ARTIGO

O PAPEL BASILAR DA CRIPTOGRAFIA NA SEGURANÇA DE DADOS

POR
Ricardo Dahab (Unicamp)
rdahab@unicamp.br

Começamos por uma questão de nomenclatura: o termo "segurança da informação" sempre teve a preferência em relação a "segurança de dados", na literatura da área. Em consonância com o currículo de referência da SBC, vamos adotar aqui a segunda forma, "segurança de dados", mesmo porque os dados de interesse podem ser totalmente desprovidos de informação, no sentido matemático da palavra. .

Os textos modernos trazem, como requisitos básicos de segurança de dados, o sigilo, a integridade, a autenticidade e a irretratabilidade. *Sigilo* refere-se à capacidade de leitura de um dado somente pelas partes autorizadas para tal. Integridade é a garantia de que um dado não tenha sido modificado desde a sua gênese. *Autenticidade* é a propriedade de que o dado seja, não somente íntegro, como também tenha sua origem comprovada. Finalmente, a *irretratabilidade* refere-se à impossibilidade de que a autoria de um dado possa ser negada posteriormente pelo seu autor. Dentre esses requisitos, a autenticidade (e integridade) é necessária com maior frequência, já que protege usuários contra dados falsos ou corrompidos. Sigilo nem sempre é necessário, assim como a irretratabilidade.

Há um grande número de outros requisitos de segurança, que são variações ou combinações desses quatro. A disponibilidade também é um requisito desejável de qualquer sistema de informação, mas envolve uma discussão mais ampla do que o nosso foco aqui, restrito à segurança.

Técnicas criptográficas para o provimento de requisitos básicos

Classicamente, o termo "Criptografia" refere-se à transformação da grafia original de um dado em outra, que oculte a informação nele contida. Mais recentemente, no entanto, a partir da década de 1970, esse termo designa todo um conjunto de técnicas matemáticas nas quais se baseia a maioria das aplicações para provimento de segurança de dados. Assim, a expressão "cifração de dados" tem a preferência sobre "criptografia de dados", evitando a ambiguidade. Alguns autores também usam a expressão "encriptação", em vez de cifração, pela sua proximidade com o termo *encryption*, em inglês.

A Criptografia dispõe de três técnicas, também básicas, para o provimento dos requisitos de segurança descritos acima: cifração de dados, resumo (*hash*) criptográfico, e assinaturas digitais.

A *cifração* consiste em substituir os caracteres (bits) de um dado, por outros, cuja relação com os caracteres originais seja impossível de se obter na prática, exceto pelos possuidores de um dado crucial, chamado de chave, que torna possível a *decifração* do dado cifrado. No caso da *cifração simétrica*, a *chave secreta* é única e de conhecimento restrito aos autorizados a cifrar/decifrar o dado. No caso da *cifração assimétrica*, ou de *chave pública*, a chave de cifração é diferente da chave de decifração- a primeira é pública, de conhecimento generalizado, e a segunda, a *chave privada*, é de conhecimento exclusivo do seu dono. A despeito de haver uma relação matemática entre as duas, é claro que a chave privada não deve ser facilmente dedutível a partir da chave pública. Primariamente, a cifração provê o requisito de sigilo, mas também pode ser usada para prover integridade e autenticidade.

Resumo (ou *hash*) é uma técnica oriunda da área de estruturas de dados. Trata-se de comprimir um texto (dado) de comprimento arbitrário, produzindo um texto de comprimento fixo, um resumo, portanto, com certa garantia de que dados diferentes resultem em resumos diferentes. É importante notar que dois ou mais textos podem ter o mesmo resumo; tais *colisões* são inevitáveis, pela grande diferença entre a dimensão do espaço de textos e a do espaço de resumos. Se colisões puderem ser evitadas, um resumo funcionará como um identificador curto de um dado arbitrariamente longo, o que é muito útil em buscas em estruturas de dados. No contexto da Criptografia, um resumo deve ter propriedades adicionais, de forma que sejam computacionalmente inviáveis: (i) produzir um texto a partir do seu resumo; e (ii) obter colisões, isto é, dois textos com o mesmo resumo, por qualquer meio. Resumos criptográficos provêm os requisitos de integridade e autenticidade.

A *assinatura digital* é a única técnica que provê irretratabilidade. A partir de um dado e da sua chave privada, um usuário *U* produz (sua) assinatura *S* do dado. Essa assinatura pode ser verificada por meio de um procedimento que combina o dado, a assinatura *S* e a chave pública de *U*, produzindo verificação positiva se, e somente se, *S* tenha sido produzida como esperado, conjugando o dado e a chave privada de *U*. O fato de a chave privada ser de conhecimento exclusivo do seu dono traz a garantia necessária para a irretratabilidade do ato da assinatura.

Algoritmos e protocolos

As técnicas descritas acima são implementadas por meio de algoritmos, que trazem em si as garantias matemáticas para o provimento dos requisitos de segurança. Tais algoritmos são combinados, na forma de um protocolo criptográfico que envolve, na maioria dos casos, duas ou mais partes realizando uma troca de mensagens. Em alguns casos a troca é muito simples; em outros, toma a forma de uma longa sequência de mensagens trocadas entre várias partes, algumas delas seres humanos e, outras, meros processos sendo executados por um dispositivo computacional.

Assim como algoritmos criptográficos derivam sua força de resultados matemáticos que necessitam ser demonstrados rigorosamente, protocolos criptográficos também devem ter sua robustez atestada, por técnicas que envolvem, muitas vezes, algum tipo de jogo entre as partes na presença de um adversário cujo objetivo é burlar o requisito de segurança em questão. O atestado de robustez deve demonstrar que a vitória do adversário só pode ocorrer com probabilidade infinitamente pequena. A isso chamamos de *segurança demonstrável* de um protocolo.

Protocolos criptográficos estão no cerne da maioria das soluções para provimento de cibersegurança. Além de proteger requisitos básicos de segurança, são usados em aplicações de suporte, como o estabelecimento, distribuição, gerenciamento e certificação de chaves criptográficas, ou requisitos mais elaborados como identificação/autenticação de entidades, e anonimato, controle de acesso e autorização, entre outros.

A matemática das técnicas criptográficas

A Criptografia clássica, sinônimo de cifração simétrica, é baseada em métodos que combinam texto e chave secreta principalmente por meio de substituições de caracteres e/ou permutações das suas posições no texto. Isso se mantém até hoje. Mesmo o método padrão dessa classe, o *AES (Advanced Encryption Standard)* usa tais operações, mas de forma muito mais sofisticada do que seus antecessores e, mais importante, usando chaves muito mais longas, de pelo menos 128 bits: o esforço computacional para encontrar a chave correta, tentando todas as possibilidades, é tarefa impensável, exigindo um número de operações da ordem de 2^{128}. Assim, os esforços de *criptoanálise*, isto é de tentativas de análise (ou "quebra") desses métodos tornando-os não efetivos, são quase sempre voltados

à redução desse esforço computacional imenso, por meio da identificação de atalhos no projeto usando técnicas como a criptoanálise *diferencial e linear*. É, em si, um trabalho também hercúleo, como buscar uma agulha num palheiro.

Algoritmos para resumos criptográficos surgiram concomitantemente ao advento da cifração de chave pública, como técnica de suporte às assinaturas digitais. Desde então encontraram um sem-número de aplicações, muito além do seu uso inicial, como a aleatorização de eventos, simulação de moedas, entre outras. Os métodos para construção de funções de resumo evoluíram nos últimos anos, dos métodos iterativos envolvendo funções de compressão, para os atuais *métodos esponja*. Seja qual for o caso, sua construção tampouco depende de conceitos matemáticos profundos. Sua criptoanálise se resume à redução da complexidade da busca por colisões.

Sem dúvida, a grande revolução nas técnicas criptográficas e teorias subjacentes foi produzida pelo advento da criptografia de chave pública na segunda metade da década de 1970, com a introdução do método RSA (das iniciais de Rivest, Shamir e Adleman). A robustez criptográfica do RSA é baseada na dificuldade do problema da *fatoração* de números inteiros muito grandes, resultantes do produto de dois números primos também de grande magnitude. Falamos aqui de números de 2 a 4 mil bits. Até esta data não se conhecem métodos eficientes para resolver essa tarefa usando computadores convencionais. Existem, no entanto, algoritmos que serão capazes de resolver eficientemente esse problema usando computadores quânticos, quando estes se tornarem uma realidade prática.

Outro problema matemático com a mesma característica de vulnerabilidade à criptoanálise quântica é o do *logaritmo discreto*. Trata-se de calcular logaritmos numa estrutura discreta (de grupos cíclicos), problema difícil no universo de grandes números e computadores convencionais. Métodos criptográficos baseados nesse problema ganharam popularidade a partir dos anos 1980, pelos ganhos de eficiência resultantes do uso de chaves menores que as dos métodos baseados em fatoração, possibilitando seu emprego em dispositivos com poucos recursos. O método mais popular dessa classe é o de *curvas elípticas*. Sua implementação é bem mais complexa do que a do RSA, necessitando de dois tipos de aritméticas em estruturas algébricas distintas. Daí, a sua implementação cuidadosa, explorando aspectos matemáticos e de engenharia de algoritmos sobre plataformas específicas, é um problema desafiador e muito interessante.

Motivada pela ameaça quântica, a comunidade de pesquisa criptográfica tem buscado ajuda em outra classe de algoritmos, a dos problemas NP-completos. Esses são problemas historicamente resistentes à resolução eficiente, mas têm a característica peculiar de que, uma vez encontrado um algoritmo eficiente para resolver um deles, então algoritmos eficientes podem ser encontrados para todos eles, mediante um esforço adicional de baixa complexidade. Essa propriedade dá a essa classe um status de alta

dificuldade. Alguns desses problemas, notadamente os oriundos da Teoria dos Reticulados e da Teoria dos Códigos (Corretores de Erros), servem de base a alguns métodos hoje em processo de padronização, após serem aprovados em uma competição pública promovida pelo *National Institute of Standards and Technology dos EUA [NIST]*. A teoria e implementação dessa nova safra de métodos ainda está sob escrutínio intenso da comunidade acadêmica, dada a sua relativa novidade. Em particular, o tamanho das chaves e a obtusidade de alguns métodos são empecilhos à sua adoção imediata. Alguns deles, após aprovação preliminar, foram posteriormente quebrados de forma simples, o que mostra o valor dessa forma de seleção aberta ao público.

Ainda na seara quântica, um método para estabelecimento de chaves secretas proposto em 1984 por Bennett e Brassard, utiliza exclusivamente fenômenos quânticos, sem recorrer a qualquer teoria matemática. Em vez de bits, a informação é codificada e transmitida como partículas elementares (fótons são as mais usadas), de forma que um adversário não consiga "ler" essas partículas sem causar distúrbios nas suas propriedades quânticas. Assim, em vez de esconder (cifrar) um dado em trânsito, o efeito é o de impedir sua leitura não autorizada sem despertar suspeitas.

Para além da criptoanálise

Como vimos, a criptoanálise é realizada, ou pela dedução sistemática de chaves secretas ou privadas, ou pela descoberta de falhas nos pressupostos matemáticos de uma técnica ou na concepção de um protocolo. Há, porém, outras formas de ataque, que visam a implementação de um método numa plataforma computacional específica. Esses são resultantes de falhas na escrita do código (programa) da implementação ou advindas do vazamento de informação sensível por canais inesperados como consumo de energia, tempo de execução e radiação eletromagnética. Tais canais são conhecidos como *canais laterais* e a prevenção de ataques deste tipo é feita pela cuidadosa implementação dos algoritmos, de forma a evitar flutuações nesses vazamentos que possam ser identificados com os bits das chaves ou outras informações sensíveis. Todo projeto atual de algoritmo criptográfico inclui a apresentação de contramedidas por meio de uma implementação imune a tais ataques. Frequentemente, tal implementação incorre em algum tipo de ineficiência quando comparada a uma implementação que vise somente eficiência.

Aplicações avançadas

Até muito recentemente, criptografia era um termo conhecido na comunidade acadêmica e profissional de computação, mas pouco conhecida fora dela. Após o advento das criptomoedas e dos comunicadores instantâneos seguros, como WhatsApp e similares, o prefixo *cripto* pode ser encontrado facilmente em qualquer portal de notícias ou mesa de bar em que se possa pagar a conta usando criptomoedas. De fato, essa *criptomania* se

deve ao desenvolvimento de protocolos mais sofisticados para não só a simulação de moedas e o sigilo ponta-a-ponta dos comunicadores instantâneos, mas também para outras aplicações sofisticadas, como:

1. computação distribuída confiável, usando protocolos conhecidos como *multi-party computation*, ou MPC, que possibilitam a interação confiável de várias partes num mesmo processo, protegendo eventuais informações sensíveis de cada parceiro, como sua identidade, chaves criptográficas, ou dados de qualquer natureza;
2. assinatura automática de contratos, usando protocolos de conhecimento zero (*zero-knowledge* protocols), que possibilitam a verificação de propriedades de um dado sem revelar o seu conteúdo;
3. computação com dados cifrados, usando cifração homomórfica (*homomorphic encryption*), que possibilita a realização de cálculos sobre dados cifrados, resguardando o sigilo do resultado e mesmo a natureza desses cálculos. Essa aplicação é muito útil para dados sensíveis armazenados em nuvens.

Considerações práticas

O emprego de técnicas criptográficas requer cuidados especiais, como já vimos, com a necessidade de prevenção de ataques por canais laterais. É possível escrever uma peça de software criptográfico totalmente aderente do ponto de vista funcional, mas recheada de vulnerabilidades. Geralmente, desenvolvedores de software não têm qualquer treinamento em criptografia, mas são deixados à vontade para desenvolverem código ad hoc ou que usam bibliotecas criptográficas, sem a assistência de alguém com o devido treinamento em criptografia. Assim, é possível que:

1. o uso de bibliotecas seja feito de forma incorreta ou com opções de compilação que introduzam vulnerabilidades como a exposição de canais laterais;
2. sejam usadas versões defasadas de bibliotecas, com vulnerabilidades corrigidas somente em versões posteriores;
3. código ad hoc seja inseguro, com métodos de cifração ingênuos, inventados por iniciativa do desenvolvedor;
4. código seja extremamente ineficiente, introduzindo atrasos insuportáveis para a aplicação a que se destina.

Esses são somente alguns dos percalços, retirados de casos reais, no desenvolvimento de software criptográfico. Por isso, é importante a observância de padrões e melhores práticas da área [NIST2].

Vale notar, finalmente, que uma tendência que cresceu muito nos últimos anos é a transferência de parte da responsabilidade pela segurança de dados para o hardware. É comum, hoje, o emprego de

hardware seguro para armazenamento e computação de dados e trechos sensíveis de aplicações. Tais peças de hardware podem ser dispositivos *stand alone*, ou placas e circuitos integrados, dependendo da criticidade da informação. *Hardware security modules*, por exemplo, são dispositivos *stand alone* dotados de mecanismos de proteção que incluem sensores diversos e provisão para destruição de dados em caso de invasão do dispositivo.

Implicações para o ensino hoje e no futuro

De toda a nossa discussão prévia, fica evidente a necessidade de introdução de disciplinas relacionadas à Criptografia nos cursos de graduação. Também fica claro que não se trata de criar uma só disciplina que abarque toda a gama de material que mal pincelamos acima.

O ideal seria termos, inicialmente, uma disciplina que cubra, de forma introdutória os princípios, técnicas fundamentais, aplicações de suporte e principais aplicações modernas em uso no mundo. Para tal tarefa, não existe um texto em português suficientemente atualizado. Em inglês, talvez o melhor texto hoje seja o de autoria de Stinson e Paterson [SP].

Uma segunda disciplina seria desejável, tendo ou não a primeira como pré-requisito, que cubra com maior detalhe um maior número de algoritmos e protocolos, ressaltando aspectos de complexidade e demonstrações de segurança necessárias em cada caso. Para tal tarefa, um bom texto, mais denso, é o de Katz e Lindell [KL], um bom complemento ao de Stinson.

Com uma ou ambas as disciplinas acima como pré-requisitos, algumas ramificações são possíveis, em oferecimentos conjuntos na pós-graduação. A literatura de apoio será, necessariamente, composta de artigos e outras publicações recentes, além de livros-texto.

1. Uma disciplina teórica para alunos que queiram seguir uma rota de pesquisa, com forte ênfase nos aspectos de complexidade, teórica e prática de algoritmos e protocolos, e suas demonstrações de segurança. Imprescindível aqui é uma boa dose de computação quântica. Um bom texto para essa disciplina é o de autoria de Hoffstein et al. [HPS].
2. Uma disciplina para os que pretendem desenvolver atividades de implementação eficiente de métodos criptográficos, explorando diversas plataformas de mercado, das mais robustas às mais restritas em recursos. Referências adicionais são de autoria de Hankerson et al. [HVM], Koç [Koç], e Menezes et al. [MOV].
3. Uma disciplina voltada a aspectos sociais relacionados ao uso indiscriminado de criptografia, como privacidade, anonimato e eleições eletrônicas.
4. Uma disciplina voltada a gestores de TI, com cobertura mais voltada à gestão de ativos criptográficos e das aplicações que dependem fortemente desses ativos.

Referências

1. [Auma] J.-P. Aumasson. Serious Cryptography: A Practical Introduction to Modern Encryption. Primeira Edição (Segunda Edição disponível em outubro de 2024). No Starch Press, 2017.
2. [HPS] J. Hoffstein, J. Pipher, J. H. Silverman. An Introduction to Mathematical Cryptography. Springer, 2014.
3. [HVM] D. Hankerson , S. Vanstone , A. Menezes. Guide to Elliptic Curve Cryptography
4. [KL] J. Katz, Y. Lindell. Introduction to Modern Cryptography, Terceira Edição. CRC Press, 2020.
5. [Koc] Ç. K. Koç, ed. Cryptographic Engineering. Springer 2008.
6. [MOV] A. J. Menezes, P. C. van Oorschot, S.Vanstone. Handbook of Applied Cryptography. CRC Press, 1996. Disponível, reeditado em https://cacr.uwaterloo.ca/hac/
7. [NIST1] Post-Quantum Cryptography Standardization. https://csrc.nist.gov/projects/post-quantum-cryptography
8. [NIST2] Cryptographic Standards and Guidelines. https://csrc.nist.gov/projects/cryptographic-standards-and-guidelines
9. [SP] D. R. Stinson, M. Paterson. Cryptography: Theory and Practice, Quarta Edição. CRC Press, 2018.
10. [Wong] D. Wong. Real-world Cryptography. Manning, 2021.

RICARDO DAHAB é professor titular do Instituto de Computação da Universidade Estadual de Campinas, UNICAMP. Tem mestrado pela UNICAMP em Criptografia e doutorado pela Universidade de Waterloo, Canadá, em Combinatória e Otimização. Seus interesses de docência e pesquisa situam-se nas áreas de Algoritmos e Protocolos Criptográficos e Segurança da Informação. Desde 1995 tem publicado trabalhos científicos e orientado teses de doutorado e mestrado nessas áreas, várias das quais receberam prêmios e distinções. Atuou também na área de Teoria dos Grafos. Participou do projeto ICP-EDU, em parceria com a RNP, UFSC, UFMG e Kryptus Tecnologias de Segurança, do qual resultou o primeiro hardware de alta segurança (HSM) totalmente nacional. Junto com as comunidades de criptografia e segurança vem cooperando ativamente na consolidação dessas áreas no Brasil e na América Latina, participando da Comissão Especial de Segurança da SBC, da organização de eventos como o SBSeg, a Escola Avançada de Criptografia da Fapesp, o Latincrypt, a CANS e o PKC. Foi agraciado, em 2011 com o prêmio Zeferino Vaz de Excelência Acadêmica da UNICAMP e, em 2019, com o Prêmio Destaque da Comissão Especial de Segurança da Informação e de Sistemas Computacionais.