Introdução à Criptografia:
Criptografia e Criptoanálise

Roteiros e tópicos para estudo por

Vinícius da Silveira Serafim
professor@serafim.eti.br
http://professor.serafim.eti.br

Palavras-chave: criptossistema, cifragem, decifragem, criptoanálise, Kerckhoffs

Criptografia

Normalmente quando pensamos em criptografia o que nos vem à mente é o objetivo (ou atributo) da segurança *confidencialidade*. No entanto, conforme vamos ver ao longo das aulas, a criptografia encontra-se na base de diversos controles que visam garantir também a *integridade*, *autenticidade* e *irrefutabilidade*.

A origem da palavra é grega e significa escrita (*grifos*) oculta (*kryptos*).

Criptografia é só parte da solução

Embora a criptografia seja a essência dos mais variados *controles* que visam garantir a segurança, deve-se ter em mente que ela por si só não constitui uma solução completa.

Conforme Ferguson e Schneier (FERGUSON, 2003): “É apenas uma pequena parte de um sistema de segurança”. Imagine que a criptografia é como uma fechadura, você pode comprar uma ótima fechadura (segura), porém a segurança da sua casa não depende apenas dela, mas sim de todo o resto da casa: janelas, paredes, da própria porta onde a fechadura será instalada, etc.

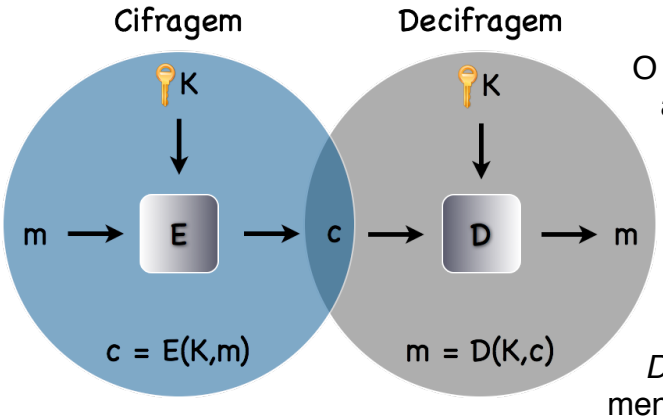
Matemática!

A matemática é a fundação da criptografia: teoria de conjuntos, complexidade computacional, probabilidade, estatística, etc. (PFLEEGER, 2006) Não precisamos nos aprofundar nessas questões para **usarmos** criptografia, o que torna viável estudarmos criptografia (em caráter introdutório) nessas duas aulas.

IMPORTANTE: **usar** é diferente de **criar!** Não invente seus próprios algoritmos e estude as formas corretas de utilizar os algoritmos que você escolher.

Criptossistema

Apesar de ser óbvia, segue a definição de um criptossistema: um sistema para cifragem (encriptação) e decifragem (decriptação). (PFLEEGER, 2006)



O modelo genérico de um criptossistema é apresentado na figura ao lado. Nela estão destacadas as duas operações: cifragem e decifragem.

Cifragem

Definição: é o processo de codificar uma mensagem de forma que seu significado não seja óbvio. (PFLEEGER, 2006)

O primeiro elemento necessário para execução desse processo é um algoritmo de cifragem (representado pela letra **E** de encriptação). O objetivo da cifragem é transformar a mensagem **m** em uma mensagem cifrada **c**. Aqui temos dois conceitos muito importantes quando falamos de criptografia:

- *Texto em claro*: termo utilizado para fazer referência à mensagem em seu formato original (**m**), Formato este que permite que qualquer um de posse da mensagem tenha conhecimento de seu conteúdo.

- *Texto cifrado*: é referente à mensagem cifrada **c**, resultante do processo de cifragem. Deve ser decifrada para que se possa ter acesso ao seu conteúdo original.

Apesar de chamarmos *texto em* claro, a mensagem **m** pode ser qualquer massa de dados como, por exemplo: um pacote de dados de uma aplicação, código fonte ou compilado de um sistema, campos ou registros de um banco de dados, etc.

Para a realização do processo de cifragem, **m** é fornecida como entrada para **E**. Além de **m**, é necessária uma chave de cifragem **K**. O resultado dessa operação é **c**.

Ficou confuso ao ler a frase anterior? Substitua as letras pelos seus significados, conforme os parágrafos anteriores, e ficará mais simples. Por que já não fiz isso? Para que você se acostume com o significado delas, pois vamos começar a descrever essas operações como fórmulas.

A operação de cifragem é representada pela fórmula: **c = E(K,m)**. Que você deve ler como: texto cifrado (ou mensagem cifrada) é o resultado da aplicação do algoritmo de cifragem em função da chave e do texto em claro (ou mensagem).

Decifragem

Definição: é o processo reverso da cifragem, transformando uma mensagem criptografada de volta em sua forma original. (PFLEEGER, 2006)

Da mesma forma que a cifragem, é necessário um algoritmo de decifragem (**D**). O processo de decifragem é assim descrito: **c** e a chave de decifragem **K** são fornecidas como entrada para **D**, se **K** estiver correta o resultado será **m**. (faça o exercício mental de ler a frase com o significado das letras!)

A operação de decifragem é representada pela fórmula: **m = D(K,c)**. E a leitura é: texto em claro (ou mensagem) é o resultado da aplicação do algoritmo de decifragem em função da chave e do texto cifrado (ou mensagem cifrada).

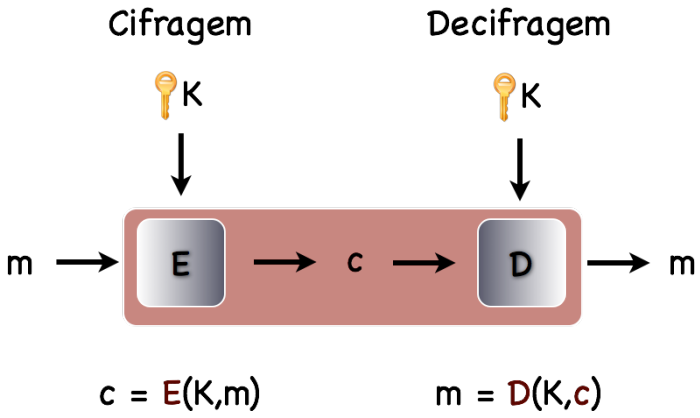
Sobre as fórmulas

Existem diversas formas de representar as operações de criptografia, praticamente cada autor usa a sua. Para nosso uso optei pela forma utilizada por Ferguson e Schneier em (FERGUSON, 2003), por eu crer que é mais fácil de ser compreendida.

Princípio de Kerckhoffs

"A segurança de um critpossistema tem que depender somente do segredo da chave e não do segredo dos algoritmos!" (FERGUSON, 2003)

O resultado prático desse princípio é que um bom algoritmo pode ser público ao invés de ter que ser mantido em segredo, pois um atacante, mesmo com conhecimento dos algoritmos **E** e **D** não consegue decifrar **c** sem o conhecimento da chave **K**. *Para garantir a segurança do criptossistema, basta que a chave **K** seja mantida em segredo.*



Na figura ao lado estão destacados em vermelho os elementos do criptossistema que o atacante conhece.

O fato de o algoritmo ser público é de grande importância. Tente imaginar como seria possível manter secretos os algoritmos em uso hoje na Internet, uma vez que os mesmos são implementados nos mais diversos programas clientes (ex: Firefox, Internet Explorer, Google Chrome e Opera) e servidores (ex.: Apache e Microsoft IIS). Não pense que, só porque um programa está compilado, que o algoritmo não é "legível" e vários softwares de código aberto também precisam implementar esses algoritmos.

Soma-se a essa dificuldade o fato de que vários algoritmos são implementados em hardware (ex.: processador criptográfico em computadores e smartcards). O que

aconteceria se os “algoritmos secretos” fossem descobertos? Todo o hardware existente teria que ser descartado?

Um bom algoritmo (entenda-se aqui como tal um algoritmo que respeite o princípio de Kerckhoffs) pode ser publicado na Internet ou em livros e pode ser implementado pelos mais diversos fornecedores de hardware e software, permitindo a interoperabilidade de seus diversos produtos.

Outra vantagem que advém da publicação dos algoritmos é a possibilidade de serem amplamente estudados pela comunidade acadêmica internacional e especialistas visando a descoberta possíveis fraquezas (criptoanálise).

Criptoanálise

Definições:

- É a ciência de recuperar o texto em claro a partir do texto cifrado sem o conhecimento da chave. (SCHNEIER, 1996)

- “Tentativas realizadas pelo criptoanalista de deduzir o significado original de uma mensagem cifrada”. (PFLEEGER, 2006)

Quebrar um texto cifrado **c** significa recuperar, a partir dele, o texto em claro **m** sem o conhecimento da chave de cifragem **K** utilizada. Em outras palavras, significa a violação do princípio de Kerckhoffs, pois o segredo da chave não foi suficiente para garantir a segurança do criptossistema.

Um ataque bastante simples (de compreender) é o chamado *ataque por força bruta*. Este consiste em simplesmente realizar tentativas de decifragem de **c** com cada uma das chaves **K** possíveis para o algoritmo. Qualquer bom algoritmo deve ser seguro contra esse tipo de ataque e essa segurança depende diretamente, porém não exclusivamente, do tamanho da chave. (Esse tema, tamanho das chaves é discutido adiante com maior profundidade.)

Para melhor compreensão, imagine um algoritmo cujo tamanho da chave é 3 bits. O número de chaves possíveis é dado por 2^3, o que resulta em apenas 8 possibilidades. Quanto tempo um atacante levaria para testar as 8 chaves possíveis? A chave, portanto, deve ser grande o suficiente para *tornar esse ataque muito caro em termos de tempo e recursos*, já que ele sempre será possível.

000	100
001	101
010	110
011	111

É importante se ter em mente que o interesse do atacante (ou criptoanalista) pode não estar restrito à descobrir uma chave e decifrar uma mensagem específica. Ele pode ter outros objetivos como: (PFLEEGER, 2006)

- deduzir a chave **K** e então decifrar futuras mensagens (e/ou mensagens antigas): sistemas que utilizam sempre a mesma chave ao longo do tempo para troca de mensagens (ou armazenamento de dados) estão sujeitos a esse tipo de ataque;

- encontrar fraquezas na implementação do algoritmo: o algoritmo define uma série de passos a serem seguidos mas, para ser executado em um computador, deve ser implementado em uma linguagem de programação por um desenvolvedor, se este cometer erros na implementação podem ser inseridas fraquezas que o algoritmo não tinha, facilitando o trabalho do atacante.

- encontrar fraquezas no ambiente onde a criptografia é utilizada: lembre-se do que foi dito no início desse roteiro: criptografia é apenas uma parte da solução. Um ótimo algoritmo criptográfico e uma boa implementação do mesmo não garantem que um determinado sistema seja seguro, a forma como a criptografia é utilizada nesse sistema pode simplesmente comprometer toda a segurança.

- inferir informações a partir do texto cifrado: algumas vezes, mesmo sem decifrar mensagem alguma, o atacante pode deduzir informações a partir de características do texto cifrado e da comunicação entre as partes. Características como tamanho e intervalo de tempo entre mensagens podem, em um determinado ambiente, permitir ao atacante identificar quando operações específicas são realizadas entre as partes.

Um exemplo elementar: A cifra de César

O imperador romano Julius César teria sido o primeiro a utilizar esse algoritmo de criptografia. Deve-se a isso o nome desta cifra. (PFLEEGER, 2006)

O algoritmo é muito simples: trata-se de um deslocamento de todas letras do texto em claro “n” posições no alfabeto. No caso do imperador esse “n” é igual a 3. Sendo assim a letra *a* seria substituída por *d*, *b* por *e* e assim por diante. Esse “n” é nossa chave de cifragem **K**.

A fórmula de cifragem desse algoritmo seria: $\mathbf{c_i = (p_i + K) \bmod 26}$. O caractere da posição *i* do alfabeto no texto em claro é igual ao caractere da posição “*(i+K) mod 26”* do alfabeto no texto cifrado. Talvez você não lembre, mas *mod* é a operação cujo resultado é o resto de uma divisão, eis alguns exemplos: 3 mod 2 = 1; 3 mod 3 = 0; 10 mod 4 = 2. Lembrou?! Qual o papel dessa operação nessa fórmula? Continue lendo...

Nota: para facilitar o entendimento, a partir deste ponto, o que for texto em claro será escrito em letras maiúsculas e em minúsculas o que for texto cifrado.

A tabela de correspondências entre letras do texto em claro para o texto cifrado para **K**=3 é:

```
Texto em claro: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Texto cifrado:  d e f g h i j k l m n o p q r s t u v w x y z a b c
```

Com **K**=3, a frase “ESTA E A CIFRA DE JULIUS CESAR”, seria assim codificada:

```
ESTA E A CIFRA DE JULIUS CESAR
hvwd h d fliud gh mxolxv fhvdu
```

Vamos entender agora o papel da operação *mod*.

Cifragem da letra *E*: a letra *E* é a quinta letra no alfabeto, assim *i* é igual a 5. Então:

$$c_5 = (5 + 3) \bmod 26 = 8 \bmod 26 = 8$$

O oitavo caractere no alfabeto (lembrando que sempre a = 1) é: *h*. Então a letra *E* no texto em claro é convertida para a letra *h* no texto cifrado. Veja acima e teste com outras letras.

Nesse caso a operação *mod* não alterou o resultado da soma. Mas vamos ver como ficaria a cifragem da letra Y. Sendo Y a penúltima letra do alfabeto, ela é a vigésima quinta letra do alfabeto, assim *i* é igual a 25. Vamos à fórmula:

$$c_{26} = (25 + 3) \bmod 26 = 28 \bmod 26 = 2$$

A segunda letra do alfabeto é *b*, assim a letra Y seria convertida para a letra *b* no texto cifrado. Faça um teste com as letras X e Z.

Para realizar a decifragem, basta que as letras do texto cifrado sejam deslocadas **K** posições no sentido contrário (uma subtração da posição). Para pensar: como seria o algoritmo de decifragem?

Criptoanálise da cifra de César

É trivial quebrar esta cifra e descobrir a chave é uma consequência inevitável. Há muitas pistas para o atacante. Uma delas são os espaços entre as palavras e, como a cifragem é letra-a-letra de forma independente, o atacante pode iniciar tentando decifrar as mais curtas.

Uma outra fraqueza que torna ainda mais fácil a quebra da cifra é que todas as ocorrências de uma letra no texto em claro são substituídas pela mesma letra no texto criptografado. Observe a letra *E* na frase do exemplo acima, todas as suas ocorrências serão cifradas para a mesma letra, *h*. O mesmo ocorre com todas as outras letras.

Apenas com essas duas pistas e alguns minutos, você deve ser capaz de quebrar o seguinte texto cifrado e descobrir a chave **K** que eu utilizei para cifrá-lo:

```
ftuvef jtup qpjt wbj tfs dpcsbep ob qspwb
```

Dificultando a criptoanálise

Para resistir à criptoanálise um bom algoritmo deve prover: (PFLEEGER, 2006)

Confusão: o atacante não deve ser capaz de prever o que vai mudar no texto cifrado (**c**) em razão da mudança de um caractere no texto em claro (**m**). Note que a cifra de César não faz isso, pois a mudança de uma letra em uma posição no texto em claro muda exatamente uma letra, na mesma posição, no texto cifrado.

Difusão: As informações do texto em claro (**m**) devem ser espalhadas por todo o texto cifrado (**c**). Mais uma vez a cifra de César não cumpre esse requisito, note que a posição

das letras é mantida. Num bom algoritmo, sem conhecimento da chave, é extremamente difícil prever as correspondências entre caracteres do texto em claro com caracteres do texto cifrado.

Bibliografia

ANDERSON, R. J. Security Engineering: A Guide to Building Dependable Distributed Systems. 2a edição. Wiley, 14 de abril de 2008. 1080 pág.

FERGUSON, N. F.; SCHNEIER, B. Practical Criptography. 1a edição. Wiley, 17 de abril de 2003. 432 pág.

MENEZES, A.; OORSCHOT P. V.; VANSTONE, S. **Handbook of Applied Cryptography**. 1a edição. CRC Press, 16 de dezembro de 1996. 780 pág.

PFLEEGER, C. P. Security in Computing. 4a edição. Prentice Hall, 23 de outubro de 2006. 880 pág.

SCHNEIER, B. Applied Cryptography: Protocols, Algorithms, and Source Code in C. 2a edição. Wiley, 18 de outubro de 1996. 758 pág.