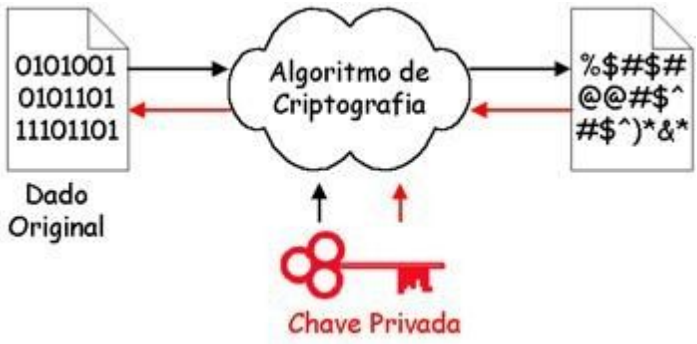
Como quebrar uma hash ou uma criptografia?

Um assunto muito abordado na Computação, em especial na Segurança Computacional, fundamental para qualquer tipo de proteção a sistemas. E hoje, abordarei um pouco mais aqui, nessa publicação.
Bom, segundo a definição da internet:

> criptografia
> *substantivo feminino*
> **1**.
> conjunto de princípios e técnicas empregadas para cifrar a escrita, torná-la ininteligível para os que não tenham acesso às convenções combinadas; criptologia.
> **2**.
> em operações políticas, diplomáticas, militares, criminais etc., modificação codificada de um texto, de forma a impedir sua compreensão pelos que não conhecem seus caracteres ou convenções.

As técnicas mais conhecidas de computação na sua compreensão de criptografia envolvem "chaves/keys". E nada mais é do que um grupo de bits baseado em um determinado algoritmo que codifica/decodifica dados e informações. Se o destinatário da informação/dado utilizar uma chave incompatível com a chave do remetente, ele não conseguirá acesso a informação/dado.



(UFRJ)

Nos príncipios da Criptografia, havia o uso apenas de um algoritmo específico que codificava a mensagem. Por tanto, o remetente necessitava apenas do conhecimento desse algoritmo para poder extrair a mensagem. O ponto negativo principal era se alguém não envolvido na comunicação tivesse posse desse algoritmo, ele também poderia decifrar a mensagem. Havia também outro problema, imagine uma pessoa X enviando um dado criptografado à pessoa Y. Esta última teria que ter o conhecimento do algoritmo usado. Agora, imagine uma pessoa N

que também precisasse receber esse dado da pessoa X, mas, a pessoa N não poderia saber qual é o dado a ser enviado à pessoa Y. Se a pessoa N tivesse acesso a informação enviada à pessoa Y, isso seria um problema. Por tanto, a solução é utilizar um diferente código/algoritmo para cada destinatário. Com o uso de chaves, um emissor pode usar o mesmo algoritmo (o mesmo método) para vários receptores. Basta que cada um receba uma chave diferente. Além disso, caso um receptor perca ou exponha determinada chave, é possível trocá-la, mantendo-se o mesmo algoritmo.

Você já deve ter ouvido falar de chave de 64 bits, chave de 128 bits e assim por diante.
Esses valores expressam o tamanho de uma determinada chave. Quanto mais bits forem utilizados, mais segura será a criptografia. Explica-se: caso um algoritmo use chaves de 8 bits, por exemplo, apenas 256 chaves poderão ser usadas na decodificação, pois 2 elevado a 8 é 256. Isso deixa claro que 8 bits é inseguro, pois até uma pessoa é capaz de gerar as 256 combinações (embora demore), imagine então um computador! Porém, se forem usados 128 ou mais bits para chaves (faça 2 elevado a 128 para ver o que acontece), teremos uma quantidade extremamente grande de combinações, deixando a informação criptografada bem mais segura.

Com a utilização de keys (chaves), um remetente pode usar o mesmo método, ou algoritmo, para vários destinatários. Apenas cada um recebendo uma key diferente. Uma outra consideração relevante é a de bits. Na internet vemos muitas chaves de 64 bits, 128 e etc. Esses valores condizem com o tamanho de uma chave. Ou seja, mais segura é uma criptografia se o tamanho da chave for longo. Na prática, vou exemplificar, se um algoritmo use chaves de 8 bits, somente 256 chaves poderão ser usadas na decodificação, afinal, matematicamente dizendo, 2 elevado a 8 é 256. Isso prova que 8 bits é inseguro, uma pessoa normal pode ser capaz de gerar as 256 combinações, mesmo que leve um bom tempo. Imagine uma super-máquina, programada para decodificar dados/quebrar criptografia...
Entretanto, se for usado 128 bits para chaves ou mais, você teria matematicamente 2 elevado a 128, ou seja, uma quantia de combinações muito grande, tornando mais segura a informação/dado criptografada.

Não entrarei muito a fundo dos tipos de criptografia e etc, quando eu mencioná-los, basta que você pesquise e terá mais informações. Vamos passar ao que interessa.

Como quebrar uma hash?

Calma, o que é uma hash?
Segundo o Wikipedia, um hash (ou escrutínio) é uma sequência de bits geradas por um algoritmo de dispersão, em geral representada em base hexadecimal, que permite a visualização em letras e números (0 a 9 e A a F), representando um nibble cada. O conceito teórico diz que "hash é a transformação de uma grande quantidade de dados em uma

pequena quantidade de informações". Essa sequência busca identificar um arquivo ou informação unicamente. Por exemplo, uma mensagem de correio eletrônico, uma senha, uma chave criptográfica ou mesmo um arquivo. É um método para transformar dados de tal forma que o resultado seja (quase) exclusivo. Além disso, funções usadas em criptografia garantem que não é possível a partir de um valor de hash retornar à informação original.

Como a sequência do hash é limitada, muitas vezes não passando de 512 bits, existem colisões (sequências iguais para dados diferentes). Quanto maior for a dificuldade de se criar colisões intencionais, melhor é o algoritmo.
Uma função de hash recebe um valor de um determinado tipo e retorna um código para ele. Enquanto o ideal seria gerar identificadores únicos para os valores de entrada, isso normalmente não é possível: na maioria dos casos, o contra-domínio de nossa função é muito menor do que o seu domínio, ou seja, x (o tipo de entrada) pode assumir uma gama muito maior de valores do que HASH(X) (o resultado da função de hash).
Os algoritmos de hash mais usados são os de 16 bytes (ou 128 bits, tamanho do message digest) MD4 e MD5 ou o SHA-1, de 20 bytes (160 bits).
Características de alguns algoritmos:

- **MD4:** Desenvolvido em 1990/91 por Ron Rivest, vários ataques foram detectados, o que fez com que o algoritmo fosse considerado frágil. Descrito na RFC 1320.
- **MD5:** O MD5 (Message-Digest algorithm 5) é um algoritmo de hash de 128 bits unidirecional desenvolvido pela RSA Data Security, Inc., descrito na RFC 1321, e muito utilizado por softwares com protocolo par-a-par (P2P, ou Peer-to-Peer, em inglês), verificação de integridade e logins. Existem alguns métodos de ataque divulgados para o MD5.
- **SHA-1 (Secure Hash Algorithm):** Desenvolvido pelo NIST e NSA. Já foram exploradas falhas no SHA.
- **WHIRLPOOL:** função criptográfica de hash desenvolvida por Paulo S. L. M. Barreto e por Vincent Rijmen (co-autor do AES). A função foi recomendada pelo projeto NESSIE (Europeu). Foi também adotado pelo ISO e IEC como parte do padrão internacional ISO 10118-3.

O processo é unidirecional e impossibilita descobrir o conteúdo original a partir do Hash. O valor de conferência ("Soma de verificação") muda se um único bit for alterado, acrescentado ou retirado da mensagem.

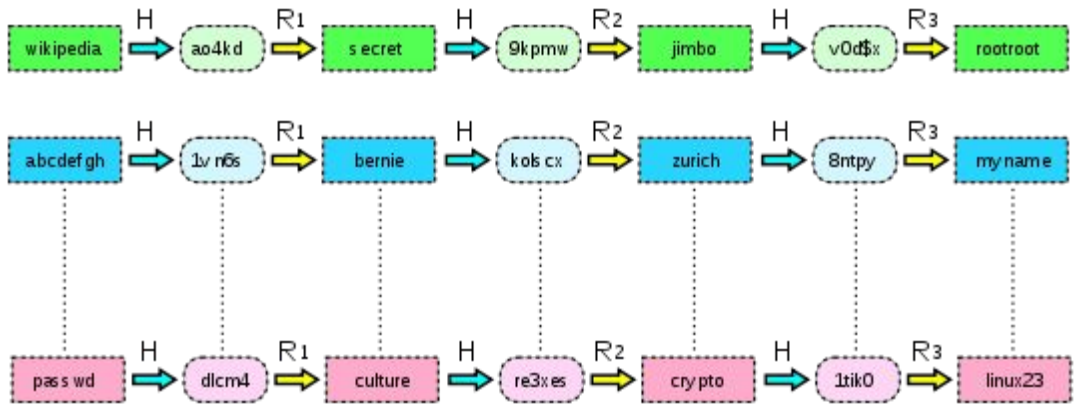
Bom, como quebramos uma hash?

Existem diversos modos para a quebra de uma hash, e cada tipo de criptografia (MD4, MD5 e etc) tem uma teoria por si só extremamente descritiva, que será impossível eu explicá-las todas aqui, por tanto, vale a pesquisa.

De fato, para cada método criptográfico, existe um método de decodificação. A maioria deles, por incrível que pareça, é "quebrável" com ferramentas online! Aliás, com o passar dos anos, mais recursos online foram programados para facilitar esse processo, computacionalmente. Matematicamente, existem infinitos modos de quebrar infinitos tipos de criptografia, chega a ser subjetivo eu mostrar várias equações aqui de quebras bem específicas de criptografia, por tanto, deixarei para você aprender em um curso ou maior pesquisa. Descreverei alguns, antes de passar para a parte computacional.

➢ Rainbow Table

Trata-se de uma tabela de consulta que utiliza um algoritmo de "transacção memória-tempo", permitindo-nos recuperar o texto original de uma password através de uma password digerida (hash), gerada por uma função criptográfica de digestão (hashing). É utilizado normalmente para que um ataque contra passwords digeridas seja possível. Uma "Salt" é normalmente aplicada em passwords digeridas para que o ataque seja mais difícil, muitas das vezes impossibilitando a sua descoberta.



(Wikipedia, enciclopédia livre)

Simplificada da tabela do arco-íris com 3 funções de redução
tabelas do arco-íris são um refinamento de um simples algoritmo anterior, por Martin Hellman que utilizaram a inversão de hashes, observando-se cadeias de hash pré-computadas.

➢ Dictionary-based

Um "ataque a dicionário" (em tradução livre), é uma técnica para quebrar um cifra ou mecanismo de autenticação apenas por tentar determinar sua chave (key) de decriptar ou a frase-chave (frase-passe) apenas tentando milhares ou talvez milhões de possibilidade factíveis, como palavras em um dicionário.
Um ataque de dicionário é baseado em tentar todas as seqüências de caracteres em uma lista pré-estabelecida, normalmente derivada de uma lista de palavras, como em um dicionário (daí o nome ataque a dicionário). Em contraste com um ataque de força bruta, onde uma grande proporção do espaço da chave é pesquisada sistematicamente, um ataque de dicionário tenta apenas as possibilidades que são considerados mais propensas a ter sucesso. Ataques a

dicionário muitas vezes são bem sucedidos, porque muitas pessoas têm uma tendência a escolher senhas curtas que são palavras comuns ou senhas comuns, ou variantes simples obtidas, por exemplo, anexando dígitos ou pontuação. Ataques a dicionário são relativamente fáceis de mitigar, por exemplo, escolhendo uma senha que não é uma simples variante de uma palavra encontrada em um dicionário ou uma lista de senhas mais usadas.

➢ Vigenerè Table

O francês Blaise de Vigenère (1523 - 1596) usou a criptografia como instrumento de trabalho durante anos. Com a idade de 39 anos resolveu abandonar a carreira diplomática e dedicar-se exclusivamente aos estudos. Em 1586 Vigenère publica seu livro de criptologia, o Traité des chiffres où secrètes manières d'escrire, no qual descreve detalhadamente sua cifra de substituição polialfabética com palavra-chave e apresenta uma tabela de alfabetos cifrantes que ficou conhecida como Carreiras de Vigenère.
O grande mérito do autor desta cifra foi o de aperfeiçoar um método que já tinha sido proposto por outros estudiosos, mas que precisava ser estruturado para oferecer a segurança necessária. Vigenère baseou-se em Alberti e Trithemius, como também em alguns contemporâneos, como Bellaso e Della Porta.
As características são: a Cifra de Vigenère pertence à classe de substituições com palavra-chave, o tipo da substituição é polialfabética monográmica (ou monográfica) porque faz uso de vários alfabetos cifrantes (polialfabética) aplicados individualmente (monográmica) aos caracteres da mensagem clara (plaintext), o método faz uso de chaves, que podem ser palavras ou frases, a segurança da cifra era alta para a época - hoje é considerada baixa. Foi somente em 1863 que o criptólogo alemão Kisiski descobriu como quebrar a cifra de Vigenère. O matemático inglês Charles Babbage já havia quebrado a cifra em 1854, fato que ficou desconhecido por muito tempo porque não publicou sua descoberta. Para se quebrar existe uma tabela de mesmo nome podendo ser encontrada em qualquer mecanismo de buscas de imagens.

➢ Caesar Cipher

Apesar de a criptologia estar bastante avançada na época, em 50 a.C. César usava um sistema bastante simples de substituição. Suetônio, escritor romano que viveu no início da era cristã (69 d.C.), em Vida dos Césares, escreveu a biografia dos imperadores romanos, de Júlio César a Domiciano.
Conta que Júlio César usava na sua correspondência particular um código de substituição no qual cada letra da mensagem original era substituída pela letra que a seguia em três posições no alfabeto: a letra A era substituída por D, a B por E, e assim sucessivamente.
Hoje em dia, porém, a denominação de Código de César é utilizada para qualquer cifra na qual cada letra da mensagem clara seja substituída por outra deslocada um número fixo de posições, não necessariamente três. Um exemplo é o código que, ainda segundo Suetônio, era usado por Augusto, onde a letra A era substituída por B, a B por C e assim sucessivamente.

Como o alfabeto romano possui 26 letras, é possível obter 26 alfabetos cifrantes diferentes, dos quais um (o do deslocamento zero) não altera a mensagem original. Cada um destes alfabetos cifrantes é conhecido como Alfabeto de César.
As características principais da Cifra de César são: classe substituição simples, tipo monoalfabético (porque usa apenas UM alfabeto cifrante), monogrâmica (porque trata cada UM dos caracteres individualmente), segurança baixíssima devido a complexidade rasa, uso aplicável apenas em textos muito curtos, e a criptoanálise baseada na característica estatística da língua, que é suficiente para decifrar o texto.

Agora vamos a considerações computacionais finais.

Pleno 2016 e ainda temos que ser criptólogos para quebrar cifras/criptografias? Não.
Hoje em dia, existem vários sites programados em PHP e Javascript, devidamente criados para facilitar a quebra de criptografia. Deixarei alguns links no final, juntamente a fonte.
Existem sites que quebram hashes MD4 (mdcrack openwall, md5decrypt, crackstation e etc), MD5 (mesmo), SHA-1 (hashkiller, md5hashing, dcode, hashcat e etc) e muito mais.

Bom, espero que tenham gostado e encarem esta publicação como um gostinho, um estopim para buscar mais e aprender mais. **CCJ OEAWL IZ URTJV!**

Autor: João Góes

Links externos:

1. www.numaboa.com.br/index.php?option=com_content&view=article&id=506&Itemid=134 (no final ainda há um tradutor do Código de Vigenère)
2. http://www.numaboa.com.br/criptografia/cifras/substituicoes/monoalfabeticas/simples/165-Codigo-de-Cesar (no final ainda há um tradutor do Código de César)
3. http://www.numaboa.com.br/escolinha/garranchos/91-informatizados/351-ascii (leitura interessante sobre ASCII - conceito importante na computação)
4. https://www.google.com.br/search?q=tabela+de+vigenere&safe=off&tbm=isch&tbo=u&source=univ&sa=X&ved=0ahUKEwjWs-LYmonKAhWIf5AKHa80ARAQsAQILA&biw=1252&bih=613 (pesquisa da tabela de Vigenère no Google Imagens)
5. http://penta.ufrgs.br/gere96/segur2/des.htm (muito interessante e completo também)
6. http://www.hardware.com.br/comunidade/criptografia-bits/769988/ (discussão interessante sobre o tempo de quebra de uma hash)

7. https://www.ucb.br/sites/100/103/TCC/12007/PalomaBarbosaFreire.pdf (abordagem matemática completa de todo o mundo criptográfico)
8. https://www.youtube.com/watch?v=vj7DpfQ_pa0 (indicação de um colega sobre os príncipios básicos da criptografia)
9. http://www.profcardy.com/logica/raciocinio.php?id=237 (exercício de criptografia disponível na internet)
10. http://md5online.org/ (quebra/crack de MD5)
11. https://danieldonda.wordpress.com/2011/04/08/criptografia-com-lgebra-linear-matriz esparte-1/ (para os amantes de matemática: criptografia com Álgebra Linear - Matrizes)
12. http://www.ebah.com.br/content/ABAAABQdwAE/criptografia-quantica-aplicacoes-militares (indicação de um amigo, já está na minha lista de próximas leituras, coloque na sua também (=)
13. http://www.mundodoshackers.com.br/ferramentas-para-quebra-de-senhas (referências simples em quebra de hashes com algumas ferramentas específicas)
14. https://crackstation.net/ (quase 18 bilhões de dados na tabela de comparação deles para quebra de criptografia e 16 tipos de criptografia suportados)
15. www.hashkiller.co.uk/ (quase 132 bilhões de dados na tabela de comparação deles para quebra de criptografia e 5 tipos de criptografia suportados)
16. http://www.onlinehashcrack.com/ (14 tipos de criptografia suportados e quebra de qualquer hash por experts)