V Bienal da SBM
Sociedade Brasileira de Matemática
UFPB - Universidade Federal da Paraíba
18 a 22 de outubro de 2010

Criptografia: códigos sem segredos

Gabriel Costa Borba de Lira *

1 INTRODUÇÃO

O termo Criptografia surgiu da fusão das palavras gregas "Kryptós"e "gráphein", que significam "oculto"e "escrever", respectivamente. Vai muito além dos filmes de 007 e outros agentes igualmente secretos. Trata-se de um conjunto de conceitos e técnicas que visa codificar uma informação de forma que somente o emissor e o receptor possam acessá-la, evitando que um intruso consiga interpretá-la. A interpretação desta informação depende crucialmente da matemática.

Na era do conhecimento e da velocidade na troca de informações, as empresas têm se valido da tecnologia criptográfica para proteger os dados de suas redes corporativas no ambiente cibernético, por onde circulam bilhões de informações por segundo, garantindo, assim, a segurança no tráfego de suas informações.

Essa segurança é fundamental já que não há formas de se saber os caminhos por onde as informações trafegarão, nem o quão seguro será este caminho (principalmente em se tratando de Internet). A utilização de criptografia em comunicações não é algo novo, e já era utilizada antes mesmo de existirem computadores.

2 CÓDIGOS FAMOSOS

2.1 CÓDIGO DE CÉSAR

Um dos códigos mais simples consiste em substituir uma letra do alfabeto pela seguinte. Por exemplo, a mensagem AMO A BIENAL seria codificada como

BNPBCJFOBM.

Foi usado pelo imperador romano Júlio César para se comunicar com suas tropas em combate. Se, ao invés da letra seguinte usássemos a quarta letra anterior, teríamos o seguinte:

UHJUVEAIUG.

É um código fácil de ser "quebrado"já que temos uma noção da freqüência com que as letras aparecem na língua portuguesa. No primeiro exemplo, quanto mais B´s aparecem, mais certeza temos de que B corresponde a A. Quanto maior o texto, mais fácil de perceber a frequência. Vejamos a Tabela 1:

Letra	%	Letra	%	Letra	%	Letra	%
A	14,64	G	1,30	N	5,05	T	4,34
B	1,04	H	1,28	O	10,73	U	4,64
C	3,88	I	6,18	P	2,52	V	1,70
D	4,10	J	0,40	Q	1,20	X	0,21
E	12,57	L	2,78	R	6,53	Z	0,47
F	1,02	M	4,75	S	7,81		

*Universidade Federal de Pernambuco, UFPE, PE, Brasil, gabrielcblira@hotmail.com

Tabela: Frequência das letras no português

Exercício1: Use o método da contagem de freqüência para decifrar a mensagem abaixo.

GENCTQSWGSWGDTCTWOCOGPUCIGORQTEQPVCIGOFGHTGSWGPEKCGCKPFCOCKUUKORNGUU
GVGOQUWOEQORWVCFQTUGCNKPIWCGEQPJGEKFCCOCKQTRCTVGFQRTQEGUUQRQFGUGTCWVQ
OCVKBCFQQSWGVQTPCKPXKCXGKUVQFQUQUEQFKIQUSWGGPXQNXGOUWDUVKVWKECQFGNGVT
CUEQOQQSWGGUVCOQUWVKNKBCPFQRCTCEQFKHKECTGUVCOGPUCIGO

2.2 CÓDIGOS EM BLOCOS

Outra maneira de criptografar mensagens é retirar os espaços e separar em blocos de n letras e depois permutar os blocos seguindo alguma regra. Quanto mais complexa for essa regra, mais difícil será decifrar a mensagem. É um código mais complexo do que o de César, pois existe uma infinidade de regras para permutar os blocos. Já os de César são apenas 26 opções de substituição de letras se considerarmos o alfabeto com as letras K, W e Y.

Vamos codificar AMO A BIENAL por blocos:

Primeiro retiramos os espaços:

AMOABIENAL

Separamos em blocos de n letras. Neste caso vamos usar $n = 2$:

AM-OA-BI-EN-AL

Permutamos as letras de cada bloco para dificultar:

MA-AO-IB-NE-LA

Agora permutamos os blocos seguindo uma regra. Neste caso a regra é trocar o primeiro com o último, o terceiro com o antepenúltimo, e assim por diante:

LA-AO-IB-NE-MA

Nossa mensagem codificada fica:

LAAOIBNEMA

Exercício2: A mensagem a seguir foi codificada pelo mesmo método do exemplo acima. Descubra a mensagem.

SIOCCROGOCBMIFOCAPMOVIOHMIDSOTEUMOCEADRAASAMNISEASDOIUQORPERTILESQASAROL
SNESEMIDAISO

Exercício3: Por que a contagem de freqüência não funciona quando usamos códigos em blocos?

3 CÓDIGOS DE CHAVE PÚBLICA

Se soubermos qual código foi usado para criptografar uma mensagem, então podemos facilmente decodificá-la fazendo o processo inverso. Certo? Errado. De modo geral a afirmação não está errada, mas a palavra "facilmente"é um equívoco. Vamos imaginar um processo que é fácil de fazer, mas difícil de desfazer.

Você já viu uma dessas armadilhas usadas para pescar lagostas? Elas consistem de uma gaiola com uma porta fechada atrás e uma entrada para a lagosta na frente. O segredo está na entrada, que tem a forma de um funil: larga na parte externa e cada vez menor à medida que a lagosta vai entrando na gaiola. A lagosta fica presa na gaiola porque, para poder sair, teria que encontrar e passar pela parte estreita do funil, que é um problema complicado demais para uma lagosta, cujo cérebro tem o tamanho aproximado de uma ervilha.

Mas que processo matemático é fácil de fazer e difícil de desfazer para que possamos utilizá-lo em criptografia? Tais códigos são conhecidos como de chave pública, já que o processo de codificação pode ser conhecido de qualquer um sem comprometer a segurança do código.

4 PRAZER, RSA!

4.1 HISTÓRIA

O mais conhecido dos métodos de criptografia de chave pública é o RSA. Este código foi inventado em 1977 por R. L. Rivest, A. Shamir e L. Adleman, que na época trabalhavam no Massachussets Institute of Technology (M.I.T.), uma das melhores universidades americanas. As letras RSA correspondem às iniciais dos inventores do código. Há vários outros códigos de chave pública, mas o RSA continua sendo o mais usado em aplicações comercias.

4.2 O MÉTODO RSA

Vamos entender como um problema é fácil de fazer, mas difícil de desfazer. Para começar, precisaremos de dois números primos p e k para multiplicá-los e obter um inteiro n. Codificamos nossa mensagem com n e decodificamos com p e k. n pode ser tornado público, mas p e k precisam ser mantidos em segredo.

Quebrar o código RSA consiste em fatorar n. Mas aí você pergunta: Só isso? Sim, mas n é a multiplicação de dois primos muito grandes, cada um com mais de 100 algarismos, logo n possui mais de 200 algarismos!

Fatorar este número mesmo em um supercomputador levaria muito mais tempo do que a idade do universo. E eu não estou brincando.

Exercício4: Fatore o número 4897. Este número tem apenas 4 algarismos.

5 CONSTRUINDO AS FERRAMENTAS PARA ENTENDER O PROCESSO RSA

As ferramentas aqui apresentadas são de suma importância para a compreensão completa do funcionamento do método RSA. Devido à complexidade de algumas delas, não demonstraremos nenhuma, até por que esse não é o objetivo principal do curso. Demonstrações dos teoremas podem ser encontradas em qualquer curso de álgebra.

5.1 O TEOREMA DA FATORAÇÃO ÚNICA

Todo número inteiro n não-nulo pode ser escrito de maneira única na forma

$$n = up_1^{a_1}p_2^{a_2}\cdots p_k^{a_k},$$

onde $u = \pm 1$, $p_1, p_2, \ldots, p_k$ são números primos tais que $p_1 < p_2 < \cdots < p_k$ e $a_1, a_2, \ldots, a_k$ são números naturais.

Exercício5: Determine se existem inteiros positivos x e y que satisfaçam a equação $30^x \cdot 35^y = 21^x \cdot 140 \cdot 5^{2x}$.

Exercício6: Determine se existem inteiros positivos x, y e z que satisfaçam a equação $2^x \cdot 3^4 \cdot 26^y = 39^z$.

5.2 CONGRUÊNCIA MODULAR

Antes de definir congruência modular, vamos à idéia básica.

Se eu faço a divisão de qualquer número inteiro por 4, quais são os possíveis restos dessa divisão? Vamos olhar a tabela abaixo:

INTEIRO	4	5	6	7	8	9	10	11	$\cdots$	$4k$	$4k+1$	$4k+2$	$4k+3$
RESTO	0	1	2	3	0	1	2	3	$\cdots$	0	1	2	3

Tabela 2: Tabela dos restos de 4

O que aconteceu quando eu dividi 5 por 4? E 9 por 4? Que relação 5 e 9 tem? Podemos dizer que estes dois números são congruentes em relação a 4. Ou seja, ambos deixam resto 1 quando divididos por 4.

Formalmente, temos:

Seja n um natural maior que 1. Diremos que dois inteiros a e b são congruentes módulo n, e denotamos por

$$a \equiv b \ (mod \ n)$$

se e somente se n divide $a - b$.

Propriedades:

reflexiva: $a \equiv a \ (mod \ n)$

simétrica: se $a \equiv b \ (mod \ n)$, então $b \equiv a \ (mod \ n)$

transitiva: se $a \equiv b \ (mod \ n)$ e $b \equiv c \ (mod \ n)$, então $a \equiv c \ (mod \ n)$

Exercício7: Calcule:

a) $57 \equiv a \ (mod \ 7)$ b) $1080 \equiv a \ (mod \ 24)$ c) $643 \equiv a \ (mod \ 11)$

5.3 RESÍDUOS

Vamos usar congruência modular para entender o que vem a ser resíduos.

Digamos que a é um inteiro positivo. Pelo algoritmo da divisão em $\mathbb{Z}$, temos

$$a = n \cdot q + r \text{ e } 0 \leq r < n.$$

Assim,

$$a - r = n \cdot q$$

Que equivale a dizer

$$a \equiv r \ (mod \ n)$$

Com isto verificamos que todo inteiro positivo é congruente módulo n ao resto de sua divisão por n, que é um número entre 0 e n.

5.4 INVERSOS MODULARES

Dois números inteiros positivos a e a' são inversos módulo n se

$$a \cdot a' \equiv 1 \ (mod \ n).$$

Neste caso, também dizemos que a' é o inverso de a módulo n e vice-versa.

Exercício8: Calcule:

a) $2a' \equiv 1 \ (mod \ 7)$ b) $7a' \equiv 1 \ (mod \ 8)$ c) $5a' \equiv 1 \ (mod \ 11)$

5.5 TEOREMA CHINÊS DO RESTO

Sejam m e n inteiros positivos primos entre si. Se a e b são inteiros quaisquer, então o sistema

$$\begin{cases} x \equiv a \ (mod \ m) \\ x \equiv b \ (mod \ n) \end{cases}$$

sempre tem solução e qualquer uma de suas soluções pode ser escrita na forma

$$a + m \cdot (m' \cdot (b - a) + n \cdot t)$$

onde t é um inteiro qualquer e m' é o inverso de m módulo n.

5.6 POTÊNCIAS

Vamos aprender a determinar os restos de uma potência quando dividida por n.

Por exemplo, qual é o resto da divisão de 10^{135} por 7? Usando congruência modular, queremos determinar X de forma que

$$X \equiv 10^{135} \ (mod \ 7).$$

Temos que observar que $10^6 \equiv 1 \ (mod \ 7)$. Dividindo 135 por 6, obtemos as seguintes congruências módulo 7:

$$10^{135} \equiv (10^6)^{22} \cdot 10^3 \equiv 1^{22} \cdot 10^3 \equiv 6 \ (mod \ 7)$$

Assim, o resto da divisão de 10^{135} por 7 é 6.

Para um caso geral precisamos fazer a divisão de a^p por n. Para isso temos que descobrir um $q < p$ que satisfaça $a^q \equiv 1 \ (mod \ n)$. Depois façamos a divisão de p por q e obtemos $p = q \cdot k + r$. Fazendo a congruência, temos:

$$X \equiv (a^q)^k \cdot a^r \ (mod \ n).$$

Note que $(a^q)^k$ é sempre 1, o que nos remete a calcular apenas $X \equiv a^r \ (mod \ n)$.

Exercício10: Calcule o resto da divisão de 2^{124512} por 31.

Exercício11: Calcule o resto da divisão de 10^{65} por 7.

Exercício12: Calcule o resto da divisão de 3^{78} por 7.

5.7 PEQUENO TEOREMA DE FERMAT

Se p é um primo e a é um inteiro que não é divisível por p, então

$$a^{p-1} \equiv 1 \ (mod\ p).$$

Exercício13: Calcule o resto da divisão de 3^{64} por 31.

Exercício14: Calcule o resto da divisão de 3^{98745} por 43.

5.8 TEOREMA DE FERMAT

Se p é um primo e a é um inteiro, então

$$a^{p} \equiv a \ (mod\ p).$$

Exercício15: Calcule o resto da divisão de 3^{34} por 17.

Exercício16: Calcule o resto da divisão de 2^{533} por 11.

6 CRIPTOGRAFIA RSA

Agora é a hora de usar as nossas ferramentas para entender o método criptográfico mais usado no mundo para transações comerciais.

6.1 PRÉ-CODIFICAÇÃO

Na pré-codificação convertemos as letras em números usando a seguinte tabela de conversão:

A	B	C	D	E	F	G	H	I	J	K	L	M
10	11	12	13	14	15	16	17	18	19	20	21	22

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
23	24	25	26	27	28	29	30	31	32	33	34	35

Tabela 3: Tabela de conversão para a pré-codificação

O espaço entre duas palavras será substituído pelo número 99. Por exemplo, a frase AMO A BIENAL é convertida da seguinte maneira:

$$10222499101118142310 21$$

Agora precisamos determinar dois primos distintos que tenham resto 5 na divisão por 6 (esta condição não é essencial, mas para o nosso exemplo tornará nossos cálculos menos complicados). Vamos escolher $p = 17$, $q = 23$ e $n = 17 \cdot 23 = 391$.

Na última parte da pré-codificação vamos separar o número que encontramos ao converter as letras em blocos. Os blocos devem ser menores do que n e não podem começar por 0. No nosso exemplo os blocos devem ser números menores do que 391.

$$102 - 224 - 99 - 101 - 118 - 142 - 310 - 21$$

6.2 CODIFICAÇÃO

Bom. Agora nós temos o n que falávamos tanto e vários blocos de números. Vamos denotar os blocos por b e o bloco codificado por $C(b)$. A receita para calculá-lo é:

$$C(b) = \text{resto da divisão de } b^3 \text{ por } n.$$

Aplicando isso ao nosso primeiro bloco do exemplo, temos:

$$102^3 \equiv 102^2 \cdot 102 \equiv 238 \cdot 102 \equiv 24276 \equiv 34(\text{ mod } 391)$$

Donde concluímos que $C(102) = 34$.

Vamos fazer o mesmo com os outros blocos:

$$\begin{aligned}
224^3 &\equiv 224^2 \cdot 224 \equiv 128 \cdot 224 \equiv 129 \pmod{391} \\
99^3 &\equiv 99^2 \cdot 99 \equiv 26 \cdot 99 \equiv 228 \pmod{391} \\
101^3 &\equiv 101^2 \cdot 101 \equiv 35 \cdot 99 \equiv 337 \pmod{391} \\
118^3 &\equiv 118^2 \cdot 118 \equiv 239 \cdot 118 \equiv 50 \pmod{391} \\
142^3 &\equiv 142^2 \cdot 142 \equiv 223 \cdot 142 \equiv 386 \pmod{391} \\
310^3 &\equiv 310^2 \cdot 310 \equiv 305 \cdot 310 \equiv 319 \pmod{391} \\
21^3 &\equiv 21^2 \cdot 21 \equiv 50 \cdot 21 \equiv 268 \pmod{391}
\end{aligned}$$

Reunindo todos os blocos, descobrimos que a mensagem codificada é:

$$34 - 129 - 228 - 337 - 50 - 386 - 319 - 268$$

É importante salientar que os blocos não devem ser juntados afim de se formar um único número, pois a decodificação seria impossível.

6.3 DECODIFICAÇÃO

A informação que precisamos para poder decodificar consiste de dois números: n e o inverso $d > 0$ de 3 módulo $(p-1)(q-1)$. Pela definição de inverso isto significa que devemos ter

$$3d \equiv 1 \pmod{(p-1)(q-1)}.$$

Chamaremos o par (n, d) de chave de decodificação. Esta chave precisa ser mantida em segredo. Quem a descobrir poderá decodificar qualquer mensagem endereçada a você. Se o bloco codificado é $C(b)$, denotaremos por $D(C(b))$ o bloco descodificado, ou seja, b. Para calculá-lo, faremos o seguinte esquema:

$$D(C(b)) = \text{resto da divisão de } C(b)d \text{ por } n.$$

Como estamos supondo que p e q deixam resto 5 na divisão por 6, temos que

$$p \equiv 5 \pmod 6 \text{ e } q \equiv 5 \pmod 6.$$

Assim,

$$(p-1)(q-1) \equiv 4 \cdot 4 \equiv 16 \equiv 4 \equiv -2 \pmod 6,$$

donde

$$(p-1)(q-1) = 6k - 2,$$

para alguma inteiro positivo k. Contudo, o inverso de 3 módulo $6k-2$ é igual a $4k-1$. Logo, podemos tomar

$$d = 4k - 1$$

No nosso exemplo, estamos considerando $p = 17$ e $q = 23$, de forma que

$$(p-1)(q-1) = 16 \cdot 22 = 352 = 6 \cdot 58 + 4$$

que é igual a

$$(p-1)(q-1) = 6 \cdot 59 - 2.$$

Portanto, neste caso, $k = 59$ e $d = 4 \cdot 59 - 1 = 235$.
Vamos provar que isso é verdade aplicando a receita ao bloco 34 para que ele volte a ser 102.

$$D(C(102)) \equiv 34^{235} \pmod{391}$$

Fazer esta conta em um computador é impossível, mas aplicando o Pequeno Teorema de Fermat e depois o Teorema Chinês do Resto, calculamos 34^{235} módulo 17 e módulo 23, que são os primos em que 391 se fatora. É por isso que devemos deixar em segredo estes dois primos, pois com 391 apenas é impossível continuar esta conta. Para começar:

$$34 \equiv 0 \ (mod 17)$$

$$34 \equiv 11 \ (mod 23)$$

Assim,

$$34^{235} \equiv 0^{235} \equiv 0 \ (mod 17)$$

Aplicando o Pequeno Teorema de Fermat à outra congruência, temos

$$11^{235} \equiv (11?^{2}2)^{10} \cdot 11^{15} \equiv 11^{15} \ (mod 23)$$

Mas

$$11 \equiv -12 \equiv -4 \cdot 3 \ (mod 23)$$

de forma que

$$11^{235} \equiv 11^{1}5 \equiv -4^{1}5 \cdot 3^{15} \ (mod 23)$$

Contudo,

$$4^{11} \equiv 1 \ (mod 23)$$

$$3^{11} \equiv 1 \ (mod 23)$$

de modo que

$$4^{15} \equiv 2^{30} \equiv (2^{11})^{2} \cdot 2^{8} \equiv 2^{8} \equiv 3 \ (mod 23)$$

$$3^{15} \equiv 3^{11} \cdot 3^4 \equiv 3^4 \equiv 12 \ (mod 23).$$

Donde podemos concluir

$$11^{235} \equiv -4^{15} \cdot 3^{15} \equiv -3 \cdot 12 \equiv 10 \ (mod 23).$$

Portanto,

$$34^{235} \equiv 0 \ (mod 17)$$

$$34^{235} \equiv 10 \ (mod 23).$$

Isto corresponde ao sistema

$$\begin{cases} x \equiv 0 \ (mod \ 17) \\ x \equiv 10 \ (mod \ 23) \end{cases}$$

que podemos resolver utilizando o Teorema Chinês do Resto. Da segunda congruência, obtemos

$$x = 10 + 23y$$

que, ao ser substituído na primeira congruência, nos dá

$$10 + 23y \equiv 0 \ (mod \ 17).$$

Assim,

$$6y \equiv 7 \ (mod \ 17).$$

Mas, 6 tem inverso 3 módulo 17, de forma que

$$y \equiv 3 \cdot 7 \equiv 4 \ (mod \ 17).$$

Portanto,

$$x = 10 + 23y = 10 + 23 \cdot 4 = 102.$$

Como seria de se esperar. Afinal, estamos decodificando 34 que é a codificação de 102.

Exercício17: Decodifique os demais blocos da mensagem

$$34 - 129 - 228 - 337 - 50 - 386 - 319 - 268$$

usando o procedimento acima.

Referências

[1] _____ - *Wikipédia.*, disponível em: http://pt.wikipedia.org/wiki/Criptografia.

[2] Coutinho, S. C. - *Programa de Iniciação Científica , Criptografia*, OBMEP 2007.

[3] Simon S. - *O Livro dos Códigos*, Editora Record, 2010.