Uma Extensão Pós-Quântica Híbrida para o Protocolo Matrix: Avaliação Experimental e Impacto Sistêmico

Marcos Ortiz[1], Vinícius Lagrota[2], Gilvan Maia[1], Rodrigo Pacheco[2] e Paulo Rego[1]

[1]Mestrado e Doutorado em Ciência da Computação (MDCC)
Universidade Federal do Ceará (UFC)
[2]Centro de Pesquisa e Desenvolvimento para a Segurança das Comunicações (CEPESC)

mdo@ufc.br, paulo@dc.ufc.br, gilvanmaia@virtual.ufc.br
vinicius.1232@abin.gov.br,rodrigo.6874@abin.gov.br

Abstract. *Quantum computing threatens E2EE messenger encryption, driving the adoption of post-quantum cryptography (PQC). This study evaluates the hybrid integration of CRYSTALS-Kyber into the Matrix protocol, extending the key agreement and the Double Ratchet. Experimental tests demonstrate that the primary bottleneck is bandwidth, with a +548% overhead in setup and +252% in key rotations, while the impact on CPU and user experience is negligible (≈16 ms to 100 ms). The cost of rotations scales quadratically, indicating that rigid security policies are inefficient and demand adaptive strategies to balance security and network consumption.*

Resumo. *A computação quântica ameaça a criptografia de mensageiros E2EE, impulsionando a adoção de criptografia pós-quântica (PQC). Este estudo avalia a integração híbrida do CRYSTALS-Kyber no protocolo Matrix, estendendo o acordo de chaves e o Double Ratchet. Testes demonstraram que o principal gargalo é a largura de banda, com overhead de +548% no setup e +252% nas rotações de chaves, enquanto o impacto na CPU e na experiência do usuário é negligenciável (≈16 ms a 100 ms). O custo das rotações escala quadraticamente, indicando que políticas de segurança rígidas são ineficientes e demandam estratégias adaptativas para equilibrar segurança e consumo de rede.*

1. Introdução

Plataformas de comunicação baseadas em Criptografia de Ponta a Ponta, *End-to-End Encryption* (E2EE) exercem papel central em sistemas distribuídos modernos, nos quais clientes heterogêneos interagem por meio de infraestruturas intermediárias potencialmente não confiáveis. Nesse contexto, protocolos como o Matrix buscam assegurar confidencialidade e autenticidade nos clientes finais, enquanto servidores se limitam ao roteamento e ao armazenamento de mensagens cifradas. A relevância desse paradigma é evidenciada por implantações em produção, como o Tchap[1], sistema de mensagens seguras do governo francês, e o msg gov[2], plataforma de comunicação segura destinada a servidores públicos e autoridades governamentais, ambos baseados no Matrix.

A segurança desses mecanismos, porém, ainda depende majoritariamente de primitivas criptográficas clássicas, como acordos de chaves baseados em *Elliptic Curve Diffie–Hellman* (ECDH), vulneráveis ao avanço da computação quântica [Shor 1994]. Nesse

[1]https://www.tchap.gouv.fr/
[2]https://agenciabrasil.ebc.com.br/tags/msg-gov

cenário, o modelo de ameaça *Harvest Now, Decrypt Later* (HNDL) reforça a urgência de defesas pós-quânticas, pois dados cifrados hoje podem ser armazenados para futura decifragem. Em contraste, primitivas simétricas e funções de *hash* não são diretamente afetadas pelo algoritmo de Shor, e o principal ganho quântico conhecido em ataques genéricos é quadrático, via algoritmo de Grover [Aydeger et al. 2024]. Assim, com tamanhos de chave adequados, espera-se que a criptografia simétrica permaneça segura no horizonte pós-quântico, razão pela qual o *National Institute of Standards and Technology* (NIST) ainda admite o uso de *Advanced Encryption Standard* (AES), embora desaconselhe mecanismos com menos de 112 bits de segurança clássica [Barker 2020].

Como resposta, vêm sendo propostas extensões híbridas que combinam primitivas clássicas e pós-quânticas, permitindo transição gradual e compatível com sistemas legados. Um exemplo representativo é o protocolo *Post-Quantum Extended Diffie-Hellman* (PQXDH), extensão pós-quântica do *Extended Triple Diffie-Hellman* (X3DH) que integra um mecanismo *Key Encapsulation Mechanism* (KEM) – como o CRYSTALS-Kyber-1024 – ao acordo ECDH, oferecendo proteção prospectiva contra ataques do tipo HNDL [Kret and Schmidt 2023]. Embora essas propostas reforcem as garantias de segurança, também introduzem custos adicionais de computação, comunicação e gerenciamento de estado, ainda pouco caracterizados em sistemas distribuídos reais.

Neste trabalho, a incorporação de criptografia pós-quântica no protocolo Matrix foi investigada por meio de extensões híbridas implementadas via *wrapper*[3] sobre o protocolo Olm, responsável pelos canais E2EE entre dispositivos [Matrix.org 2019]. Essas extensões abrangem o estabelecimento de chaves e o mecanismo de Double Ratchet da biblioteca `vodozemac`[4], preservando propriedades fundamentais como *Forward Secrecy* (FS) e *Post-Compromise Security* (PCS) [Cohn-Gordon et al. 2016]. Em contraste, o Megolm não foi alterado, pois a cifração de mensagens em salas é realizada por criptografia simétrica (AES-256-CBC), cujo impacto pós-quântico pode ser tratado por dimensionamento apropriado de chave [Matrix.org 2022]. Foi conduzida uma análise experimental sistemática com foco no impacto da adoção de *Post-Quantum Cryptography* (PQC), considerando largura de banda, redistribuição de chaves, frequência de rotação de sessões e políticas de segurança em padrões realistas de uso do Matrix.

As principais contribuições deste trabalho são: implementação de um *wrapper* PQC integrado à biblioteca `vodozemac`, viabilizando a execução das variantes *Classical* e *Hybrid* do protocolo Matrix; extensão do protocolo de estabelecimento de chaves e do mecanismo de Double Ratchet para suportar primitivas PQC em um ambiente real; e análise experimental controlada dos impactos sistêmicos da adoção de PQC, considerando diferentes *workloads*, políticas de rotação de chaves e cenários de execução.

O restante deste artigo está organizado da seguinte forma: a Seção 2 apresenta a arquitetura do protocolo Matrix e os fundamentos de segurança pós-quântica. A Seção 3 discute os trabalhos relacionados. A Seção 4 descreve o wrapper pós-quântico proposto. A Seção 5 detalha a metodologia experimental e os artefatos. A Seção 6 apresenta e discute os resultados. Por fim, a Seção 7 encerra o artigo com as conclusões.

[3] `https://github.com/mdo-br/matrix_pqc`

[4] `https://matrix-org.github.io/vodozemac/vodozemac/index.html`

2. Arquitetura do Protocolo Matrix e Segurança Pós-Quântica

O Matrix é um protocolo aberto para comunicação em tempo real, com uma arquitetura descentralizada e federada que permite a publicação, persistência e assinatura de dados de forma segura. Sua especificação define um ecossistema onde a comunicação entre domínios ocorre sem um ponto único de controle, suportando desde mensagens instantâneas até sinalização de *Voice over IP* (VoIP) e dispositivos inteligentes [Matrix.org 2024]. No modelo de federação do Matrix, cada servidor doméstico, *homeserver* (HS), é responsável pela gestão das contas dos seus usuários, sendo a identidade de cada participante globalmente única e composta pela combinação de um identificador local e do domínio do servidor (e.g., `@bob:ufc.br`), como ilustrado na Figura 1. Essa estrutura permite a interação entre usuários de servidores distintos por meio da cooperação entre os HS, onde, embora o transporte entre clientes e servidores seja protegido por *Transport Layer Security* (TLS), as garantias fundamentais de confidencialidade e autenticidade são providas por uma camada robusta de E2EE.

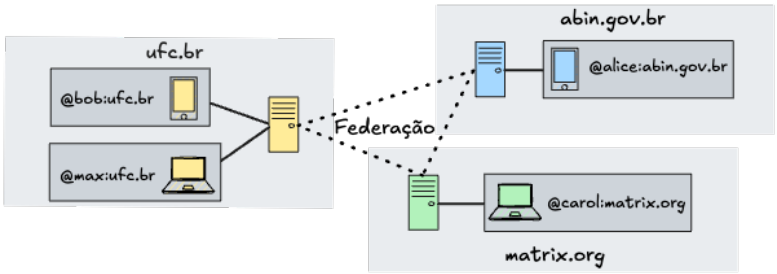

Figura 1. Visão conceitual da federação no Matrix

A implementação da E2EE no Matrix assegura que apenas os participantes finais de uma conversa acessem o conteúdo das mensagens, mitigando riscos associados a servidores intermediários não confiáveis. Esse modelo se apoia nos protocolos criptográficos Olm e Megolm, cujas implementações de referência em Rust estão na biblioteca `vodozemac`. O protocolo Olm é utilizado na comunicação ponto a ponto e no estabelecimento inicial de sessões seguras, sendo baseado no mecanismo de Double Ratchet inspirado no Signal Protocol [Matrix.org 2019] e empregando o protocolo *Triple Diffie-Hellman* (3DH) no acordo inicial de chaves. Já o protocolo Megolm é adotado nas comunicações em salas, cifrando as mensagens e utilizando o Olm como canal seguro para distribuir as chaves de sessão entre os dispositivos participantes. Nessa arquitetura, todas as mensagens, inclusive em conversas individuais (1:1), são organizadas em salas e cifradas via Megolm, facilitando a expansão de diálogos diretos para grupos sem renegociação criptográfica completa.

A eficácia desses protocolos baseia-se em mecanismos de *ratcheting* que viabilizam dois atributos críticos de segurança: a FS, que garante que o comprometimento de chaves de longo prazo não afete a confidencialidade de mensagens passadas; e a PCS, que permite a recuperação automática da confidencialidade após um comprometimento temporário do estado interno de um participante. No contexto do Matrix, tais propriedades são mantidas pelo avanço das catracas do Olm e pela rotação constante das sessões Megolm, conforme detalhado em [Martins et al. 2025, Martins et al. 2026]. Entretanto, o advento da computação quântica impõe uma ameaça direta às primitivas criptográficas clássicas

utilizadas nestes protocolos, especialmente àquelas destinadas ao estabelecimento de chaves. Essa vulnerabilidade motiva o desenvolvimento da PQC e impulsiona o processo de padronização de algoritmos resistentes a adversários quânticos conduzido pelo NIST. Entre as soluções emergentes, destacam-se os mecanismos de encapsulamento de chaves (KEM), como o CRYSTALS-Kyber [NIST 2022]. Embora promissores para a segurança a longo prazo, a integração desses algoritmos em sistemas distribuídos como o Matrix introduz desafios práticos, visto que os custos adicionais de comunicação e processamento podem impactar métricas sensíveis, como largura de banda e latência operacional.

3. Trabalhos Relacionados

A literatura recente tem avançado na análise da adoção de PQC em protocolos de mensagens seguras, com ênfase no equilíbrio entre FS, PCS e eficiência de comunicação sob restrições de largura de banda. A tese de [Duits 2019] figura entre os primeiros estudos experimentais sobre a integração de troca de chaves pós-quântica em mensageria segura. Foram avaliadas variantes do protocolo Signal com algoritmos como SIDH e CRYSTALS-Kyber, reportando latência, consumo de energia e largura de banda em dispositivos móveis. Os resultados evidenciaram a viabilidade de abordagens híbridas em ambientes restritos, servindo de referência metodológica para avaliações posteriores.

Em produção, o protocolo PQ3 da Apple para o iMessage consolidou a adoção de criptografia híbrida. O estabelecimento e a renovação de chaves combinam primitivas clássicas e pós-quânticas, com política adaptativa para reduzir impacto de banda, acionando o componente pós-quântico com menor frequência. A segurança do PQ3 foi analisada formalmente em [Stebila 2024], considerando diferentes modelos de adversário e propriedades como FS pós-comprometimento.

Com foco explícito em eficiência de comunicação, [Dodis et al. 2025] propuseram o protocolo *Triple Ratchet*, introduzindo o KEM Katana (otimizado a partir do CRYSTALS-Kyber) e o uso de códigos de apagamento para fragmentação incremental do material criptográfico. Essas técnicas têm sido incorporadas ao *ML-KEM Braid* do Signal [Schmidt and Connell 2025], reforçando sua aplicabilidade prática. Complementarmente, [Auerbach et al. 2025] propõem uma metodologia formal para comparação de protocolos sob restrições de banda, baseada no modelo *Sparse Continuous Key Agreement* (SCKA) e na métrica *Vulnerable Message Set* (VMS), que quantifica exposição temporal após comprometimentos.

Em conjunto, esses trabalhos estabelecem a base conceitual e metodológica que orienta esta pesquisa: viabilidade experimental de híbridos [Duits 2019], adoção em produção com renovação adaptativa [Stebila 2024], técnicas de redução de banda [Dodis et al. 2025, Schmidt and Connell 2025] e métricas comparativas [Auerbach et al. 2025]. A Tabela 1 consolida essa comparação e evidencia que a avaliação empírica no ecossistema Matrix, com implementação reprodutível e métricas de custo sistêmico, permanece necessária para subsidiar políticas de rotação de chaves mais eficientes e contextualizadas.

4. Wrapper Pós-Quântico Proposto

Nesta seção, é descrito o *wrapper* pós-quântico proposto, no qual primitivas PQC foram incorporadas ao plano de controle criptográfico do Matrix. A integração foi concen-

Tabela 1. Comparação deste estudo com os principais trabalhos relacionados

Trabalho	Contexto	PQC	Métricas	Relação
Duits (2019)	Signal, X3DH	SIDH, Kyber	Latência, energia, banda	Viabilidade de acordos híbridos; base para análise Olm/Megolm.
Apple PQ3 (2024)	iMessage	Kyber768, ML-KEM1024	Renovação adaptativa	Implementação prática de renovação adaptativa em produção.
Dodis et al. (2025)	Triple Ratchet	Katana, codes	Redução de banda	Mitigação de banda via fragmentação; integrado ao ML-KEM Braid (Signal).
Auerbach et al. (2025)	SCKA, CKA	KEMs genéricos	Modelo SCKA, VMS	Métrica VMS como referência futura; SCKA integrado ao ML-KEM Braid (Signal).
Este trabalho (2026)	Vodozemac (Matrix)	Kyber-1024, Kyber-768	Banda, tempo, escalabilidade	Avaliação empírica de políticas de rotação (R=25–250) em grupos heterogêneos (N=2–150); quantificação de overhead sistêmico e identificação de gargalos de banda.

trada em (i) estabelecimento inicial de canais Olm via acordo híbrido (PQXDH) e (ii) distribuição de sessões Megolm durante rotações, por meio de um Double Ratchet estendido com material pós-quântico. Em contraste, o Megolm não foi modificado, uma vez que a cifração do tráfego de dados na sala é realizada por criptografia simétrica (AES-256-CBC), conforme a especificação [Matrix.org 2024].

4.1. Visão Geral

O *wrapper* foi implementado como uma extensão direta da biblioteca `vodozemac`, permitindo que a adoção de PQC fosse realizada na camada criptográfica, sem demandar alterações na API cliente–servidor definida pela especificação do Matrix [Matrix.org 2024]. Com isso, a compatibilidade com clientes existentes é preservada, ao mesmo tempo em que o custo das primitivas pós-quânticas é mantido restrito a operações pontuais de controle (acordo e distribuição de chaves), evitando *overhead* pós-quântico por mensagem no tráfego de sala.

A Figura 2 sumariza a arquitetura proposta e evidencia três fluxos: ❶ o estabelecimento inicial de canais Olm com PQXDH híbrido entre dispositivos; ❷ a distribuição de sessões Megolm via canais Olm, com avanço controlado do Double Ratchet estendido; e ❸ o envio das mensagens cifradas na sala com Megolm (AES-256-CBC), sem *overhead* pós-quântico por mensagem na sessão. As rotações são acionadas por uma política definida no escopo da sala, que determina quando uma nova sessão Megolm deve ser criada e redistribuída.

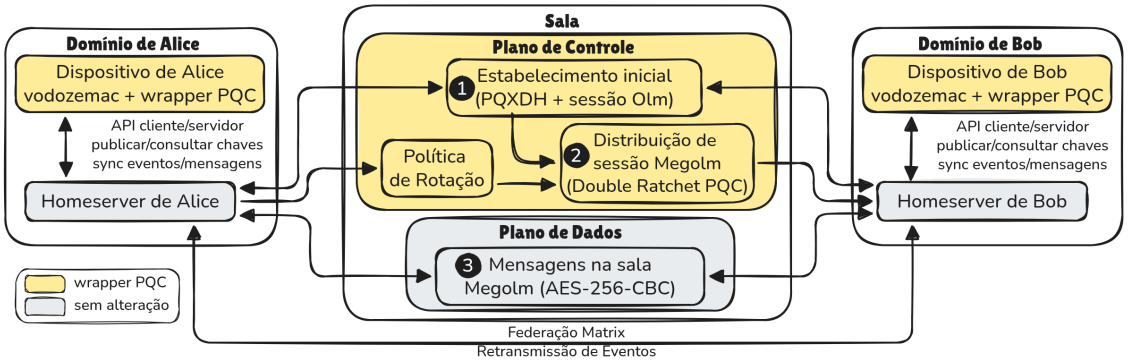

Figura 2. Arquitetura do *wrapper* pós-quântico proposto.

4.2. Acordos Clássico e Híbrido

No Olm clássico, a sessão 1:1 é estabelecida por um acordo de chaves do tipo 3DH, seguido pela inicialização do Double Ratchet [Matrix.org 2019]. Nesse modo, o componente assimétrico é instanciado exclusivamente por ECDH sobre X25519 (Curve25519), derivando o segredo inicial que alimenta a chave-raiz do *ratchet*. A partir desse ponto, a evolução do estado criptográfico emprega primitivas simétricas e funções de derivação, usando HKDF-SHA-256 na derivação inicial e na atualização do *root-key*, e HMAC-SHA-256 para avanço das *chain keys* como definido na especificação do Olm [Matrix.org 2019]. Esse comportamento do modo *Classical* é descrito na Tabela 2.

No *wrapper* proposto, o acordo inicial foi estendido para PQXDH [Kret and Schmidt 2023], compondo um acordo híbrido em que o segredo inicial é derivado pela combinação de (i) segredos clássicos via X25519 e (ii) um mecanismo KEM pós-quântico. Em particular, o componente KEM adotado no acordo foi o CRYSTALS-Kyber-1024, conforme definido para o modo *Hybrid* na Tabela 2. O segredo híbrido resultante foi utilizado como entrada para as mesmas funções de derivação do Olm, de modo que a integração de PQC seja realizada no estabelecimento de sessão (fluxo ❶ na Figura 2), sem alterar o comportamento do tráfego de dados na sala.

4.3. Avanço de Chaves via Double Ratchet

No Double Ratchet clássico, a evolução de chaves decorre da composição de uma catraca assimétrica (baseada em ECDH com X25519) e de uma catraca simétrica (derivações de chaves de cadeia e de mensagem), provendo FS e PCS [Cohn-Gordon et al. 2016]. Na prática, a catraca simétrica permanece no caminho crítico por mensagem, enquanto a catraca assimétrica atualiza a chave-raiz em eventos discretos.

No *wrapper*, o Double Ratchet foi estendido para incorporar material pós-quântico de forma controlada, preservando sua estrutura, mas incluindo uma contribuição de KEM na atualização da chave-raiz. Em conformidade com a Tabela 2, o modo *Hybrid* injeta CRYSTALS-Kyber-768 no mecanismo de Double Ratchet, enquanto o componente clássico permanece baseado em X25519. Dessa forma, a atualização da chave-raiz passa a refletir uma composição híbrida (X25519 + Kyber-768), concentrando o custo pós-quântico em eventos específicos e restritos ao plano de controle. Na arquitetura proposta, esses eventos ocorrem durante a distribuição e a rotação das chaves de sessão Megolm (fluxo ❷ na Figura 2), reduzindo a necessidade de operações pós-quânticas frequentes. Essa decisão de projeto está alinhada ao racional de amortização observado em protocolos híbridos recentes, como o PQ3, nos quais o componente pós-quântico é acionado apenas em momentos específicos, limitando o impacto sobre largura de banda e sobrecarga computacional.

4.4. Políticas de Rotação de Sessão Megolm

Como complemento ao *wrapper* criptográfico, políticas simples de rotação de sessões Megolm foram incorporadas para controlar a janela temporal de exposição criptográfica em salas. Essas políticas determinam quando uma nova sessão Megolm deve ser criada e redistribuída aos participantes por meio de canais Olm (clássicos ou híbridos).

As políticas implementadas são baseadas em critérios fixos (e.g., número de mensagens ou intervalos temporais) e não consideram, neste momento, sinais contextuais do

usuário, do dispositivo ou do ambiente. Ainda assim, elas permitem explorar de forma controlada o *trade-off* entre segurança temporal e custo operacional. Nesse contexto, rotações mais frequentes reduzem a quantidade de mensagens potencialmente expostas sob comprometimento de chave de sessão, ao custo de maior tráfego de controle e maior frequência de redistribuição (fluxo ❷).

4.5. Considerações

A implementação foi estruturada para permitir comparação direta entre as variantes *Classical* e *Hybrid*, mantendo a mesma base de código e diferenciando apenas os pontos de integração das primitivas pós-quânticas (acordo e atualização híbrida do *ratchet*). Para viabilizar a avaliação em condições controladas, foi desenvolvido um módulo para simular salas e padrões de comunicação, uma vez que o *wrapper* foi restrito à extensão da `vodozemac`, sem incorporar a camada de API cliente–servidor[5]. Os detalhes do ambiente, fatores e métricas coletadas são apresentados na Seção 5.

5. Metodologia Experimental e Artefatos

A metodologia experimental foi estruturada para quantificar o impacto sistêmico da adoção de criptografia pós-quântica no protocolo Matrix, comparando diretamente as variantes *Classical* e *Hybrid*. A avaliação foi conduzida com foco em três dimensões: (i) tempo de execução em fases críticas do protocolo, (ii) *overhead* de largura de banda associado ao plano de controle criptográfico, e (iii) efeito de políticas de rotação de chaves sobre o *trade-off* entre segurança e eficiência. Para garantir coerência com a contribuição descrita na Seção 4, foi utilizada a implementação do *wrapper* PQC integrada à biblioteca `vodozemac`.

5.1. *Design* Experimental

Adotou-se um *design* pareado, executando cada configuração em duas condições: uma em modo *Classical* e uma em modo *Hybrid*. Essa estratégia foi empregada para reduzir variabilidade causada por fatores externos (como efeitos de *cache* e flutuações de CPU e do sistema operacional), garantindo maior precisão na análise comparativa. Foram realizadas 30 repetições pareadas, totalizando 60 execuções (30 *Classical* + 30 *Hybrid*), com a ordem de execução dos modos alternada sistematicamente para mitigar efeitos de ordem.

A Tabela 2 detalha os fatores experimentais, seus respectivos níveis, o ambiente de execução e a carga de trabalho. Para cada execução, foram coletadas métricas que permitem decompor o custo do modo *Hybrid* em relação ao modo *Classical*. Do ponto de vista temporal, foram medidos: (i) o tempo de *setup* (estabelecimento inicial de sessões Olm e primeira Distribuição de chaves Megolm), (ii) o tempo acumulado associado às rotações (redistribuições subsequentes de chaves Megolm ao longo da execução), e (iii) os tempos de cifração e decifração das mensagens de sala (excluindo os períodos de rotação), de forma a separar custos de dados e de controle.

Quanto à largura de banda, foram contabilizados separadamente os bytes transmitidos no *setup* e nas rotações (plano de controle), bem como o tamanho das mensagens em sala cifradas (plano de dados). Por fim, métricas de rotação foram registradas para validar o comportamento das políticas e correlacionar frequência de rotação com *overhead*

[5] `https://github.com/matrix-org/matrix-rust-sdk`

observado, incluindo o número total de rotações efetivadas e o volume de redistribuição associado a cada rotação.

Tabela 2. Fatores e níveis da avaliação de desempenho e largura de banda.

Fator	Níveis	Descrição
Modo Criptográfico	`Classical` `Hybrid`	Variante clássica usa apenas X25519 (ECDH). Variante híbrida adiciona CRYSTALS-Kyber-1024 ao protocolo de estabelecimento de chaves (PQXDH) e CRYSTALS-Kyber-768 ao mecanismo de Double Ratchet, garantindo resistência prospectiva a adversários quânticos.
Tipo de Sala	`DM` `SmallGroup` `MediumGroup` `LargeChannel`	Conversas diretas (2 usuários), grupos pequenos (7 usuários), grupos médios (25 usuários) e canais grandes (150 usuários). O experimento cria 11 salas simultâneas: 5 DM, 3 SmallGroup, 2 MediumGroup e 1 LargeChannel.
Política de Rotação[1]	`Paranoid` `PQ3` `Balanced` `Relaxed`	Paranoid: rotação a cada 25 mensagens (máxima segurança). PQ3: 50 mensagens (Apple iMessage). Balanced: 100 mensagens (padrão Matrix). Relaxed: 250 mensagens (eficiência prioritária).
Carga de Trabalho[2]	500–1250 mensagens	Número de mensagens enviadas varia por tipo de sala: DM (500), SmallGroup (750), MediumGroup (1000), LargeChannel (1250). Mensagens geradas com tipos e tamanhos variados (texto: 50-500B; imagem: 10-500kB; arquivo: 100kB-5MB; voz: 10-200kB) [Seufert et al. 2023]. Garante múltiplas rotações em todas as políticas.
Número de Senders	1 sender	Single-user profile: apenas 1 membro envia mensagens, simulando experiência típica do usuário. Todos os membros recebem mensagens (N-1 receptores).
Repetições	30 pares	30 repetições pareadas (Classical $\leftrightarrow$ Hybrid), totalizando 60 execuções. Ordem alternada entre repetições para controlar efeitos de cache/aquecimento.
Ambiente	`Notebook`	Intel Core i7-1165G7 (11th Gen), 16GB RAM, Ubuntu 24.04 LTS.

[1] Políticas de rotação controlam frequência de redistribuição de chaves Megolm via canais Olm. Rotações mais frequentes aumentam PCS, mas geram maior *overhead* de controle.
[2] Escalonada: garante pelo menos 2 rotações em política Relaxed até 50 rotações em política Paranoid para LargeChannel.

5.2. Análise Estatística

A análise estatística seguiu metodologia inspirada nas práticas de [Duits 2019] e [Auerbach et al. 2025]. Para cada configuração, analisou-se a diferença pareada entre os valores observados nos modos *Hybrid* e *Classical*. Aplicou-se o teste de Shapiro-Wilk às diferenças para orientar a escolha entre teste t pareado e Wilcoxon *signed-rank*. Para múltiplas comparações, aplicou-se a correção de Holm-Bonferroni aos *p-values*. Intervalos de confiança foram calculados a 95% (*t-Student* para dados normais ou *bootstrap* com 10.000 reamostragens para dados não-normais). As estatísticas descritivas reportadas incluem mediana, *Interquartile Range* (IQR) e P95; e o *overhead* percentual é definido como a razão entre as medianas *Hybrid* e *Classical*.

6. Resultados

Os resultados da avaliação experimental são apresentados em três dimensões: (i) *overhead* de largura de banda, quantificando o custo adicional de comunicação imposto por primitivas PQC no plano de controle; (ii) *overhead* de tempo de processamento, avaliando o impacto computacional em CPU; e (iii) *trade-off* segurança temporal vs eficiência, analisando a relação entre frequência de rotação de chaves, janela de exposição criptográfica e custo operacional acumulado.

6.1. Overhead de Largura de Banda e Tempo de Processamento

A Figura 3 apresenta uma visão consolidada do *overhead* PQC em ambas as dimensões (largura de banda e tempo de processamento), decomposto por fase do protocolo. O

overhead percentual médio (diferença *Hybrid* − *Classical*) agregado sobre todas as configurações experimentais (4 tipos de sala × 4 políticas de rotação), com todos os resultados estatisticamente significativos (p-*value* $< 0,001$).

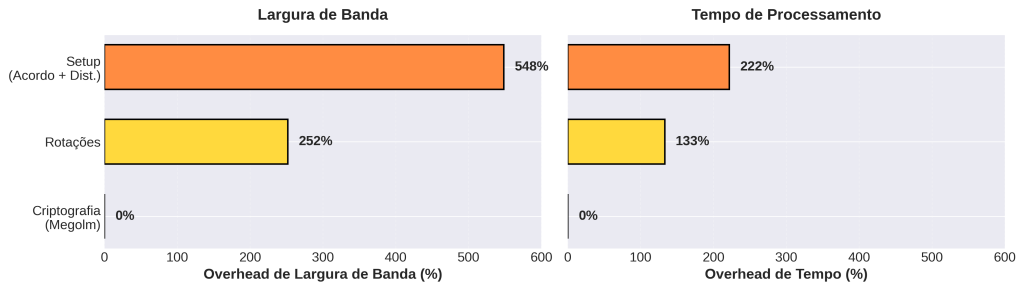

Figura 3. *Overhead* PQC: largura de banda vs tempo no plano de controle.

A análise comparativa revela três achados principais. Primeiramente, observa-se que o plano de dados permanece virtualmente inalterado, uma vez que a criptografia Megolm, baseada em AES-256-CBC simétrico, não apresenta *overhead* em largura de banda ou tempo de processamento. Esse resultado confirma que as mensagens do usuário final não são penalizadas pela adoção de PQC, concentrando-se o impacto tecnológico exclusivamente no plano de controle, especificamente no estabelecimento de canais Olm e nas rotações de chaves. Paralelamente, nota-se uma assimetria acentuada entre o *overhead* de banda e o tempo de execução durante a fase inicial (acordo + distribuição inicial). Enquanto o aumento no tempo de processamento é de 222%, o impacto na largura de banda é substancialmente superior, atingindo 548% (um fator de 2,5× em relação ao tempo). Essa disparidade decorre da natureza das primitivas CRYSTALS-Kyber, que, embora computacionalmente eficientes, geram *payloads* significativos, variando entre 1,184 kB (CRYSTALS-Kyber-768) e 1,568 kB (CRYSTALS-Kyber-1024). Tal cenário sugere que a largura de banda, e não a capacidade de processamento da CPU, constitui o gargalo principal para a viabilização de PQC em protocolos de mensageria federada.

Por fim, as operações de rotação de chaves apresentam um *overhead* moderado quando comparadas à fase inicial (+252% em banda e +133% em tempo), contudo, seu impacto sistêmico é amplificado pela recorrência do processo. Dado que cada rotação exige a redistribuição da chave Megolm para todos os receptores via canais Olm individuais, o custo operacional torna-se cumulativo, reforçando a necessidade de estratégias de otimização no gerenciamento de chaves em ambientes de larga escala.

6.1.1. Valores Absolutos e Experiência do Usuário

A Tabela 3 detalha o *overhead* absoluto por fase, tipo de sala e política de rotação, permitindo contextualizar o impacto na experiência do usuário. Os dados são decompostos em duas fases (*setup*: acordo + distribuição inicial; e rotações) para cada dimensão (largura de banda e tempo), quatro tipos de sala (DM, Small, Medium, Large) e quatro políticas de rotação (Paranoid, PQ3, Balanced, Relaxed). Os valores correspondem à mediana do *overhead* (diferença *Hybrid* − *Classical*) observado em 30 repetições pareadas, com todos os resultados estatisticamente significativos (p-*value* $< 0,001$ via teste de Wilcoxon).

Tabela 3. Overhead de Largura de Banda e Tempo por Fase, Sala e Política

Sala	Política	Overhead de Largura de Banda		Overhead de Tempo	
		Setup	Rotação	Setup	Rotação
DM (2)	Paranoid (25)	4,1 kB	32,7 kB	0,86 ms	2,51 ms
DM (2)	PQ3 (50)	4,1 kB	16,3 kB	0,84 ms	1,26 ms
DM (2)	Balanced (100)	4,1 kB	8,2 kB	0,83 ms	0,68 ms
DM (2)	Relaxed (250)	4,1 kB	3,3 kB	0,84 ms	0,28 ms
Small (7)	Paranoid (25)	24,9 kB	293,9 kB	5,14 ms	19,77 ms
Small (7)	PQ3 (50)	24,9 kB	147,0 kB	4,97 ms	9,51 ms
Small (7)	Balanced (100)	24,9 kB	68,6 kB	4,94 ms	4,54 ms
Small (7)	Relaxed (250)	24,9 kB	29,4 kB	4,96 ms	1,95 ms
Medium (25)	Paranoid (25)	99,5 kB	1,53 MB	20,62 ms	106,46 ms
Medium (25)	PQ3 (50)	99,5 kB	783,8 kB	20,04 ms	50,88 ms
Medium (25)	Balanced (100)	99,5 kB	391,9 kB	19,89 ms	25,14 ms
Medium (25)	Relaxed (250)	99,5 kB	156,8 kB	19,95 ms	10,20 ms
Large (150)	Paranoid (25)	617,8 kB	11,88 MB	128,09 ms	834,22 ms
Large (150)	PQ3 (50)	617,8 kB	5,94 MB	124,50 ms	403,87 ms
Large (150)	Balanced (100)	617,8 kB	2,85 MB	122,51 ms	188,93 ms
Large (150)	Relaxed (250)	617,8 kB	1,19 MB	122,60 ms	78,89 ms

Apesar do *setup* ter *overhead* percentual elevado (≈548%), as rotações consomem mais banda absoluta devido ao volume acumulado de redistribuições. Em SmallGroup ($N = 7$) com política PQ3, o *setup* consome 24,9 kB, mas as rotações acumulam 147,0 kB (6× maior que o *setup*). Para Large (150) com Paranoid, as rotações atingem 11,88 MB, enquanto o *setup* consome apenas 617,8 kB. Esse padrão confirma que **rotações dominam o consumo de largura de banda** em sessões com múltiplas mensagens. A escalabilidade revela que o *setup* escala quadraticamente com $N \times (N - 1)$ sessões Olm estabelecidas. Large (150) apresenta *overhead* absoluto 149× maior que DM (2), mas *overhead* percentual consistente (≈548% em ambos), indicando que a complexidade do protocolo Matrix escala proporcionalmente ao número de sessões.

Do ponto de vista da experiência do usuário, o *overhead* de controle é significativo em termos absolutos: em Large/Paranoid, cada usuário envia 12,5 MB adicionais de controle para transmitir 1.250 mensagens de dados (aproximadamente 10 kB de *overhead* de controle por mensagem de usuário). A política de rotação tem impacto direto: comparando Paranoid (25 mensagens) e Relaxed (250 mensagens) para SmallGroup, o *overhead* de rotações varia de 293,9 kB a 29,4 kB, uma redução de 10× que reflete diretamente a janela de segurança temporal 10× maior do Relaxed.

O *overhead* de tempo é moderado e imperceptível na maioria dos cenários. Em Large (150) com PQ3, a fase inicial consome 124,5 ms de *overhead* (tempo adicional para estabelecer 22.350 sessões Olm via PQXDH), enquanto Rotações acumulam 403,9 ms ao longo de 25 redistribuições. A cada 50 mensagens, ocorre uma pausa de ≈16 ms para redistribuir a chave Megolm. Esse atraso é **imperceptível em interações humanas** (latência típica de rede Wi-Fi: 10-50 ms), mas pode ser relevante em aplicações de alta frequência (e.g., bots automatizados enviando centenas de mensagens por segundo).

A assimetria entre *overhead* de banda e tempo é evidente: *setup* tem razão BW/Time ≈ 2,5 (largura de banda cresce mais que tempo), indicando que CRYSTALS-Kyber gera *payloads* grandes, apesar de rápido. Rotações têm razão ≈ 1,9 (mais equilibrado). O *overhead* de tempo de rotações (2,3× ou +133%) é significativamente menor

que o de largura de banda (3,5× ou +252%), confirmando que operações PQC incrementais (CRYSTALS-Kyber-768 Encaps) são rápidas apesar dos *payloads* grandes. Em DM (2) com Relaxed, o *overhead* total de tempo é de apenas 1,12 ms (0,84 ms setup + 0.28 ms rotações), i.e., **PQC não comprometeria a experiência de usuário em cenários típicos**.

6.1.2. Análise de Componentes: Além das Primitivas Isoladas

A avaliação sistêmica neste trabalho vai além da caracterização de primitivas criptográficas isoladas. A Tabela 4 decompõe as partes de largura de banda para contextualizar a origem do *overhead*: primitivas PQC (chaves públicas CRYSTALS-Kyber, *ciphertexts* KEM) contribuem com 71% do custo total, enquanto o *overhead* de protocolo (serialização JSON, codificação Base64, headers Matrix) adiciona 29%.

Tabela 4. Decomposição de primitivas criptográficas e *overhead* de protocolo por fase (SmallGroup, N=7, single-user profile).

Fase	Primitivas		*Overhead*	
	Classical	**Hybrid**	**Classical**	**Hybrid**
Acordo				
Identity keys	64 B	64 B	–	–
OTK	32 B	32 B	–	–
CRYSTALS-Kyber-1024 PK	–	1,568 B	–	–
Protocolo	–	–	14 B	336 B
Total (7 bundles*)*	*672 B*	*1.664 B*	*98 B*	*3.145 B*
TOTAL	**662 B**		**4.809 B**	
***Setup*/rotação**				
Megolm *key*	308 B	308 B	–	–
X25519 *ratchet*	32 B	32 B	–	–
CRYSTALS-Kyber-768 *ratchet*	–	1.184 B	–	–
MAC + *headers*	32 B	32 B	101 B	151 B
Base64 *encoding*	–	–	191 B	629 B
Total (6 msgs)	*2.232 B*	*9.336 B*	*1.752 B*	*4.680 B*
TOTAL	**3.984 B**		**14.016 B**	
Overhead PQC				
Acordo	–	+3.136 B	–	+3.047 B
Setup/rotação	–	+7.104 B	–	+2.928 B

Interpretação: Primitivas = componentes criptográficos puros. *Overhead* = serialização JSON, Base64, *headers*, MACs. Acordo usa CRYSTALS-Kyber-1024 (NIST L5), *setup*/rotação usam CRYSTALS-Kyber-768 (NIST L3). *Single-user profile*: 1 *sender* × $(N-1)$ *receptores*.

Essa decomposição evidencia que o *overhead* é predominantemente inerente aos algoritmos pós-quânticos (CRYSTALS-Kyber-1024 no *setup* com 1,568 kB de chave pública e CRYSTALS-Kyber-768 nas rotações com 1,184 kB de chave pública) e não decorre de ineficiências na implementação do protocolo Matrix. Contudo, a análise sistêmica completa (Tabela 3) integra esses componentes ao comportamento operacional real: número de sessões Olm estabelecidas — $N \times (N-1)$ —, frequência de rotações (controlada por política), e escalabilidade com tamanho de sala. Essa abordagem permite responder questões práticas como, por exemplo, o custo de largura de banda para manter uma sala de 150 participantes com rotações a cada 50 mensagens, que seria de 617.8 kB *setup* + 5.94 MB rotações, totalizando 6.56 MB de controle para 1.250 mensagens.

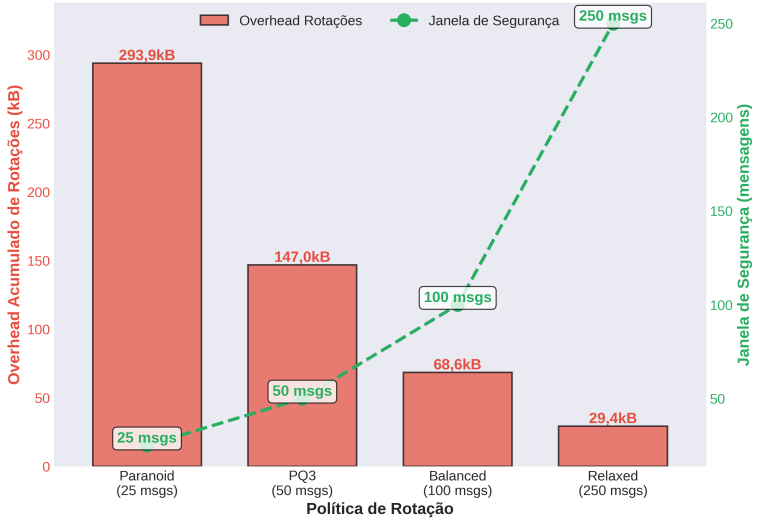

Figura 4. *Trade-off* entre segurança e custo de rotações para $N = 7$.

6.2. *Trade-off*: Segurança Temporal vs Custo Operacional

Políticas de rotação controlam explicitamente o *trade-off* entre segurança temporal (frequência de atualização de chaves) e custo operacional (largura de banda e tempo de CPU). A Figura 4 sumariza esse *trade-off* para SmallGroup ($N = 7$), quantificando: (i) largura de banda acumulada de rotações (eixo y esquerdo, barras vermelhas) e (ii) janela de segurança (eixo y direito, curva verde), definida como o número de mensagens enviadas entre rotações consecutivas. Quanto menor a janela, menor o conjunto de mensagens potencialmente expostas sob comprometimento de uma chave Megolm, conceito análogo ao *Vulnerable Message Set* de [Auerbach et al. 2025].

Há três padrões fundamentais. (i) *Relação inversa entre segurança e eficiência*, Paranoid oferece janela de segurança de 25 mensagens, mas consome 293,9 kB de banda acumulada em rotações (barras vermelhas). Relaxed expande a janela 10× (250 mensagens, curva verde) mas reduz o custo para 29,4 kB, um *trade-off* bidirecional de 10×: a janela de segurança cresce linearmente com o intervalo de rotação, enquanto o custo de banda (barras vermelhas) decresce na proporção inversa. (ii) *Escalabilidade amplifica o trade-off*, o custo absoluto varia dramaticamente com o tamanho da sala: Large (150) com Paranoid consome 11,88 MB de banda em rotações, enquanto DM (2) consome apenas 32,7 kB (diferença de 363×). Isso evidencia que políticas conservadoras são economicamente viáveis apenas em salas pequenas. (iii) *Linearidade do custo*, o *overhead* acumulado escala linearmente com o número de rotações — $O(R)$ — e com o número de receptores — $O(N)$ —, resultando em custo total $O(N \times R)$. Para 1.250 mensagens em Large/Paranoid (50 rotações × 149 receptores), o custo atinge 11,88 MB.

6.3. Implicações para Políticas Adaptativas

Os resultados evidenciam que políticas fixas de rotação não são ideais em cenários heterogêneos. Em salas pequenas (DM, SmallGroup), mesmo políticas conservadoras como Paranoid impõem *overhead* aceitável: SmallGroup/Paranoid consome apenas 293,9 kB de banda e 19,77 ms de tempo acumulado para rotações, recursos negligenciáveis em dispositivos modernos. Em contraste, em salas grandes (Medium, Large), políticas con-

servadoras tornam-se proibitivas: Large/Paranoid consome 11,88 MB de banda e 834 ms de tempo, *overhead* que pode degradar a experiência de usuário em conexões lentas (e.g., 3G rural com 500 kbps de *upload* levaria ≈190 s apenas para enviar controle de rotações).

Uma estratégia adaptativa poderia ajustar dinamicamente a frequência de rotação com base em: (i) **tamanho da sala**, rotações mais frequentes em salas pequenas — onde o custo $O(N)$ é baixo —, menos frequentes em salas grandes para mitigar o custo quadrático; (ii) **taxa de envio de mensagens**, se a taxa é baixa (< 10 mensagens/hora), rotações baseadas em tempo (e.g., a cada 1 hora) evitam rotações desnecessárias; (iii) **restrições de rede**, priorizar eficiência (Relaxed/Balanced) sobre segurança temporal em conexões de baixa largura de banda; (iv) **modelo de ameaça**, salas com requisitos de alta segurança (e.g., comunicações governamentais, *whistleblowers*) justificam políticas conservadoras (Paranoid/PQ3) apesar do custo, enquanto salas públicas podem tolerar janelas de segurança maiores (Balanced/Relaxed) em troca de eficiência.

7. Conclusão

Os resultados evidenciam que adotar PQC no Matrix é tecnicamente viável, sem *overhead* em mensagens de usuário, mas impõe custos concentrados no plano de controle criptográfico. O *overhead* de largura de banda é o gargalo principal no plano de controle: +548% na fase de *setup* e +252% em rotações, enquanto é moderado no tempo (+222% *setup*, +133% rotações). Essa assimetria decorre das características das primitivas CRYSTALS-Kyber, que geram *payloads* grandes, de 1,568 kB para CRYSTALS-Kyber-1024 e 1,184 kB para CRYSTALS-Kyber-768, mas computacionalmente rápidas.

A análise sistêmica revelou que rotações de chaves dominam o consumo de largura de banda em sessões com múltiplas mensagens, com custos escalando quadraticamente com o número de participantes — $O(N \times R)$. Para SmallGroup ($N = 7$) com política Paranoid, rotações acumulam 293,9 kB, enquanto para Large ($N = 150$) o custo atinge 11,88 MB. Políticas fixas de rotação mostraram-se inadequadas para cenários heterogêneos: salas pequenas suportam políticas conservadoras com *overhead* aceitável, mas salas grandes tornam-se proibitivas (190 s para enviar controle em conexões 3G rurais).

Trabalhos futuros incluem: (i) implementação de políticas adaptativas de rotação baseadas em contexto (tamanho de sala, taxa de mensagens, restrições de rede, modelo de ameaça); (ii) avaliação experimental em dispositivos IoT e móveis com recursos restritos; (iii) integração com a especificação oficial do Matrix para validação em produção; e (iv) análise formal das propriedades de segurança do *wrapper* híbrido proposto.

Disponibilidade de Artefatos

Os artefatos (código-fonte, dados brutos e *scripts* de análise) foram organizados e disponibilizados em `https://github.com/mdo-br/matrix_pqc` seguindo os princípios e práticas de Ciência Aberta.

Referências

Auerbach, B., Dodis, Y., Jost, D., Katsumata, S., and Schmidt, R. (2025). How to compare bandwidth constrained two-party secure messaging protocols: a quest for a more efficient and secure post-quantum protocol. In *Proceedings of the 34th USENIX Conference on Security Symposium*, pages 6717–6736.

Aydeger, A., Zeydan, E., Yadav, A. K., Hemachandra, K. T., and Liyanage, M. (2024). Towards a quantum-resilient future: Strategies for transitioning to post-quantum cryptography. In *2024 15th International Conference on Network of the Future (NoF)*, pages 195–203. IEEE.

Barker, E. (2020). Recommendation for key management: Part 1 – general (sp 800-57 part 1, rev 5). Technical report, National Institute of Standards and Technology (NIST).

Cohn-Gordon, K., Cremers, C., Dowling, B., Garratt, L., and Stebila, D. (2016). A formal security analysis of the signal messaging protocol. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 1339–1353.

Dodis, Y., Jost, D., Katsumata, S., Prest, T., and Schmidt, R. (2025). Triple ratchet: A bandwidth efficient hybrid-secure signal protocol. In *Annual Int. Conference on the Theory and Applications of Cryptographic Techniques*, pages 302–331. Springer.

Duits, I. (2019). The post-quantum signal protocol: Secure chat in a quantum world. Master's thesis, University of Twente.

Kret, E. and Schmidt, R. (2023). The pqxdh key agreement protocol. *https://signal.org/docs/specifications/pqxdh/pqxdh.pdf*.

Martins, J. A., Rego, P. A., de Macêdo, J. A., Silva, F. A., and Lagrota, V. (2026). Matrix protocol: a comprehensive systematic mapping study. *Journal of Cloud Computing*, 15(1):20.

Martins, J. A. P., Rego, P. A. L., Macêdo, J. A. F. d., Andrade, R. M. C., Bonfim, M. S., Ivo, R. F., Costa, V. L. R. d., Pacheco, R. P., and Silva, F. A. P. d. (2025). Protocolo matrix: Conceitos, arquitetura, aplicações e desafios. In *Jornada de Atualização em Informática 2025*, pages 1–46. Sociedade Brasileira de Computação (SBC).

Matrix.org (2019). Olm: A cryptographic ratchet. `https://gitlab.matrix.org/matrix-org/olm/-/blob/master/docs/olm.md`. libolm repository.

Matrix.org (2022). Megolm group ratchet. `https://gitlab.matrix.org/matrix-org/olm/-/blob/master/docs/megolm.md`. libolm repository.

Matrix.org (2024). Matrix specification. `https://spec.matrix.org/`.

NIST (2022). Post-quantum cryptography standardization. Technical report, National Institute of Standards and Technology. Acesso em: `https://csrc.nist.gov/projects/post-quantum-cryptography`.

Schmidt, R. and Connell, G. (2025). The ml-kem braid protocol. Signal Messenger Specification. Available at: `https://signal.org/docs/specifications/mlkembraid/`.

Seufert, M., Wassermann, S., and Casas, P. (2023). Share and multiply: Modeling communication and generated traffic in private whatsapp groups. In *Proceedings of the 23rd ACM Internet Measurement Conference*, pages 642–657.

Shor, P. W. (1994). Algorithms for quantum computation: discrete logarithms and factoring. *35th Annual Symposium on Foundations of Computer Science*, pages 124–134.

Stebila, D. (2024). Security analysis of the imessage pq3 protocol. *Cryptology ePrint Archive*.