Universidade de São Paulo/Faculdade de Educação **abril/2012**
Seminários de Ensino de Matemática (SEMA-FEUSP)

Coordenador: Nílson José Machado
Responsável: Marisa Ortegoza da Cunha
marisa.ortegoza@gmail.com

Simon Singh: O livro dos códigos[1]

Neste livro, o autor de *O último teorema de Fermat* apresenta um resumo da história da codificação de mensagens e descreve a evolução das técnicas utilizadas, desde a antiguidade até os dias atuais. Mostra que a codificação é, hoje, mais importante e estratégica do que nunca – fundamental, por exemplo, para a existência de mercado digital.

Singh cita historiadores e analistas políticos que afirmam que, se a primeira guerra mundial foi a guerra dos químicos, devido ao uso de cloro e gás mostarda; a segunda, dos físicos, pela construção da bomba atômica, a terceira, se houver, certamente será dos matemáticos, pelo controle que terão sobre a próxima grande arma de guerra: a informação.

Historicamente, a criptografia é precedida pela esteganografia, que consiste na ocultação da mensagem. Em "As Histórias", de Heródoto (480 a.C.), constam relatos de mensagens tatuadas nas cabeças raspadas de escravos para posteriormente serem escondidas pelos cabelos, mensagens escritas em tabuletas, ocultas sob camadas de cera, ou mesmo a utilização de tintas invisíveis na escrita sobre a casca de ovos cozidos. Em 1941, o FBI descobriu que agentes alemães trocavam mensagens ocultas em micropontos, que tomavam o lugar do ponto final do texto de uma carta aparentemente inocente. A esteganografia continua muito presente, hoje em dia, por exemplo, nas marcas ocultas das cédulas de dinheiro, como fator de segurança, assim como, afirmam muitos, em mensagens e imagens divulgadas pela Internet.

A esteganografia oculta a mensagem, mantendo sua forma; a criptografia altera a forma da mensagem, ocultando seu significado.

[1] Singh, Simon. *O livro dos códigos*. Rio de Janeiro: Record, 1999. (446páginas)

O autor descreve o empenho, ao longo dos tempos, de um lado, dos criptógrafos, na busca de códigos mais e mais difíceis de serem "quebrados" e, de outro, dos chamados criptoanalistas, os decodificadores, no esforço em descobrir qual a estratégia adotada em cada método. Singh chega a comparar os códigos às bactérias e os decifradores aos antibióticos: o fortalecimento de uns leva ao fortalecimento dos outros, num processo sem fim.

Uma cifra fraca pode ser pior do que nenhuma cifra, pela falsa sensação de segurança gerada. Como exemplo disso, Singh relata a condenação e execução de Maria Stuart, em 1587: a correspondência que ela trocava com seus seguidores, que conspiravam a morte de sua prima, a Rainha Elizabeth, e sua posterior ascensão ao trono da Inglaterra, foi interceptada e corretamente decifrada pelo secretário de segurança do palácio e chefe da espionagem inglesa, Sir Francis Walsingham.

Concomitantemente a deliciosos relatos de fatos históricos, o livro apresenta um vasto repertório de técnicas de codificação/decodificação. Veremos algumas delas.

Basicamente, a codificação pode ser feita por transposição ou por substituição. Na primeira, cada letra mantém sua identidade, mudando de posição; na segunda, as letras mudam de identidade, retendo a posição. Muitos métodos combinam as duas técnicas.

Cifra por transposição – a mensagem é trocada por um anagrama. Um exemplo simples desse método:

Mensagem original:	ISTO DEVE SER CODIFICADO
Passo 1:	ISTODEVESERCODIFICADO
Passo 2:	I T D V S R O I I A O S O E E E C D F C D
Passo 3:	I T D V S R O I I A O S O E E E C D F C D
Mensagem codificada:	ITDVSROIIAOSOEEECDFCD

Que tal descobrir o que está escrito, a seguir?

MSUAPRECNEITRRAASODR

Cifra de Cesar – cifra por substituição

Cada letra da mensagem é substituída por outra, deslocada uma quantidade fixa de casas, em relação ao alfabeto original. Exemplo:

Mensagem original:	ATAQUE AO AMANHECER
Mensagem cifrada:	DWDTXH DR DPDQKHFHU
(Cifra de Cesar de passo 3)	

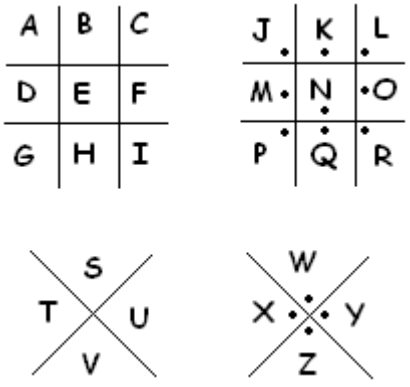
Fácil, não é? Então: NYPYZCLQ NPY TMAC !

O autor destaca o fato de que cada codificação consiste num *algoritmo* e numa *chave*, sendo que esta determina os detalhes exatos de uma particular codificação. Segundo Singh, a importância da chave, em oposição ao algoritmo, é um princípio constante da criptologia, definido em 1883, pelo linguista holandês Auguste Kerckhoff, e que leva seu nome: "A segurança de um criptossistema não deve depender da manutenção de um criptoalgoritmo em segredo. A segurança depende apenas de se manter em segredo a chave."

Em outras palavras, se a segurança da mensagem codificada depende do segredo do método usado, então o método não é seguro.

Na substituição monoalfabética, em que uma letra é substituída por outra, (da qual a cifra de Cesar é um caso particular), a chave é definida pelo alfabeto cifrado, que pode ser um rearranjo qualquer do alfabeto original. Logo, existem $26! \approx 4 \times 10^{26}$ chaves possíveis. Se cada letra é substituída por um símbolo, escolhido num determinado conjunto, a cifra também é chamada monoalfabética.

Um exemplo é a **Cifra do chiqueiro**, usada pelos maçons no século XVIII:



Uma mensagem para você:

Na tentativa de decodificar uma mensagem cifrada por substituição monoalfabética, é comum proceder-se a uma análise de frequência de letras. Pela observação de longos textos escritos na língua usada na mensagem, tabela-se a frequência de ocorrência de cada letra. Faz-se o mesmo estudo no texto encriptado e comparam-se os elementos de frequências semelhantes. A mais antiga descrição conhecida da técnica de análise de frequências data do século IX, e é atribuída ao estudioso árabe al-Kindi. Ele mostrou que era possível quebrar a cifra monoalfabética, que foi considerada invulnerável por vários séculos.

A tabela[2] abaixo refere-se à língua portuguesa.

LETRA	A	B	C	D	E	F	G	H	I	J	K	L	M
Freq%	14,63	1,04	3,88	4,99	12,57	1,02	1,30	1,28	6,18	0,40	0,02	2,78	4,74
LETRA	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Freq%	5,05	10,73	2,52	1,20	6,53	7,81	4,34	4,63	1,67	0,01	0,21	0,01	0,47

Cifra por substituição polialfabética

Diante da fraqueza apresentada pelas cifras monoalfabéticas, por volta de 1640, o italiano Leon Alberti propôs o uso de dois ou mais alfabetos, usados alternadamente. Levando essa ideia adiante, o francês Blaise de Vigenère criou a cifra que leva seu nome, e que usa 26 alfabetos cifrados distintos, para criar a mensagem cifrada. Para decifrar a mensagem, o destinatário precisa saber que alfabeto usar para cada letra da mensagem, e isso é previamente informado por uma palavra-chave. A enorme vantagem da cifra de Vigenère é que ela é imune à análise de freqüências; por esse fato, ficou conhecida, por quase dois séculos, como a "cifra indecifrável".

Mesmo sendo tão mais complexa, a cifra de Vigenère foi quebrada pelo matemático inglês Charles Babbage, por volta de 1850, que fez um estudo do padrão que a palavra-chave criava ao ser repetidamente utilizada, ao longo do texto.

[2] Fonte: www.gta.ufrj.br (*Decifrando textos em português)*

Máquinas de cifragem: a Enigma

Em 1918, o alemão Arthur Scherbius desenvolveu uma máquina criptográfica, que, segundo o autor, tornar-se-ia o mais terrível sistema de cifragem da história: a Enigma. Sua forma básica consistia de um teclado para a entrada de cada letra do texto original, uma unidade misturadora, que cifrava cada letra e um mostrador, que iluminava a letra cifrada.

Em dado momento, seu funcionamento já era plenamente conhecido pelos ingleses, que até dispunham de algumas unidades, mas qual configuração – dentre as cerca de 10^{16} possíveis – teria sido usada para cifrar determinada mensagem? Conhecia-se o algoritmo, mas qual a chave?

O governo britânico reuniu um grupo de especialistas (em lingüística, matemática, xadrez, entre outros) para descobrir como quebrar o código Enigma. Credita-se a Alan Turing o maior mérito nesse feito. Mas os decodificadores que trabalhavam para o serviço de inteligência britânico faziam juramento de sigilo, e muitos morreram antes de terem o devido reconhecimento pelo seu trabalho. Foi o caso de Turing. Mesmo depois da guerra, o governo manteve o segredo, para que a Enigma continuasse a ser usada nas colônias britânicas, e o poder devidamente exercido.

Criptografia com chave pública (RSA)

A grande fraqueza de um código reside no fato de a chave ter de ser transmitida ao destinatário da mensagem. Como fazer isso, sem sofrer riscos de que ela seja descoberta?

Dois grupos distintos de matemáticos se debruçaram sobre o problema e chegaram à mesma conclusão: tudo estaria resolvido se a chave pudesse ser pública, livremente transmitida.

A ideia é que a mensagem fosse cifrada por uma função matemática de "mão única", isto é, de fácil aplicação, mas de difícil reversão. Haveria duas chaves: uma pública, para que todos pudessem codificar e enviar a mensagem para uma determinada pessoa, e uma privada, conhecida apenas por essa pessoa, destinatária da mensagem, para sua decodificação.

O nome deste criptossistema - RSA - é uma homenagem a seus inventores: Ronald Rivest, Adi Shamir e Leonard Adleman, pesquisadores do MIT (Massachusetts Institute of Technology). A dificuldade de se quebrar o sistema RSA deriva da dificuldade em se fatorar grandes números – atualmente, com cerca

de 150 algarismos. Trata-se de um sistema cuja utilização exige o uso de computadores.

Inicialmente devem ser escolhidos dois números primos quaisquer p e q. Quanto maior o número escolhido, mais seguro será o algoritmo. O produto n = pq faz parte da chave pública de codificação. Os primos p e q, da chave privada. Logo, a segurança do sistema está na dificuldade em fatorar n.

Computadores digitais processam apenas números (escritos na forma binária). Por isso, cada mensagem a ser codificada é transformada num número, antes de cifrada.

Vejamos um exemplo de aplicação do RSA:

Bob quer enviar a mensagem M = "88" para Alice.

Codificação:

Alice escolhe os primos p = 11 e q = 17.

Alice calcula n = pq = 187 (esta informação será pública)

Alice calcula a função de Euler de n: $\varphi(77) = 10 \times 16 = 160$. [3]

Alice precisa de um número *e*, primo com $\varphi(n)$, assim, *e* será inversível módulo 160. Por exemplo, seja $e = 7$.

O par (n , *e*) é a chave pública de codificação RSA.

Alice envia (187, 7) para Bob.

Bob codifica a mensagem; a mensagem codificada, C(M), será o resto da divisão de M^e, por n, ou seja:

$C(M) \equiv 88^7 (\text{mod } 187) = 88^2 \times 88^2 \times 88^2 \times 88 (\text{mod } 187) =$

$= (77 \times 77 \times 77 \times 88) (\text{mod } 187) = 40.174.904 (\text{mod } 187) = 11$

Bob envia 11 para Alice.

Decodificação:

Alice recebe a mensagem cifrada C(M) = 11.

[3] A função de Euler de um número inteiro n é a quantidade de números inteiros entre 1 e n-1 primos com n. Se p é primo, então $\varphi(p) = p-1$; se n = pq, com p e q primos, então $\varphi(n) = (p-1)(q-1)$.

Alice determina o elemento inverso de *e*, módulo $\varphi(n)$, ou seja, o elemento d tal que $ed \equiv 1 \pmod{\varphi(n)}$. No caso, $7d \equiv 1 \pmod{160}$.

Para encontrar d, Alice pode ir por tentativas, ou usar o algoritmo de Euclides. Alice encontra d = 23.

O par (n , d) é a chave privada de decodificação RSA.

A mensagem original M será o resto da divisão de $[C(M)]^d$ por n, ou seja:

$D(C(M)) = 11^{23} \pmod{187} = 11^9 \times 11^9 \times 11^5 (187) = 11^3 \times 11^3 \times 11^3 \times 11^9 \times 11^5 (187) =$

$= (22 \times 22 \times 22) \times 11^9 \times 11^5 (187) = 176 \times 176 \times 11^5 (187) = 121 \times 44 (187) =$

$= 5324(187) = 88 = M$, que é a mensagem original.

Por que o método RSA funciona?

Seja M o número inteiro que representa a mensagem a ser codificada, com $1 \leq M < n = pq$. Seja C(M) a mensagem M codificada. Chamando de D a função decodificadora, é preciso que

$$D(C(M)) = M$$

Temos:

$D(C(M)) \equiv (M^e)^d \pmod{n} \equiv M^{ed} \pmod{n}$

Como d é o inverso de e, módulo $\varphi(n)$, temos $ed \equiv 1 \pmod{\varphi(n)}$. Daí, existe inteiro k tal que

$ed = 1 + k.\varphi(n) = 1 + k(p-1)(q-1)$

Como $n = pq$, $p \neq q$, para calcular $M^{ed} \pmod{n}$, podemos calcular $M^{ed} \pmod{p}$ e $M^{ed} \pmod{q}$.

$M^{ed} \pmod{p} \equiv M^{1 + k(p-1)(q-1)} \pmod{n} \equiv M(M^{p-1})^{k(q-1)} \pmod{p}$

Agora: ou p divide M ou p não divide M.

Se p divide M, então $M \equiv 0 \pmod{p}$, ou seja, $M^{ed} \pmod{p} \equiv M \pmod{p}$, para todo M inteiro.

Se p não divide M, pelo Pequeno Teorema de Fermat[4], temos

$M^{p-1} \equiv 1 \pmod{p}$, ou seja, $M^{ed} \pmod{p} \equiv M.1 \pmod{p} \equiv M \pmod{p}$.

[4] Pequeno Teorema de Fermat:
Dados inteiros a e p, com p primo, se p não divide a, então $a^{p-1} \equiv 1 \pmod{p}$.

Analogamente, podemos chegar a $M^{ed} \pmod q \equiv M \pmod q$.

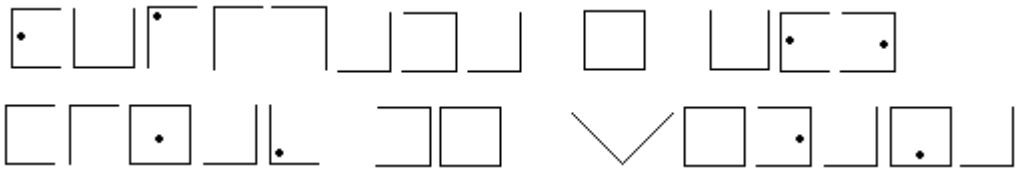
Logo, p divide $(M^{ed} - M)$ e q divide $(M^{ed} - M)$.

Como p e q são primos distintos, concluímos que pq = n divide $(M^{ed} - M)$.

Isto é: $M^{ed} \equiv M \pmod n$, para todo M inteiro.

Como M^{ed} e M estão no intervalo de 1 a n-1, podemos concluir que $M^{ed} = M$, o que prova a corretude do RSA.

***** *****



***** *****