

Modelagem Markoviana da Confiabilidade do Sistema Pix a partir de Dados Reais de Incidentes

Vitor Braço¹, Daniel Sadoc Menasché¹, Alberto Avritzer²

¹ Instituto de Computação – Universidade Federal do Rio de Janeiro (UFRJ)

²eSulabSolutions – Princeton, NJ, USA

Resumo. O sistema Pix, implementado pelo Banco Central do Brasil, é um dos principais meios de pagamento em tempo real no país, exigindo altíssima disponibilidade e desempenho. Neste artigo, propomos um modelo baseado em Cadeias de Markov de Tempo Contínuo (CTMC) para analisar a confiabilidade do sistema Pix com base em registros reais de incidentes. Os dados foram extraídos e classificados a partir de uma base contendo falhas operacionais, que foram codificadas em cinco estados distintos de operação e falha. A matriz de transição foi estimada empiricamente, e os tempos médios de recuperação foram comparados com os Acordos de Nível de Serviço (SLAs) definidos pelo Banco Central. Os resultados mostram que os tempos observados podem estar acima dos limites normativos estabelecidos, sugerindo a necessidade de melhorias nos mecanismos de recuperação automática.

Abstract. Pix is a real-time payment system developed by the Central Bank of Brazil, requiring strict availability and performance guarantees. This paper proposes a Continuous-Time Markov Chain (CTMC) model to assess the reliability of Pix based on real incident reports. A dataset comprising operational failures was pre-processed and classified into five distinct system states, including normal operation and various failure modes. Transition rates were estimated empirically, and average recovery times were compared with official Service Level Agreements (SLAs) published by the Central Bank. Results indicate that observed recovery durations may exceed regulatory limits, highlighting the need for improved automation mechanisms.

1. Introdução

Desde seu lançamento em novembro de 2020, o sistema Pix tornou-se o principal meio de pagamento no Brasil, sendo amplamente adotado tanto por pessoas físicas quanto jurídicas. O volume crescente de transações e a criticidade do serviço tornaram sua disponibilidade um requisito fundamental. No entanto, incidentes recentes chamaram a atenção da imprensa e da sociedade devido a falhas pontuais em sua operação.

Em 2023, uma falha generalizada em uma instituição participante causou a interrupção parcial do serviço para milhares de usuários por mais de uma hora, impossibilitando transações de envio e recebimento de Pix [G1 2023]. Outro caso ocorreu em dezembro do mesmo ano, quando instabilidades na camada de consulta de chaves do DICT resultaram em atrasos significativos em pagamentos via QR Code [Banco Central do Brasil 2024]. Em 2025, também foram reportadas reportadas falhas em larga escala.^{1 2} Essas ocorrências reforçam a importância de

¹<https://g1.globo.com/economia/noticia/2025/02/07/problemas-para-pagamento-com-Pix-em-diversos-bancos.ghtml>

²<https://epocanegocios.globo.com/brasil/noticia/2025/03/Pix-fora-do-ar-hoje-usuarios-relatam-instabilidade-no-sistema.ghtml>

modelos quantitativos que permitam analisar, prever e mitigar os impactos da indisponibilidade do sistema.

Neste contexto, este trabalho apresenta três contribuições principais:

1. **Análise de dados reais.** Coletamos e analisamos dados reais de incidentes relacionados ao sistema Pix, observados do ponto de vista de uma empresa intermediária entre o usuário final e o Banco Central;
2. **Modelo analítico.** Propomos um modelo baseado em Cadeias de Markov de Tempo Contínuo (CTMC) para representar o comportamento do sistema Pix, distinguindo entre falhas leves, falhas severas e mecanismos de recuperação;
3. **Parametrização do modelo com dados reais e análise.** Desenvolvemos uma metodologia prática de parametrização do modelo a partir de dados reais de operação, utilizando regras heurísticas para codificação de estados e estimativas empíricas de transição.

Cabe destacar que o sistema Pix possui requisitos rígidos de disponibilidade e desempenho [Banco Central do Brasil 2023, Banco Central do Brasil 2022]. Neste trabalho, usando modelos, também visamos avaliar com dados reais a confiabilidade operacional do Pix, estimando métricas como tempo médio para reparo (MTTR) e comparando com os requisitos.

O restante deste artigo está estruturado da seguinte forma. Na Seção 2, descrevemos a base de incidentes utilizada. A Seção 3 apresenta a construção da cadeia de Markov e o processo de estimativa da matriz de transição é apresentado na Seção 4. A Seção 5 compara os tempos médios de recuperação com os SLAs do Manual de Tempos do Pix. Por fim, as Seções 6 e 7 discutem trabalhos relacionados e conclusões.

2. Introdução ao Pix e Base de Dados

Nesta seção, apresentamos a terminologia e descrevemos nossa base de dados de incidentes.

2.1. Terminologia e Funcionamento do Sistema Pix

O fluxo típico de uma transação Pix envolve três partes principais:

- **Usuário final:** pessoa física ou jurídica que inicia ou recebe um pagamento via Pix, geralmente por meio de um aplicativo bancário.
- **Instituição financeira (ou de pagamento):** intermediária entre o usuário e a infraestrutura do Pix. É responsável por validar a transação, coletar os dados necessários (como chave Pix ou QR Code) e comunicar-se com o Banco Central.
- **Banco Central (Bacen):** provê a infraestrutura central de liquidação e registro das transações, operando dois sistemas principais: o SPI e o DICT.

O **SPI** (Sistema de Pagamentos Instantâneos) é a espinha dorsal do Pix. Trata-se de uma infraestrutura de liquidação operada pelo Banco Central, que realiza a compensação e liquidação final dos pagamentos entre contas de diferentes instituições, em tempo real e com alta disponibilidade.

O **DICT** (Diretório de Identificadores de Contas Transacionais) é o componente responsável por manter o mapeamento entre as chamadas *chaves Pix* (como CPF, CNPJ, telefone, e-mail ou chave aleatória) e as contas transacionais de destino.

Quando um usuário digita uma chave Pix no aplicativo, a instituição consulta o DICT para descobrir os dados da conta associada. Em seguida, o pagamento pode ser encaminhado corretamente via SPI.

Relação entre os Componentes. O sistema funciona de forma integrada: 1) O usuário inicia um Pix (ex: escaneando um QR Code); 2) A instituição intermediária consulta o DICT para obter a conta de destino; 3) A transação é enviada ao SPI para liquidação; 4) O resultado da transação é devolvido à instituição e ao usuário.

Essa arquitetura exige alta disponibilidade e baixa latência tanto na resolução de chaves (DICT) quanto na liquidação (SPI), sendo ambos monitorados por SLAs rígidos definidos pelo Banco Central [Banco Central do Brasil 2023].

2.2. Dados Disponíveis e Análise Qualitativa dos Incidentes

A base de dados consiste de registros de incidentes do ponto de vista de uma instituição de pagamento, ocorridos entre janeiro de 2024 e janeiro de 2025. Alguns incidentes se referem a questões internas à instituição e outros se referem a questões externas.

Utilizamos uma planilha para organizar os dados dos incidentes. Cada linha representa um incidente e inclui informações como: 1) Descrição e causa raiz; 2) Classe (tipo de falha); 3) Tempo para recuperação; 4) Tipo de intervenção (manual ou automática); 5) Possibilidade de rejuvenescimento.

As descrições textuais dos incidentes revelam padrões recorrentes de falha em componentes diversos da infraestrutura do Pix, conforme brevemente ilustrado a seguir:

- **Falhas de infraestrutura de rede ou comunicação:** Incluem perda de pacotes, problemas de roteamento, falhas em links de operadoras ou perda de conectividade entre serviços internos. Ex: “Falha de comunicação com o Bacen” e “Problemas com Pix dinâmico QR Code”.
- **Indisponibilidade ou lentidão no DICT/BACEN:** Diversos incidentes estão relacionados à degradação ou interrupção do serviço de consulta de chaves, cuja operação depende do DICT. Ex: “Pix - Falha na consulta de chaves (DICT)” e “Instabilidade no envio transacional do DICT”.
- **Problemas internos de aplicação ou banco de dados:** Compreendem erros de timeout, contenções, broker de mensagens degradado, jobs mal dimensionados e concorrência em bancos relacionais. Ex: “Timeout no DB” e “Falha em cadastro de clientes”.
- **Automação incompleta ou falhas em recuperação automatizada:** Referem-se a casos nos quais mecanismos de failover, health check ou escalonamento automático não funcionaram corretamente. Ex: “AWS ASG” e “Reinício não disparado apesar de alta carga”.
- **Degradações de performance sem interrupção completa:** Incluem queda de volume, aumento de tempo de resposta ou lentidão esporádica. Ex: “Queda no volume de envio Pix” e “Atraso na consulta de saldo”.

3. Modelo Markoviano

Definimos os seguintes estados para o sistema Pix:

Assumimos que, antes de cada incidente, o sistema estava operando normalmente (estado S_0). Assim, para cada incidente registrado, modelamos dois tipos de transições:

1. $S_0 \rightarrow S_i \rightarrow S_0$ (início da falha ou manutenção e conclusão após tempo de permanência no estado S_i);
2. $S_0 \rightarrow S_i \rightarrow S_j \rightarrow S_0$ (início da falha ou manutenção e conclusão após escalonamento para nível superior, com recuperação após tempo de permanência nos estados S_i e S_j).

Tabela 1. Definição dos Estados da Cadeia de Markov

Código	Estado	Interrupção
0	Operação normal	-
1	Degradação de desempenho	mínima
2	Indisponibilidade total	total
3	Manutenção com reestabelecimento automático	parcial
4	Manutenção com reestabelecimento manual	parcial

A partir dessas transições, estimamos a matriz de taxas de transição Q , onde:

$$q_{ij} = \frac{\text{número de transições de } i \rightarrow j}{\text{tempo total gasto no estado } i}, \quad i \neq j \quad (1)$$

$$\text{e } q_{ii} = -\sum_{j \neq i} q_{ij}.$$

Geração da Matriz de Transição a partir da Tabela Final. Com base na tabela final construída a partir dos incidentes analisados, foram extraídas explicitamente as durações e ocorrências das seguintes transições relevantes para o modelo markoviano:

- $1 \rightarrow 0$ — Recuperação direta de degradação;
- $1 \rightarrow 2$ — Escalonamento de degradação para falha total;
- $2 \rightarrow 0$ — Recuperação a partir de falha total;
- $3 \rightarrow 0$ — Recuperação automática bem-sucedida;
- $3 \rightarrow 4$ — Automação falha, escalando para manutenção manual;
- $4 \rightarrow 0$ — Recuperação manual.

A matriz de transição de taxas Q foi então construída conforme ilustrado na Figura 1.

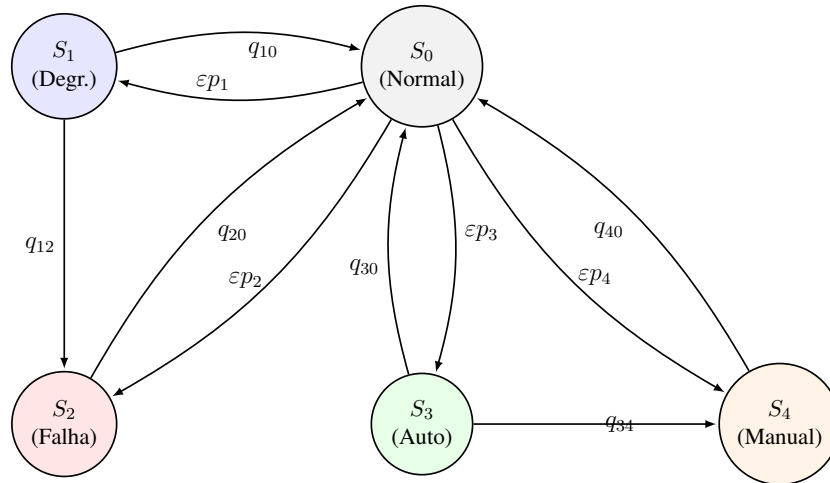


Figura 1. Diagrama da Cadeia de Markov com os estados operacionais e de falha do sistema Pix.

3.1. Ajuste de Transições a partir do Estado Normal

Para representar o comportamento do sistema em operação normal, o modelo markoviano foi ajustado para incluir uma taxa constante de entrada em estados de falha ou manutenção. O

estado S_0 , correspondente à operação normal, possui agora uma linha especial na matriz Q definida por:

$$Q_{0j} = \varepsilon \cdot p_j, \quad Q_{00} = -\varepsilon$$

onde:

- ε é a taxa de saída do estado normal, S_0 ;
- p_j é a probabilidade empírica de entrada no estado de falha S_j , calculada com base nas amostras da planilha.

A primeira linha da matriz Q resultante é:

$$Q_0 = \varepsilon \cdot \begin{bmatrix} -1 & p_1 & p_2 & p_3 & p_4 \end{bmatrix}.$$

Esse ajuste permite simular a ocorrência espontânea de falhas durante a operação do sistema de forma coerente com os dados históricos.

Como Determinar ε ? O valor de $1/\varepsilon$ caracteriza o tempo médio de permanência no estado normal (S_0), por visita. Sua escolha influencia diretamente a frequência com que o sistema simulado entra em falha. Apresentamos a seguir quatro abordagens possíveis para definir esse valor:

- **Abordagem 1 — Sobrevivência (*Survivability*):** assume-se que o sistema já inicia em estado de falha. O parâmetro ε é definido como um valor muito alto (no limite teórico, ele é definido como tendendo ao infinito), de modo a representar que o tempo de detecção de falha é desprezível e que o objetivo principal é modelar o comportamento de recuperação.
- **Abordagem 2 — Tempo entre falhas muito longo:** considera-se que o sistema permanece funcional por longos períodos entre falhas. Neste caso, ε é fixado como um valor pequeno (por exemplo, $\varepsilon = 10^{-4}$ ou inferior), modelando cenários de alta confiabilidade.
- **Abordagem 3 — Estimativa baseada em dados reais:** se estiverem disponíveis registros suficientemente grandes e confiáveis de tempo médio entre falhas (MTBF), é possível estimar ε como o inverso desse tempo. No entanto, essa abordagem requer acesso a uma base mais ampla de incidentes, o que não foi possível neste estudo.
- **Abordagem 4 — Análise de sensibilidade:** explora-se uma faixa de valores de ε e avalia-se o impacto nas métricas globais do sistema (por exemplo, disponibilidade ou taxa de incidentes por hora). Esta abordagem é útil para análises exploratórias e cenários hipotéticos.

Neste trabalho, consideramos as abordagens 1 e 4. A primeira permite analisar o comportamento do sistema sob a perspectiva de sobrevivência, isto é, assumindo que ele já se encontra em falha. A quarta abordagem será utilizada para explorar, via simulação, o impacto de diferentes valores de ε nos indicadores de desempenho, como disponibilidade e tempo médio de recuperação.

3.2. Algoritmo para Construção da Matriz Q

1. Para cada incidente registrado, considera-se uma transição do estado normal S_0 para o estado de falha correspondente S_i ;
2. Transições do tipo $S_0 \rightarrow S_i \rightarrow S_0$ (início da falha ou manutenção e conclusão após tempo de permanência no estado S_i);

- O tempo de permanência em S_i , nos casos em que o sistema transiciona de S_0 diretamente para S_i e, em seguida, de S_i diretamente para S_0 , é extraído da coluna “tempo para reparo” da base de dados (convertido para minutos);
 - Este tipo de incidente gera duas transições, $S_0 \rightarrow S_i \rightarrow S_0$;
3. Para cada estado S_i , calcular o tempo médio de permanência no estado S_i , por visita a S_i , denotado por m_i , usando os tempos colhidos na etapa acima;
 4. Transições do tipo $S_0 \rightarrow S_i \rightarrow S_j \rightarrow S_0$ (início da falha ou manutenção e conclusão após tempo de permanência nos estados S_i e S_j);
 - Para incidentes que resultam em três transições consecutivas — isto é, da forma $S_0 \rightarrow S_i \rightarrow S_j \rightarrow S_0$ —, o tempo total de recuperação observado é distribuído entre os estados intermediários S_i e S_j de forma proporcional aos seus tempos médios de permanência estimados.
 - Especificamente, suponha que os tempos médios de permanência previamente estimados nos estados S_i e S_j sejam m_i e m_j , respectivamente. Dado um incidente com tempo total de recuperação r , a duração atribuída ao estado S_i será $r \cdot m_i / (m_i + m_j)$, enquanto a duração correspondente ao estado S_j será $r \cdot m_j / (m_i + m_j)$.
 - Essa aproximação preserva a proporção entre os tempos médios característicos dos estados, respeitando a duração total observada do incidente.
 5. A taxa de transição q_{ij} é calculada como:

$$q_{ij} = \frac{N_{ij}}{T_i}$$

onde N_{ij} é o número de transições observadas de i para j , e T_i é o tempo total de permanência no estado i , levando em conta todas as visitas ao estado i .

4. Classificação de Estados a partir de Incidentes

Nesta seção, consideramos o problema de mapear os incidentes reais observados em estados do modelo markoviano. Para tal, consideramos uma heurística, descrita a seguir. Embora a heurística seja simples, e possa ser aprimorada em trabalhos futuros, ela serve para ilustrar a viabilidade de traduzir dados colhidos a partir de incidentes reais para modelos analíticos que possam ser usados para fazer estudos de caso e extrapolações para novos cenários.

4.1. Heurística para Mapear Incidentes Reais em Estados do Modelo

A codificação do estado de falha foi feita com base nas seguintes regras:

- S_1 , se a falha corresponde a um problema de concorrência de banco de dados ou, de forma mais geral, a um problema de impacto leve;
- S_2 , se a falha corresponde a exceção não tratada;
- S_3 , se há degradação de desempenho, com possibilidade de manutenção automática e/ou recuperação por rejuvenescimento;
- S_4 , se nenhuma das condições acima for satisfeita, por exemplo, se há degradação, mas sem possibilidade de automação ou rejuvenescimento.

4.2. Estimativas de Disponibilidade: Curto vs. Longo Prazo

Consideremos um intervalo de missão de 8 horas (por exemplo, o período de operação bancária entre 9h e 17h). Sejam λ a taxa de falhas totais por hora e $\tau = 8$ o tempo da missão. Com base em 7 incidentes com indisponibilidade total observados ao longo de um ano, estimamos $\lambda = 7 / (365 \cdot 24)$. A seguir, apresentamos duas abordagens clássicas para estimar a disponibilidade do sistema [Trivedi 2001]:

Tabela 2. Número de ocorrências observadas para cada estado (base de dados). Nesta tabela, levamos em conta apenas o primeiro estado visitado após o estado S_0 , por incidente. Temos um total de 38 incidentes.

Estado	# Ocorrências
1 — Degradação de desempenho	8
2 — Indisponibilidade total	7
3 — Manutenção com reestabelecimento automático	14
4 — Manutenção com reestabelecimento manual	9
Total de incidentes	38

- **Confiabilidade (*reliability*):** probabilidade de o sistema não falhar durante uma missão de duração τ , assumindo tempo entre falhas exponencial, conforme estabelecido no trabalho seminal [Musa 1975]. A expressão é dada por:

$$R = \mathbb{P}(\text{sem falha durante janela } \tau) = \exp(-\lambda\tau) = \exp\left(-\frac{7}{365 \cdot 24} \cdot 8\right) \approx 0,9936.$$

- **Disponibilidade (*availability*):** fração do tempo em que o sistema está disponível,

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

onde $MTTF$ (Mean Time To Failure) é o tempo médio até falha e $MTTR$ (Mean Time To Repair) é o tempo médio até recuperação. Com $MTTF = \frac{365 \cdot 24}{7} \approx 75085$ minutos e $MTTR = 30$ minutos (vide Seção 4.5):

$$A = 75085 / (75085 + 30) \approx 0,9996.$$

Enquanto a confiabilidade corresponde à probabilidade de o sistema operar sem falhas durante uma janela de tempo específica — sendo, portanto, uma *métrica de intervalo* (*interval measure*) —, a disponibilidade é uma *medida pontual* (*point measure*) que representa a fração do tempo em que o sistema está operacional no regime estacionário. Quando a propriedade PASTA é satisfeita, a disponibilidade também corresponde à probabilidade de um usuário, e.g., que chega segundo um fluxo de Poisson, encontrar o sistema disponível [Trivedi 2001].

4.3. Distribuição dos Estados de Falha

Com base em regras heurísticas definidas previamente, cada incidente foi classificado em um dos seguintes estados: S_1 (degradação leve), S_2 (falha total), S_3 (recuperação automática) ou S_4 (recuperação manual). A Tabela 2 apresenta a contagem observada.

As probabilidades observadas foram:

$$p_1 = 21.1\%, \quad p_2 = 18.4\%, \quad p_3 = 36.8\%, \quad p_4 = 23.7\%$$

A Figura 2(a) ilustra tais probabilidades. A Figura 2(b) indica o tempo médio até recuperação, ou seja, até retornar ao estado S_0 , em função do estado inicial, com intervalos de confiança de 95%.

A ocorrência de transições para o estado S_4 indica que, embora o sistema conte com mecanismos de automação, muitos incidentes ainda requerem recuperação manual. Além disso, a

presença significativa de degradações leves e indisponibilidades totais reforça a importância de se tratar de forma diferenciada a severidade e escalabilidade dos impactos.

Essa análise qualitativa fundamenta a parametrização do modelo markoviano e ajuda a estabelecer prioridades de mitigação na engenharia do sistema.

4.4. Matriz de Taxas

A matriz resultante (valores em **transições por minuto**) é:

$$Q = \begin{bmatrix} -\varepsilon & \varepsilon p_1 & \varepsilon p_2 & \varepsilon p_3 & \varepsilon p_4 \\ 0.0203 & -0.0231 & 0.0029 & 0 & 0 \\ 0.0195 & 0 & -0.0195 & 0 & 0 \\ 0.0213 & 0 & 0 & -0.0243 & 0.0030 \\ 0.0294 & 0 & 0 & 0 & -0.0294 \end{bmatrix}$$

Essa matriz reflete a dinâmica observada nos dados reais, incluindo diferentes caminhos de recuperação e escalonamento de falhas.

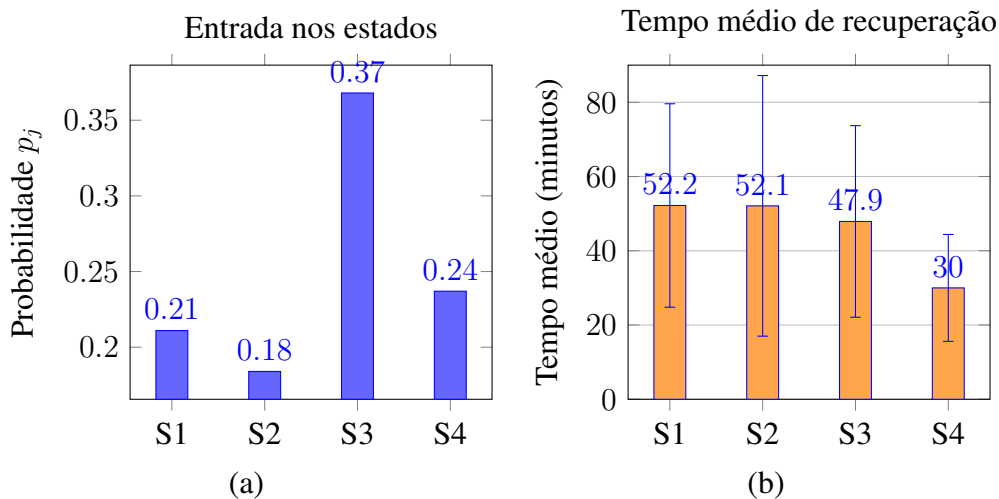


Figura 2. (a) Probabilidade de entrada em estados de falha; (b) tempo médio de recuperação em função do primeiro estado visitado durante incidente.

4.5. Análise de Sobrevivência (*Survivability*)

Nesta seção, conduzimos uma análise de *survivability*, ou seja, o comportamento do sistema após o início de uma falha (Figura 3). Assumimos que o sistema já se encontra em estado de falha, e que o objetivo é modelar o tempo necessário até a recuperação. Para isso, consideramos o limite $\varepsilon \rightarrow \infty$, implicando que o tempo de permanência no estado normal S_0 é desprezível. Consequentemente, a primeira linha da matriz de transição Q (correspondente ao estado S_0) torna-se irrelevante para esta análise, já que assumimos que a falha já ocorreu e o sistema encontra-se em um dos estados S_1 a S_4 .

Tempo médio por visita a cada estado. O primeiro gráfico da figura 3 (a) exibe o tempo médio de permanência em cada estado de falha, calculado diretamente a partir do inverso do valor absoluto da entrada da diagonal da matriz Q . Ou seja, para um estado S_j , temos:

$$\text{Tempo médio por visita a } S_j = \frac{1}{-q_{jj}}$$

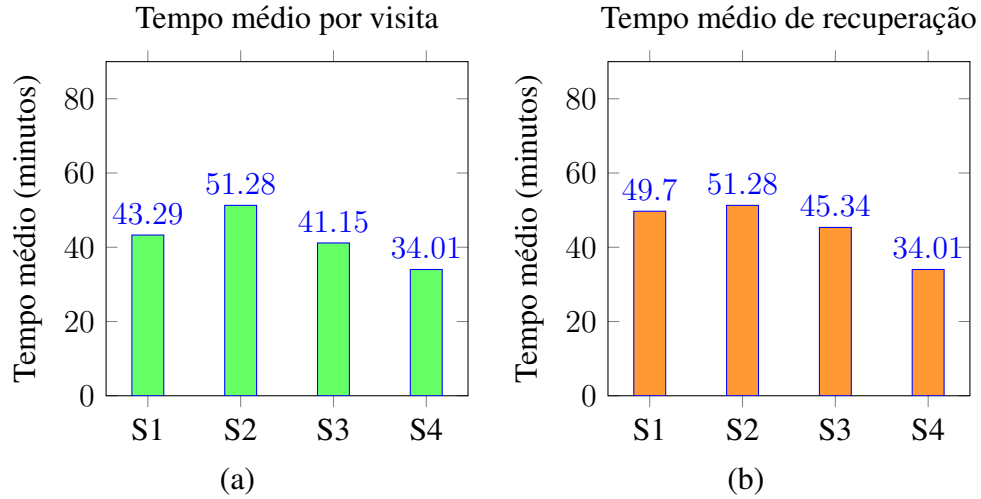


Figura 3. Gráficos obtidos diretamente a partir da matriz Q : (a) tempo médio de visita por estado; (b) tempo médio de recuperação em função do primeiro estado visitado durante incidente.

Tempo médio total até recuperação. Alguns incidentes exigem múltiplas transições antes da recuperação completa. Para esses casos, é necessário considerar o tempo adicional decorrente de transições intermediárias. O segundo gráfico mostra o tempo médio total de recuperação, considerando essas transições. Em particular:

- Para S_1 : além do tempo médio no próprio estado, há uma chance de escalonamento para S_2 . A correção aplicada foi:

$$43,29 + \frac{0,0029}{0,0203 + 0,0029} \cdot 51,28 = 49,70 \text{ minutos}$$

- Para S_3 : há possibilidade de escalonamento para S_4 , então usamos:

$$41,15 + \frac{0,0030}{0,0030 + 0,0213} \cdot 34,01 = 45,34 \text{ minutos}$$

- Para S_2 e S_4 , não há transições intermediárias, e os tempos permanecem inalterados.

A diferença observada entre os tempos médios apresentados na Figura 2(b) e na Figura 3(b) advém da forma como os dados foram tratados em cada abordagem. Na Figura 2(b), os tempos de recuperação foram extraídos diretamente da base de incidentes, considerando o tempo total desde a entrada no primeiro estado de falha até o retorno ao estado normal (S_0), sem desagregar os tempos de permanência em estados intermediários. Em outras palavras, a média é feita diretamente sobre os incidentes como um todo, tratando cada recuperação como uma “caixa preta”.

Já na Figura 3(b), a estimativa do tempo médio de recuperação foi feita com base exclusivamente na matriz de taxas Q , cuja construção foi detalhada na Seção 3.2. Nessa abordagem, a duração total de cada incidente é decomposta nos tempos de permanência individuais de cada estado visitado, mesmo nos casos em que o sistema percorre múltiplos estados antes de retornar a S_0 . Isso permite incorporar, de forma mais refinada, o impacto das transições intermediárias (como escalonamentos) nos tempos totais de recuperação. Por exemplo, os tempos associados a S_1 e S_3 foram ajustados para refletir a probabilidade de escalonamento para S_2 e S_4 , respectivamente, conforme demonstrado anteriormente. Essa diferenciação metodológica explica as discrepâncias entre os tempos médios observados nas duas figuras.

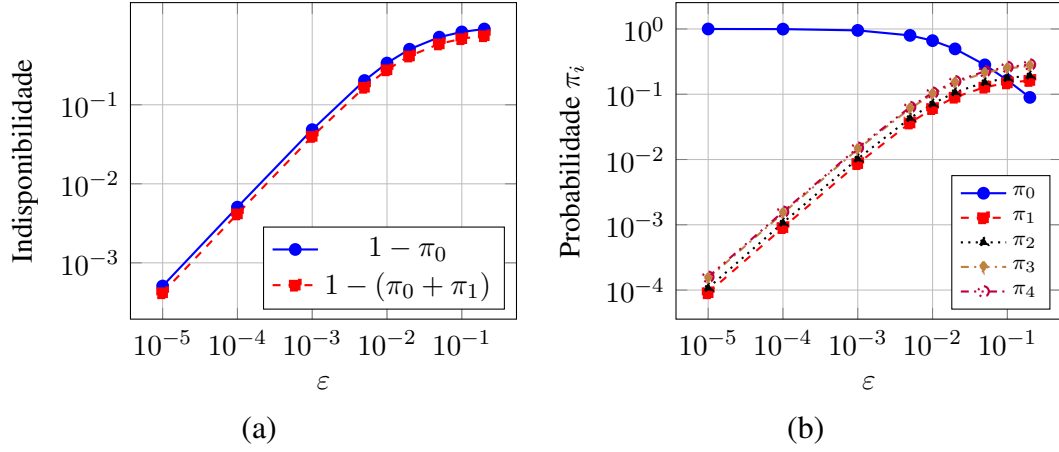


Figura 4. (a) Indisponibilidade em função de ε com dois critérios; (b) probabilidades estacionárias π_i dos estados S_0 a S_4 (cadeia de Markov apresentada na Figura 1).

4.6. Análise de Sensitividade

A Figura 4 foi construída com base em simulações estacionárias da cadeia de Markov descrita anteriormente, variando-se o parâmetro ε e calculando-se a distribuição de probabilidade de estados no regime estacionário. Diferentemente da análise de *survivability* realizada na seção anterior — onde assumimos $\varepsilon \rightarrow \infty$ para modelar apenas o comportamento de recuperação após falha — aqui consideramos valores finitos de ε , que representam diferentes intensidades de ocorrência de falhas ao longo do tempo. Para cada valor de ε , computamos a probabilidade estacionária π_0 de o sistema estar no estado S_0 (operação normal), bem como a soma $\pi_0 + \pi_1$, onde S_1 representa degradação leve sem impacto direto ao usuário. A Figura 4(a) exibe a indisponibilidade do sistema, definida como $1 - \pi_0$ ou $1 - (\pi_0 + \pi_1)$, respectivamente, em escala logarítmica, evidenciando a sensibilidade do sistema à frequência de falhas. A Figura 4(b) ilustra a distribuição estacionária completa.

Observa-se que, para $\varepsilon < 10^{-5}$, a indisponibilidade permanece abaixo de 10^{-3} (ou seja, menor que 0,1%), o que é compatível com os requisitos estabelecidos pelo Manual de Tempos do Pix. Como ε representa a taxa de entrada em falha por minuto, esse valor implica que falhas ocorrem, em média, a cada 100.000 minutos — ou aproximadamente uma falha a cada 70 dias. À medida que ε aumenta, a fração de tempo em que o sistema permanece em estados distintos de S_0 cresce rapidamente. No entanto, ao considerar que o estado S_1 ainda representa uma condição aceitável de operação (sem impacto perceptível ao usuário), a disponibilidade percebida melhora substancialmente. Essa distinção é essencial para análises realistas de SLA, como discutido na próxima seção.

5. Comparação com os SLAs do Manual de Tempos do Pix

O Manual de Tempos do Pix estabelece metas de desempenho e disponibilidade para os diversos serviços relacionados ao ecossistema do Pix. Estes requisitos de qualidade de serviço (SLAs) servem como parâmetros normativos com os quais é possível comparar os tempos médios de recuperação estimados por meio do modelo markoviano desenvolvido neste trabalho.

5.1. Disponibilidade do SPI e do DICT

De acordo com o documento oficial, o **índice de disponibilidade do SPI** deve ser de, no mínimo, **99,9%**, considerando uma janela móvel de três meses. Já o **índice de disponibilidade do DICT** estabelece metas de **99,9% para consultas** e **99,8% para atualizações**.

Considera-se uma janela móvel de três meses. Assim, essa meta de 99,9% de disponibilidade equivale, em termos práticos, a uma tolerância máxima de indisponibilidade mensal, que pode ser calculada como:

$$\text{Indisponibilidade máxima} = (1 - 0,999) \times 30 \times 24 \times 60 = 43,2 \text{ minutos/mês}$$

Ou seja, o sistema SPI (e, por extensão, os serviços críticos do ecossistema Pix) devem apresentar indisponibilidade total inferior a **43 minutos por mês**, sob risco de descumprir o Acordo de Nível de Serviço (SLA) estabelecido pelo Banco Central. Para efeito de mensuração, considera-se indisponibilidade qualquer situação na qual um percentual igual ou superior a 80% das requisições de liquidação (SPI) ou consultas/atualizações (DICT) seja respondido com erros causados pelo sistema e que persista por mais do que **36 segundos** (SPI) ou **30 segundos** (DICT).

Tempos de resposta para consultas e atualizações. O tempo máximo esperado para resposta em operações realizadas via DICT está descrito em percentis. Exemplos: 1) **Consulta de chaves no DICT:** tempo de resposta no percentil 99 deve ser menor que **1 segundo**. 2) **Atualizações no DICT:** tempo de resposta no percentil 99 deve ser inferior a **5 segundos**.

5.2. Comparação com o Modelo Markoviano

A análise mostra que os tempos médios de recuperação observados nos estados de falha variam entre 34 e 51 minutos, que estão acima dos limites estabelecidos no Manual de Tempos do Pix, que prevê indisponibilidade mensal total inferior a 43 minutos.

Essa observação indica possíveis violações dos SLAs e evidencia a necessidade de aprimoramento nos mecanismos de detecção, resposta e escalonamento automático.

Seria desejável haver maior transparência por parte do Banco Central sobre o que ocorre durante esses períodos de indisponibilidade, especialmente no que diz respeito às causas das falhas, à efetividade dos mecanismos de recuperação e às medidas adotadas para mitigar impactos futuros. Essa transparência contribuiria ainda mais para o fortalecimento da confiança no ecossistema Pix e para uma melhor colaboração entre os diversos agentes participantes na identificação e correção de vulnerabilidades sistêmicas.

6. Trabalhos Relacionados

A literatura sobre avaliação de confiabilidade, disponibilidade e recuperação de sistemas é extensa, especialmente no contexto de sistemas de missão crítica e serviços de alta disponibilidade [Trivedi 2001]. Em particular, diversos trabalhos abordam técnicas de modelagem estocástica, como Cadeias de Markov, para descrever a dinâmica de falhas e estratégias de rejuvenescimento de software. Em particular, técnicas de *survivability* são discutidas com exemplos do sistema bancário em [Somesh and Wing 2001].

[Jacques-Silva et al. 2015] propõem um modelo de *survivability* com rejuvenescimento automatizado para avaliar diferentes alternativas de implantação em sistemas distribuídos. Esse trabalho destaca a importância de mecanismos automáticos de recuperação para garantir a continuidade de serviço. Na mesma linha, [Avritzer et al. 2020] discutem o uso de técnicas de monitoramento de envelhecimento de software e estratégias de rejuvenescimento como ferramentas para melhorar a confiabilidade de sistemas de alta disponibilidade. Além disso, Koutras e Platis [Koutras and Platis 2020] analisam o desempenho de modelos de rejuvenescimento em sistemas sujeitos a múltiplos níveis de degradação, enfatizando a necessidade de distinguir entre diferentes severidades de falha para decisões mais eficazes de manutenção.

No presente trabalho, focamos especificamente na modelagem da confiabilidade do sistema Pix. Para tal, combinamos dados operacionais reais com modelagem matemática para analisar a resiliência de um sistema de pagamentos em tempo real.

7. Conclusão

Este trabalho apresentou um modelo markoviano de tempo contínuo para avaliar a confiabilidade do sistema Pix com base em dados reais de incidentes. A partir da classificação dos eventos em cinco estados operacionais, foi possível estimar uma matriz de transição e derivar métricas como tempos médios de recuperação e indisponibilidade. Os resultados indicam que os tempos médios de recuperação variam entre 34 e 51 minutos. Mostramos que, para manter a indisponibilidade mensal abaixo de 43 minutos, é necessário que a taxa de falhas (ε) seja inferior a 10^{-5} , o que equivale a uma falha a cada 70 dias, em média.

A distinção entre falhas severas e degradações leves permite análises mais realistas de SLA, destacando a importância de mecanismos automáticos de recuperação. Como trabalho futuro, propomos o refinamento da heurística de classificação de estados a partir de dados reais e a incorporação de novos dados para tornar o modelo mais preciso.

8. Artefatos e Reproducibilidade

A planilha (anonimizada) contendo todos os incidentes considerados neste trabalho encontra-se em: <https://tinyurl.com/wperf2025PIX>

Referências

- [Avritzer et al. 2020] Avritzer, A., Grottke, M., and Menasché, D. S. (2020). Using software aging monitoring and rejuvenation for the assessment of high-availability systems. In *Handbook of Software Aging and Rejuvenation: Fundamentals, Methods, Applications, and Future Directions*, pages 197–228. World Scientific.
- [Banco Central do Brasil 2022] Banco Central do Brasil (2022). Manual Operacional do DICT. Technical report, Banco Central do Brasil. <https://www.bcb.gov.br/content/estabilidadefinanceira/spi/ManualdeTemposdoPix-versao5-0.pdf>.
- [Banco Central do Brasil 2023] Banco Central do Brasil (2023). Manual de Tempos do Pix - Versão 5.0. Technical report, Banco Central do Brasil. https://www.bcb.gov.br/content/estabilidadefinanceira/pix/Regulamento_Pix/IX_ManualdeTemposdoPix.pdf.
- [Banco Central do Brasil 2024] Banco Central do Brasil (2024). Relatório de gestão do pix – ano 2023. https://www.bcb.gov.br/content/estabilidadefinanceira/pix/relatorio_de_gestao_pix/relatorio_gestao_pix_2023.pdf.
- [G1 2023] G1 (2023). Usuários reclamam de instabilidade no sistema do PIX nesta sexta-feira. <https://g1.globo.com/economia/noticia/2023/02/03/reclamacao-instabilidade-pix.ghtml>.
- [Jacques-Silva et al. 2015] Jacques-Silva, G., Avritzer, A., Menasche, D. S., Koziolk, A., Happe, L., and Suresh, S. (2015). Survivability modeling to assess deployment alternatives accounting for rejuvenation. In *ISSREW*, pages 114–119. IEEE.
- [Koutras and Platis 2020] Koutras, V. P. and Platis, A. N. (2020). On the performance of software rejuvenation models with multiple degradation levels. *Software Quality Journal*, 28.
- [Musa 1975] Musa, J. D. (1975). A theory of software reliability and its application. *IEEE Transactions on Software Engineering*, 1(03):312–327.
- [Somesh and Wing 2001] Somesh, J. and Wing, J. M. (2001). Survivability analysis of networked systems. In *Proceedings of the 23rd International Conference on Software Engineering. ICSE 2001*, pages 307–317. IEEE.
- [Trivedi 2001] Trivedi, K. S. (2001). *Probability and statistics with reliability, queuing, and computer science applications*. John Wiley & Sons.