}, G_{f2}, G_{f3}	$(\#P_{dc\_op} = 0)$
G_{fn1}	$(\#P_{dc\_des} = 1)$
$G_{dc\text{-}lb\_rep}$, $G_{dc\text{-}db\_rep}$, $G_{dc\text{-}web\_rep}$, G_{nt1_rep}	$(\#P_{dc\_op} = 1)$
G_{fc}	$(\#P_{bkpS\_op} = 1)$
$G_{fdetectar}$	$(\#P_{dc\_des} = 1)$
G_{fn2}	$(\#P_{dc2\_dis} = 1)$
G_{f4}	$(\#P_{dc2\_op} = 0)$
G_{bkpS_rep}, G_{nt2_rep}	$(\#P_{dc2\_op} = 1)$
G_{fw}	$(\#P_{bkpS\_op} = 1)$ AND $(\#P_{ntw2\_op} = 1)$ AND $(\#P_{ntw1\_op} = 1)$ AND $(\#P_{dc\_op} = 1)$ AND $(\#P_{dc\text{-}lb\_op} > 0)$ AND $(\#P_{dc\text{-}web\_op} > 0)$ AND $(\#P_{dc\text{-}db\_op} > 0)$
G_{f1bkp}, G_{f2bkp}	$(\#P_{bkpS\_op} = 0)$ OR $(\#P_{ntw2\_op} = 0)$ OR $(\#P_{ntw1\_op} = 0)$ OR $(\#P_{dc\_op} = 0)$ OR $(\#P_{dc\text{-}lb\_op} = 0)$ OR $(\#P_{dc\text{-}web\_op} = 0)$ OR $(\#P_{dc\text{-}db\_op} = 0)$
G_{wrst}	$(\#P_{bkpS\_op} = 1)$ AND $(\#P_{ntw2\_op} = 1)$ AND $(\#P_{ntw1\_op} = 1)$ AND $(\#P_{dc\_op} = 1)$ AND $(\#P_{dc\text{-}lb\_op} > 0)$ AND $(\#P_{dc\text{-}web\_op} > 0)$ AND $(\#P_{dc\text{-}db\_op} > 0)$
G_{f1rst}, G_{f2rst}	$(\#P_{bkpS\_op} = 0)$ OR $(\#P_{ntw2\_op} = 0)$ OR $(\#P_{ntw1\_op} = 0)$ OR $(\#P_{dc\_op} = 0)$ OR $(\#P_{dc\text{-}lb\_op} = 0)$ OR $(\#P_{dc\text{-}web\_op} = 0)$ OR $(\#P_{dc\text{-}db\_op} = 0)$

e quando não há *token* no lugar $P_{relogio\_ini}$. Note que a transição $T_{relogio}$ representa o comportamento da periodicidade das operações de backup (por exemplo, realizar um backup a cada 24 horas). No disparo da transição T_{esp_bkp} um *token* é consumido do lugar P_{esp_bkp} e um *token* é gerado no lugar P_{ini_bkp}. Em seguida, o backup pode ser ativado (T_{ativar}) ou uma falha pode ocorrer (T_{f1bkp}). Se o backup estiver ativo ($P_{ativado}$), a operação de backup pode ser executada ($T_{executar\_bkp}$) ou outra falha pode acontecer (T_{f2bkp}). As funções de guarda G_{f1bkp} e G_{f2bkp} verificam as falhas que podem tornar impraticável a operação de backup (por exemplo, falhas de conexão de rede). Supondo a realização do backup, um *token* é gerado no lugar $P_{verifica\_bkp}$. Neste lugar, há uma chance de falha (T_{falha_bkp}) ou sucesso ($T_{aceitar\_bkp}$) para a operação de backup. As incertezas de falha ou sucesso para a operação de backup são representadas através dos *pesos* dessas transições: *[prob-error]* e *[prob-ok]*.

A operação de restauração (Figura 2 (j)) funciona de forma semelhante à operação de backup. No entanto, a operação de restauração não tem uma periodicidade definida. As tarefas de restauração de dados apenas são realizadas em caso de um desastre no DC primário. A função guarda $G_{fdetectar}$ é responsável por monitorar a ocorrência de desastres no DC primário (ver Tabela 1).

6 Resultados e Discussões

6.1 Resultados dos experimentos

Primeiramente, apresentamos os resultados da *fase 1* dos experimentos realizados no *testbed* montado. Em seguida, discutimos os resultados da *fase 2* (injeção de falhas). Na *fase 1* obtivemos o tempo médio de execução das operações de backup e restauração de dados (ver Capítulo 4). Depois de coletar os valores dessas operações, calculamos o tempo médio por GB para executar um backup e o tempo médio por GB para restaurar os dados. Nós escolhemos adotar o tempo médio por GB para flexibilizar e tornar facilmente adaptável os modelos DSPN e as soluções propostas. A Tabela 2 exibe os resultados obtidos na *fase 1*. Esses resultados foram usados como parâmetros de entrada para os modelos DSPN (ver os parâmetros B_{tpGB} e R_{tpGB} descritos na Tabela 3). Além dos valores mostrados nesta tabela, nesta fase dos experimentos foram computadas as taxas de erro e sucesso das operações de backup e restauração. Essas taxas também foram utilizadas como parâmetro de entrada para os modelos através dos pesos *[prob-erro]* e *[prob-sucesso]* (ver Tabela 3). Vale mencionar que os resultados têm uma margem de erro de 5% com 95% de confiança (IC).

Na *fase 2* realizamos injeções de falhas artificiais no *testbed*, almejando confrontar os resultados desses experimentos com os resultados obtidos através dos modelos DSPN. A Tabela 3 exibe os parâmetros de entrada utilizados na *fase 2* e os parâmetros de entrada atribuídos a cada transição temporizada dos modelos DSPN da Figura 2. Na Tabela 4 (coluna Experimentos) são apresentados os resultados dos experimentos de injeção de falhas em relação as métricas de disponibilidade, RPO e RTO. A disponibilidade da solução foi calculada pela probabilidade de se ter um balanceador de carga, ao menos um servidor web, um servidor de banco de dados e a rede operando no DC primário. O RPO da solução foi obtido calculando o tempo médio decorrido entre o último backup válido e um evento de desastre. Por fim, o RTO foi calculado através do tempo médio decorrido entre um desastre e a recuperação completa do DC primário.

Nos experimentos, a solução BaaS alcançou uma disponibilidade de 0,9965243, variando entre 0,999496 e 0,9975440 para 95% de IC. Isso significa um *downtime*/ano médio de 30,447 horas. Já os valores médios computados nos experimentos para o RPO e RTO foram 21 e 22,791 horas, respectivamente. Considerando o IC de 95%, o RPO variou entre 13,896 e 28,103 horas, enquanto o RTO variou entre 15,438 e 30,144 horas.

Tabela 2 – Resultados dos experimentos da fase 1

Métrica	Parâmetro	Valor (*h*)
Média de tempo por *GB* para backup de dados	B_{tpGB}	0,0304
Média de tempo por *GB* para restauração de dados	R_{tpGB}	0,0253

Tabela 3 – Parâmetros usados nos experimentos (fase 2) e modelos DSPN

Parâmetro	Transição associada	Valor (h)
MTTF DCs (F_{dc1}, F_{dc2})	T_{dc_falhar}, T_{dc2_falhar}	8760
MTTR DCs (R_{dc1}, R_{dc2})	T_{dc_rep}, T_{dc2_rep}	4
MTTD[1] (D_{edc1}, D_{edc2})	T_{dc_des}, T_{dc2_des}	17520
MTTRD[2] (D_{rdc1}, D_{rdc2})	$T_{dc\_rep\text{-}des}$, $T_{dc2\_rep\text{-}des}$	12
MTTF servidor backup (F_{bs})	T_{bkpS_falhar}	8760
MTTR servidor backup (R_{rbs})	T_{bkpS_rep}	4
MTTF Balanceador de carga (LB_f)	$T_{dc\text{-}lb\_falhar}$	8760
MTTR Balanceador de carga (LB_r)	$T_{dc\text{-}lb\_rep}$	0.5
MTTF servidores Web/BD (VM_f)	$T_{dc\text{-}web\_falhar}$, $T_{dc\text{-}db\_falhar}$	2654
MTTR servidores Web/BD (VM_r)	$T_{dc\text{-}web\_rep}$, $T_{dc\text{-}db\_rep}$	1.25
MTTF redes (N_{fi}, N_{fe})	T_{ntw1_falhar}, T_{ntw2_falhar}	10000
MTTR redes (N_{ri}, N_{re})	T_{ntw1_rep}, T_{ntw2_rep}	1
(Des)ativar backup/restaração (A_t)	T_{ativar}, T_{ativar_rst}, T_{bkp_fim}	0.015
Tamanho Backup (B_s)	-	400 (GB)
Tempo médio de Backup	$T_{perform\_bkp}$	$B_s \times B_{tpGB}$
Tempo médio de Restauração	$T_{perform\_rst}$	$B_s \times R_{tpGB}$
Intervalo entre Backups (B_i)	$T_{relogio}$	24
Probabilidade de erro (*[prob-error]*)	T_{falhar_bkp}, T_{falhar_rst}	0.015 (%)
Probabilidade de sucesso (*[prob-ok]*)	$T_{aceitar\_bkp}$, $T_{aceitar\_rst}$	0.985 (%)

Tabela 4 – Resultados dos experimentos de injeção de falhas e DSPNs

Métrica	Experimentos	Modelos DSPN
RPO	21 h (13,896 - 28,103)	27,9026 h (27,9020 - 27,9033)
RTO	22,791 h (15,438 - 30,144)	24,2394 h (24,2158 - 24,2630)
Disponibilidade	0,9965243 (0,9949956 - 0,9975440)	0,997248 (0,997247 - 0,997249)

6.2 Análise dos modelos DSPN

Para a análise dos modelos DSPN propostos (ver Figura 2) foi adotada a ferramenta *Mercury* (OLIVEIRA et al., 2017). Os modelos DSPN utilizaram os mesmos parâmetros de entrada adotados no experimento de injeção de falhas (ver Tabela 3). As métricas computadas pelos modelos DSPN são similares às métricas do experimento anterior. A Tabela 5 exibe as expressões, na sintaxe do *Mercury*, usadas para calcular a disponibilidade, o RPO e o RTO dos modelos DSPN. Por fim, a Tabela 4 (coluna Modelos DSPN) apresenta os resultados obtidos através da simulação dos modelos DSPN.

Os resultados das simulações dos modelos DSPN revelaram uma disponibilidade de 0,997248 para a solução BaaS, variando entre 0,997247 e 0,997249 em um IC de 95%. Portanto, o *downtime*/ano médio computado foi de 24,11 horas. Já os re-

Tabela 5 – Expressões para calcular as métricas nos modelos DSPN

Métrica	Expressão
RPO	$(E\{\#P_{ini\_bkp}\} + E\{\#P_{ativado}\} + E\{\#P_{verifica\_bkp}\}) / (E\{\#P_{ini\_bkp}\} \times (1/A_t))$
RTO	$(E\{\#P_{des\_detectado}\}+E\{\#P_{ini\_rst}\}+E\{\#P_{ativado\_rst}\}+E\{\#P_{verifica\_rst}\}) / (E\{\#P_{ini\_rst}\} \times (1/A_t))$
Disponibilidade	$P\{(\#P_{dc\_op} = 1)$ AND $(\#P_{dc\text{-}lb\_op} = 1)$ AND $(\#P_{dc\text{-}web\_op} > 0)$ AND $(\#P_{dc\text{-}db\_op} = 1)$ AND $(\#P_{ntw1} = 1)\}$

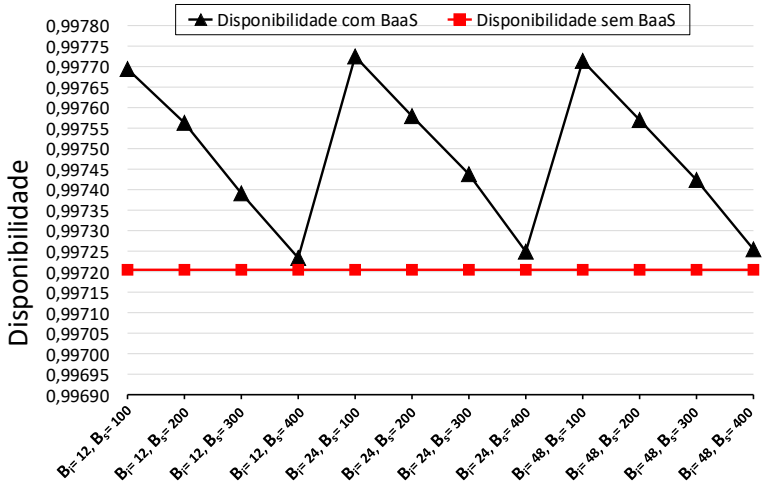

Figura 3 – Avaliação de disponibilidade em diferentes cenários utilizando BaaS

sultados para o RPO e RTO variaram um pouco mais em relação ao resultado dos experimentos. O RPO médio obtido pelos modelos DSPN foi de 27,9026 horas, variando entre 27,902 e 27,9033 horas, para um IC de 95%. Já o RTO médio foi de 24,2394 horas, variando entre 24,2158 e 24,2630 horas também em um IC de 95%. Comparando os resultados dos experimentos e dos modelos DSPN, é possível notar que são correspondentes, pois os intervalos de confiança dos resultados se sobrepõem. Dessa forma, não se pode afirmar que exista diferença estatística entre os resultados fornecidos pelos experimentos e pelos modelos DSPN.

Após verificar a eficácia dos modelos DSPNs propostos, cenários distintos foram analisados para demonstrar a flexibilidade dos modelos e encontrar situações que poderiam apresentar melhores resultados. A Figura 3 exibe um gráfico comparando os resultados de 12 cenários analisados. Nestes 12 cenários foram variados o intervalo da operação de backup (B_i = 12, 24 e 48 horas) e o tamanho do backup (B_s = 100, 200, 300 e 400 GB). O melhor resultado para a disponibilidade foi no cenário considerando um intervalo de backup de 24 horas e um tamanho de backup de 100 GB. Nesse cenário, a disponibilidade média alcançada do sistema foi de 0,9977247, representando um *downtime* médio de 19,931 horas/ano. Além disso, é possível notar que em todos os cenários testados, a disponibilidade do sistema foi mais alta quando o BaaS foi utilizado como estratégia de DR.

Já a Figura 4 apresenta a análise dos cenários para as métricas de RTO e RPO.

Tabela 6 – Top-3 Ranking de sensibilidade dos parâmetros em relação as métricas analisadas

	RPO		RTO		Disponibilidade	
Ordem	**Param.**	**I.S.**	**Param.**	**I.S.**	**Param.**	**I.S.**
1º	B_i	4.982E-1	D_{rdc1}	4.025E-1	D_{edc1}	2.080E-3
2º	B_{tpGB}	2.483E-1	R_{tpGB}	3.431E-1	F_{dc1}	9.021E-4
3º	B_s	2.209E-1	B_s	2.991E-1	R_{tpGB}	8.473E-4

Observando o gráfico nota-se uma relação dessas métricas com os parâmetros B_i e B_s. Os melhores resultados para o RTO aconteceram quando o tamanho do backup (B_s) foi composto de uma quantidade de dados menor (100 GB). Diferentemente, os melhores valores para o RPO ocorreram não só quando o intervalo de backup (B_i) foi curto (12 horas), mas também quando o sistema teve uma pequena quantidade de dados para backup/restauração. Em outras palavras, o RTO degradou à medida que a quantidade de dados para backup/restauração aumentou, enquanto o RPO degradou quando o intervalo de backup e a quantidade de dados para backup/restauração aumentaram.

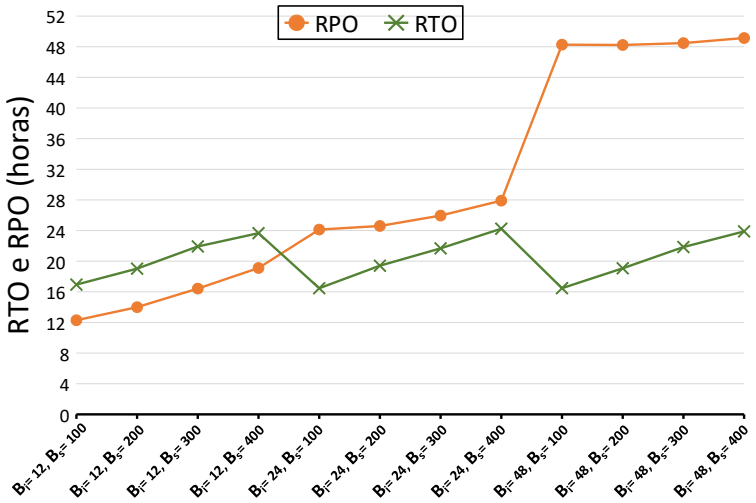

Figura 4 – Avaliação de RTO e RPO em diferentes cenários utilizando BaaS

Por fim, para identificar o comportamento dos parâmetros em relação as métricas, realizamos uma análise de sensibilidade utilizando a técnica de diferença percentual (Owen Hoffman; MILLER; DISCLAIMER, 1983) nos modelos DSPN. Essa técnica apresenta um índice de sensibilidade (I.S.) para cada parâmetro em relação com cada métrica. O I.S. é dado entre 0 e 1, onde 1 indica que um parâmetro tem a maior influência sobre a métrica e 0 a menor influência (Owen Hoffman; MILLER; DISCLAIMER, 1983). Dessa forma, variamos os valores dos parâmetros de entrada originais (ver Tabela 3) de -50% a +50% e calculamos o I.S. para cada um deles. A Tabela 6 exibe os resultados da análise de sensibilidade, mostrando em ordem decrescente os 3 parâmetros mais influentes sobre cada métrica.

7 Conclusões e Trabalhos Futuros

Este trabalho apresentou uma abordagem baseada em experimentos e modelagem para avaliar um ambiente de BaaS para fins de DR. Realizamos experimentos em um ambiente real de BaaS para calcular quatro importantes métricas de DR: disponibilidade, *downtime*, RPO e RTO. Utilizando o cenário adotado no estudo e os resultados dos experimentos, foram propostos modelos DSPN para análise de soluções BaaS. As análises realizadas através dos modelos DSPNs mostraram que os modelos propostos fornecem resultados correspondentes aos obtidos a partir dos experimentos conduzidos em um ambiente real de BaaS. Os resultados também indicaram que a adoção do BaaS melhorou o RTO quando a quantidade de dados de backup foi pequena. Em relação ao RPO, os resultados foram melhores quando se adotaram intervalos de backup menores. Destacamos ainda que houve uma melhora geral para todos os casos testados para a disponibilidade. Além disso, identificamos os parâmetros que mais influenciam as métricas RPO, RTO e a disponibilidade através da análise de sensibilidade. Nos trabalhos futuros, pretendemos considerar outras políticas de backup para o ambiente de BaaS.

Considerações Finais

Este trabalho foi publicado nos Anais do IX Simpósio Brasileiro de Engenharia e Sistemas Computacionais (QUEIROZ et al., 2019).

Referências

ANDRADE, E. et al. Availability modeling and analysis of a disaster-recovery-as-a-service solution. *Computing*, Springer, v. 99, n. 10, p. 929–954, 2017. Citado na página 8.

Bacula Systems. *Corporate Data Backup and Recovery Features in Bacula Enterprise Edition*. 2018. Disponível em: <https://www.baculasystems.com/architecture>. Citado na página 13.

CASSANDRAS, C.; LAFORTUNE, S. *Introduction to Discrete Event Systems*. 2nd. ed. [S.l.]: Springer Publishing Company, Incorporated, 2010. 800 p. ISBN 1441941193. Citado 2 vezes nas páginas 10 e 11.

MARSAN, M.; CHIOLA, G. On petri nets with deterministic and exponentially distributed firing times. In: *Advances in Petri Nets 1987*. [S.l.: s.n.], 1987. v. 266. ISBN 978-3-540-18086-9. Citado na página 9.

NGUYEN, T. A.; KIM, D. S.; PARK, J. S. Availability modeling and analysis of a data center for disaster tolerance. *Future Generation Computer Systems*, Elsevier, 2016. Citado na página 8.

OLIVEIRA, D. et al. Advanced stochastic petri net modeling with the mercury scripting language. In: ACM. *Proceedings of the 11th EAI International Conference on Performance Evaluation Methodologies and Tools*. [S.l.], 2017. p. 192–197. Citado na página 19.

Owen Hoffman, F.; MILLER, C. W.; DISCLAIMER, D. *Umcertainties in Environmental Radiological Assessment models and their Implications*. [S.l.], 1983. 57 p. Disponível em: <http://bit.ly/30XNchr>. Citado na página 21.

QUEIROZ, E. et al. Analisando o backup-as-a-service como uma estratégia de recuperaçao de desastres. In: SBC. *Anais Estendidos do IX Simpósio Brasileiro de Engenharia de Sistemas Computacionais*. [S.l.], 2019. p. 95–100. Citado na página 22.

REESE, G. *Cloud application architectures: building applications and infrastructure in the cloud*. [S.l.]: ”O’Reilly Media, Inc.”, 2009. Citado na página 10.

ROONEY, W. J.; MCBRIDE, G. E.; HANIF, T. IBM TotalStorage Productivity Center for Replication for z/OS. *IBM Systems Journal*, v. 47, n. 4, p. 681–694, 2008. Citado na página 8.

SOUZA, D. et al. EucaBomber: Experimental Evaluation of Availability in Eucalyptus Private Clouds. In: *2013 IEEE International Conference on Systems, Man, and Cybernetics*. [S.l.: s.n.], 2013. Citado na página 13.

UNITRENDS. *The State of Cloud and Data Protection 2018*. [S.l.], 2018. Disponível em: <https://bit.ly/2VsiieB>. Citado 3 vezes nas páginas 8, 10 e 12.

Unix Tutorial. *dd command*. 2019. Disponível em: <http://bit.ly/2Y5zHPU>. Citado na página 13.

XIA, R. et al. Performance and availability modeling of it systems with data backup and restore. *IEEE Transactions on Dependable and Secure Computing*, IEEE, 2014. Citado na página 8.

YIN, X. et al. Availability modeling and analysis for data backup and restore operations. In: *IEEE 31st Symposium on Reliable Distributed Systems (SRDS)*. [S.l.: s.n.], 2012. Citado na página 8.

ZETTA, I. *State of Disaster Recovery 2016*. 2016. <https://bit.ly/2H6TwhN>. [Online]. Citado na página 8.