Monitor do Debate Político no Meio Digital
Grupo de Políticas Públicas para o Acesso à Informação
Escola de Artes Ciências e Humanidades - USP

Nota Técnica 10

Rastreamento de mensagens virais no WhatsApp

Márcio Moretto Ribeiro

17 de Agosto de 2020

Resumo: *No dia 30 de junho de 2020 o senado aprovou o Projeto de Lei 2.630/2020, PL das Fake News, e o encaminhou para a Câmara dos Deputados. O projeto tem como intenção declarada o combate a disseminação de desinformação e a promoção da transparência em relação a conteúdos patrocinados pelo poder público. Seu artigo mais polêmico, o décimo, estabelece a obrigação de retenção de metadados de mensagens com encaminhamento massivo. Esses metadados devem permitir a identificação dos disseminadores de conteúdo criminoso*[1].

Nesta nota técnica explicitamos os princípios de segurança implementados nos principais aplicativos de transmissão privada de mensagens, descrevemos o modelo implementado pelo WhatsApp - aplicativo mais popular para esse fim - e argumentamos que a retenção desses metadados não é contraditória com os princípios descritos.

[1] https://www25.senado.leg.br/web/atividade/materias/-/materia/141944

1 Criptografia ponta a ponta

No modelo tradicional de comunicação pela internet, os usuários confiam em um intermediário para proteger sua comunicação. Por exemplo, quando uma pessoa, a chamaremos de Alice, manda uma mensagem por email para outra pessoa, Bob, a mensagem passa por um provedor de emails. A comunicação entre Alice e o provedor é protegida contra a interferência de terceiros. Alice troca de maneira segura uma *chave* com o provedor que é usada para trancar a mensagem. De maneira análoga, a comunicação entre o provedor e Bob também é criptografada. Nesse modelo o provedor é responsável por gerenciar as chaves da comunicação em seus servidores.

Quando em 2013 Edward Snowden denunciou que a agência de espionagem estadunidense NSA praticava tinha acesso a comunicação privada de bilhões de usuário no mundo, acadêmicos e ativistas avaliaram que isso era possível por conta da concentração do tráfego da comunicação de uma gigantesca parcela de usuários de internet pelos servidores de poucas empresas. Essa concentração criava *pontos únicos de falha* – ou seja, um lugar central no qual uma falha comprometeria toda comunicação. A solução defendida por esses atores foi desenvolver e popularizar ferramentas de comunicação com *criptografia ponta a ponta*.

No modelo com criptografia ponta a ponta, quando Alice envia uma mensagem para Bob por meio de um provedor, as chaves da criptografia ficam armazenadas nos dispositivos dos próprios usuários e não nos servidores do provedor do serviço. Assim o conteúdo das mensagens que trafegam por esses servidores não seria acessível nem mesmo pelas empresas que fornecem o serviço. A aposta é que, suprimindo os pontos únicos de falha, a NSA seria forçada a investir na vigilância contra alvos específicos ao invés da vigilância em massa.

2 Protocolos PGP e OTR

A criptografia ponta a ponta não é uma tecnologia nova. O protocolo PGP (*Pretty Good Privacy*), por exemplo, permite esse grau de segurança. O protocolo, desenvolvido no começo dos anos 90, estabelece que cada usuário gere uma par de chaves: uma chamada *chave pública* e outra chamada *chave privada*.

Quando Alice envia uma mensagem para Bob ela deve obter sua chave pública que tipicamente fica disponível em um *servidor de chaves*. Alice então verifica que a chave obtida é de fato a chave de Bob[2] e a utiliza para criptografar a mensagem. Opcionalmente Alice pode usar sua chave privada para assinar digitalmente a mensagem. Quando Bob recebe a mensagem de Alice ele usa sua chave privada, que fica guardada no seu dispositivo, para descriptografá-la. Opcionalmente Bob pode obter a chave pública de Alice para verificar sua assinatura.

O protocolo PGP busca garantir três propriedades:

1. *Confidencialidade:* uma mensagem criptografada com a chave pública de Bob deve ser compreensível apenas por quem possui a chave privada de Bob.

2. *Integridade:* quando Bob verifica uma mensagem assinada digitalmente por Alice, ele deve ter uma garantia de que essa mensagem é idêntica a que foi assinada.

3. *Autenticidade:* quando Bob verifica uma mensagem assinada digitalmente por Alice, ele deve ter uma garantia de que ela foi de fato produzida por Alice.

[2]O modelo implementado pelo protocolo PGP para verificar a autenticidade da chave é chamado rede de confiança. Embora essencial para a comunicação segura, julgamos que a explicação sobre autenticidade das chaves foge ao escopo desta nota.

No começo dos anos 2000, com a popularização de softwares para comunicação síncrona, um grupo de pesquisa canadense propôs um novo protocolo, o OTR (*Off The Record*)[3]. Esse protocolo difere do anterior em dois aspectos. Como a comunicação por meio do protocolo PGP sempre utiliza uma mesma chave, um ator malicioso que a obtenha poderá ler toda a comunicação passada. O protocolo OTR busca superar essa limitação. Além disso, quando Bob recebe uma mensagem assinada usando o protocolo PGP ele pode usá-la como prova de que ela foi escrita por Alice. O protocolo PGP simula a comunicação assinada por escrito em que o remetente guarda uma prova da comunicação que pode ser apresentada em público. O OTR busca simular uma conversa ao pé do ouvido em que, embora as partes tenham garantias da autenticidade da comunicação, essa garantia não pode ser usada como prova da comunicação para um terceiro. Seguindo o jargão técnico, além de garantir a confidencialidade, a autenticidade e a integridade, o protocolo OTR busca as seguintes garantias:

4. *Sigilo futuro:* mesmo que uma chave de comunicação seja comprometida, toda a comunicação passada continua protegida.

5. *Negação plausível:* quando Bob recebe uma mensagem de Alice, ela não pode ser usada para convencer um terceiro da autoria de Alice.

3 Encaminhamento de arquivos pelo WhatsApp

Originalmente o WhatsApp foi concebido como uma ferramenta de comunicação interpessoal (um para um). Seguindo esse espírito, com a crise provocada pelas acusações de Edward Snowden, o protocolo desenvolvido pela WhisperSystems para o Signal[4] foi implementado no WhatsApp em 2016. O protocolo do Signal é uma evolução do OTR que garante as mesmas cinco noções de segurança. Com o tempo outras funcionalidades foram implementadas no WhatsApp o que descaracterizou seu caráter interpessoal: grupos, envio de mensagens para múltiplos usuários e encaminhamentos de mensagens, por exemplo. São essas funções que preocupam os legisladores que propuseram o PL das Fake News. Isso porque elas permitem que um pequeno grupo opere uma cadeia de grupos e espalhe conteúdo criminoso para milhões de usuários.

A preocupação com a disseminação de conteúdos nocivos gira principalmente em torno do encaminhamento de arquivos de mídia (imagem, áudio, vídeo etc.). No WhatsApp o encaminhamento de arquivos funciona da seguinte forma: o remetente gera as chaves para criptografar um o arquivo. Esse arquivo criptografado é armazenado no servidor. O remetente então envia uma mensagem criptografada ponta a ponta ao destinatário com um link para o tal arquivo e com as chaves para descriptografá-lo. O destinatário pode então recuperar o arquivo no servidor e localmente descriptografá-lo. Dessa forma, embora o arquivo esteja armazenado no servidor, o intermediário da comunicação (o WhatsApp no caso) não tem acesso a seu conteúdo[5] (Figura 3).

Recentemente o WhatsApp buscou limitar o encaminhamento de arquivos guardando nas mensagens um contador do número de encaminhamentos. Quando esse contador registra que um mesmo conteúdo foi encaminhamento mais de cinco vezes ele limita a capacidade de encaminha-

[3]https://otr.cypherpunks.ca/otr-wpes.pdf

[4]https://signal.org/docs/

[5]https://scontent.whatsapp.net/v/t61.22868-34/68135620_760356657751682_6212997528851833559_n.pdf/Artigo-t%C3%A9cnico-sobre-seguran%C3%A7a-do-WhatsApp.pdf?_nc_sid=2fbf2a&_nc_ohc=oXR1BTgKzeAAX-jKnQT&_nc_ht=scontent.whatsapp.net&oh=8ec85baf94bf091d67149447ed824ef4&oe=5F3C8D0C

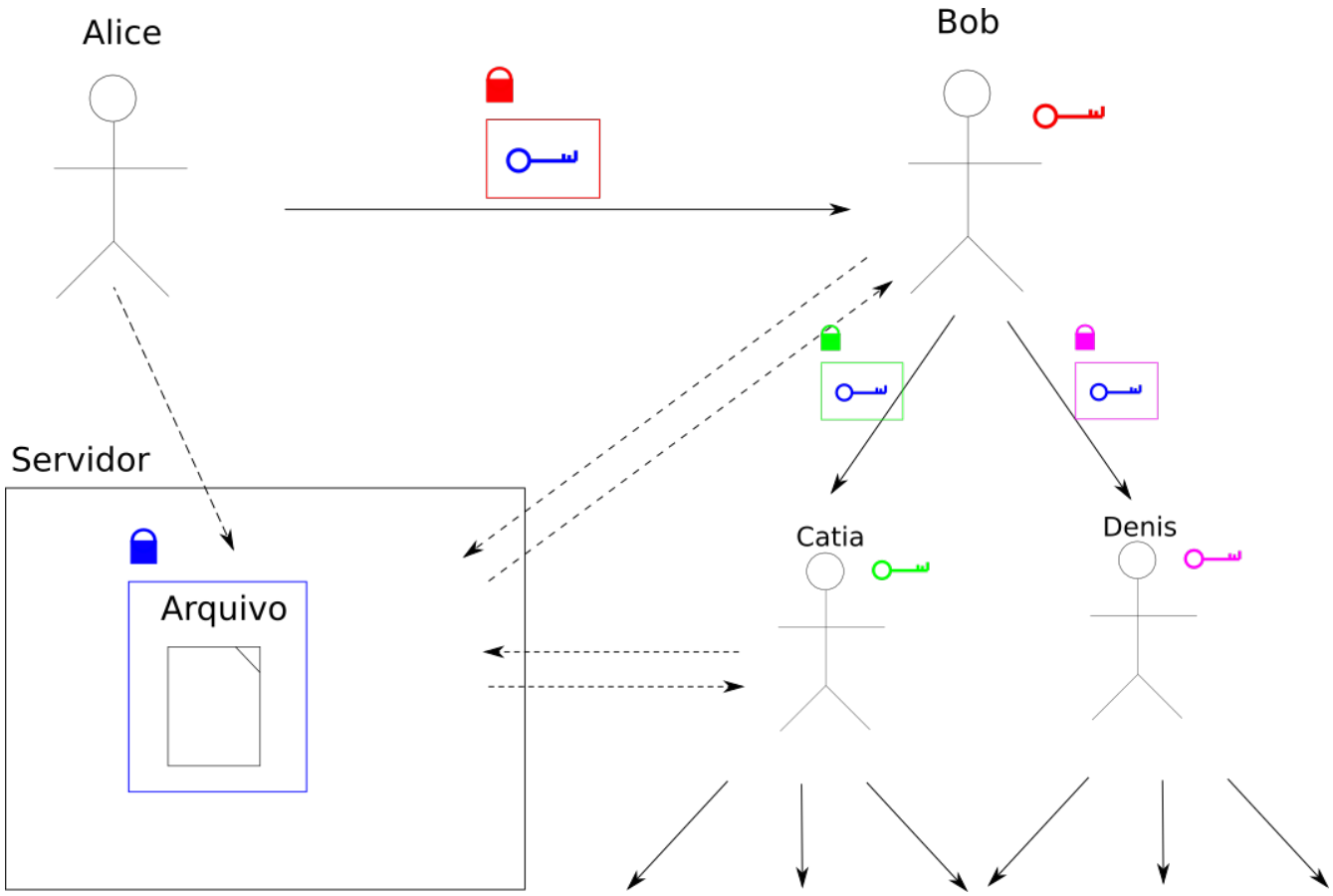

Figura 1: Os arquivos de mídia ficam armazenados no servidor criptografados. A chave para abri-los é encaminhada com as mensagens protegida pela critpografia ponta a ponta.

mento para apenas um novo destinatário por vez. Como esse contador fica guardado dentro da mensagem ele fica protegido pela criptografia ponta a ponta e não pode ser acessado pela empresa[6].

4 Dois modelos de rastreamento

O projeto de lei da forma como aprovada pelo senado impõe ao WhatsApp a retenção de metadados de cada usuário que requisitou acesso a algum arquivo armazenado nos seus servidores. Os metadados retidos devem conter um identificador do usuário e o horário da requisição. Isso pode ser feito sem comprometer a criptografia da comunicação conforme explicado na seção anterior.

Essa solução não fere nenhum dos princípios criptográficos que o protocolo busca garantir. A criptografia ponta a ponta não precisa ser violada e a empresa continuaria não tendo acesso ao conteúdo da comunicação dos usuários. Assim, apenas quem recebe uma mensagem com conteúdo suspeito poderia fazer uma denúncia. Neste caso uma autoridade poderia requerer ao WhatsApp os metadados armazenados e com isso identificar os primeiros usuários que encaminharam o arquivo. Como os metadados armazenados são dos que requisitaram acesso ao arquivo, o usuário que enviou-o ao servidor não terá seus dados expostos. Isso é desejável porque quem envia um arquivo pode fazê-lo para fins de uma conversa interpessoal e, portanto, deve ter seus dados protegidos. Quem encaminha o arquivo tira-o do contexto privado e assim assume o risco de viralizá-lo. Dessa forma, é ao mesmo tempo possível responsabilizar quem opera esquemas de disseminação de conteúdo criminoso e proteger a comunicação interpessoal.

Os termos de uso do aplicativo não impedem que a empresa armazene e processe os metadados da comunicação. Apesar disso, poderíamos argumentar que a legislação não deveria exigir que a empresa guarde mais dados pessoais. Tal base de dados é de fato sensível e poderia ser usada para fins indevidos.

A forma como a empresa implementou a restrição do número de encaminhamentos sugere uma outra solução que exigiria uma pequena mudança no texto da lei. As identidades dos usuários poderiam ficar armazenadas dentro da proteção da criptografia ponta a ponta. Assim a responsabilidade por essa informação não estaria sob responsabilizada centralizada da empresa, mas distribuída entre aqueles que recebem a mensagem encaminhada.

5 Conclusão

A implementação da criptografia ponta a ponta pelo WhatsApp foi uma resposta ao abuso da NSA denunciado em 2013. Ao armazenar as chaves de criptografia diretamente nos dispositivos dos usuários, evita-se a formação de pontos únicos de falha. Embora a decisão tenha sido tomada pela empresa, essa solução foi construída e amplamente debatida pela comunidade ativista e acadêmica em conferências e fóruns como as criptoparties e criptorraves. O protocolo implementado respeita princípios estabelecidos e testados ao longo dos anos e, inclusive, foi primeiramente implementado em um aplicativo de código aberto. A quebra desse protocolo seria um grande retrocesso.

Por outro lado, funcionalidades que facilitam a viralização de conteúdo não seguiram a mesma lógica. As capacidades de enviar mensagens para múltiplos usuários de uma vez e de encaminhar mensagens foram implementadas *a posteriori*. Em particular, a decisão de implementar o encaminhamento de mensagens preservando a identidade dos intermediários não passou pelo mesmo processo de debate público.

[6] https://faq.whatsapp.com/general/coronavirus-product-changes/about-forwarding-limits/

Do ponto de vista técnico, o WhatsApp não implementa técnicas de anonimização. Há acúmulo científico e técnico sobre as formas mais seguras de proteger metadados, mas a proteção deles nunca foi uma preocupação da empresa. Do ponto de vista político, não debatemos o suficiente se o encaminhamento anônimo é desejável. Uns defendem que deve ser preservada a identidade dos intermediários em uma comunicação viral. Esses entendem que a viralização de conteúdo é uma ferramenta empoderadora para grupos subalternos. Outros defendem que ao proteger a identidade dos intermediários facilitamos que um pequeno grupo de atores possa espalhar conteúdo criminoso para milhões de pessoas sem ser responsabilizado. Para esses, a funcionalidade deturpa o caráter originalmente interpessoal da ferramenta.

Essa nota técnica buscou contribuir com o debate mostrando que seria possível rastrear uma mensagem sem violar os princípios de segurança construídos pela comunidade ao longo dos anos. Apresentamos duas soluções para isso: na primeira as informações ficam sob responsabilidade da empresa, na segunda, esses dados ficam disponíveis para os usuários que receberem as mensagens.