De: Routo Terada (web-site http://www.ime.usp.br/~rt)
    Prof Titular do Depto. de Ciência da Computação da Univ. de S. Paulo

Para: Ciência Hoje na Escola
Data: maio de 1999.
Versão: 2

Título: Uma carta cifrada pela Internet? Como decifrar?

Vou falar sobre um assunto muito comum e atual na Internet. Vamos supor que um rapaz chamado Beto quer enviar uma carta de amor para uma garota chamada Alice, mas pela linha telefônica ligada ao seu computador pessoal. Mas há um problema: Beto teme que um outro rapaz, Carlos, seja capaz de "capturar" a linha e consiga ler a sua carta, ainda mais porque Carlos está interessada na Alice também!

Beto, muito esperto, já havia lido um artigo como esse aqui, e sabe de um truque chamado "ciframento" de um texto, que é muito usado na Internet, e pode solucionar o seu problema com Carlos. Vamos ver aqui um tipo de ciframento que é fácil de usar.

Ciframento consiste em tornar a carta ilegível, sem sentido para Carlos, ANTES de Beto transmití-la pela linha para Alice. Por exemplo, Beto pode trocar cada letra "a" por "g", e então uma frase na carta como:
   "A minha felicidade fica infinita quando recebo sua carta"
fica assim:
   "G minhg felicidgde ficg infinitg qugndo recebo sug cgrtg"

Algumas palavras como "cgrtg" ficam impronunciáveis e ilegíveis, mas outras como "recebo" ficam intatas. Aí então Beto decide complicar mais a vida de Carlos e troca também as outras três vogais por outras letras da seguinte forma:
  "e" por "w", "i" por "x", "o" por "z", "u" por "k"

E então aquela frase fica assim:
  "G mxnhg fwlxcxdgdw fxcg xnfxnxtg qkgndz rwcwbz skg cgrtg"

Que tal? Ficou bastante ilegível, não? Agora você vê que todas as palavras são ilegíveis para Carlos. E Alice? O que ela recebe também é ilegível? A resposta é: sim, também é, a não ser que ANTES de receber ela tivesse combinado com Beto quais as trocas de letras Beto usaria SEM Carlos ficar sabendo. Alice consegue então "destrocar" as letras e recuperar a frase original, legível. Por exemplo, Alice recupera "felicidade" após ter

recebido "fwlxcxdgdw".

Esse tipo de troca e destroca de letras é chamado CIFRAMENTO ou codificação, e a correspondência das letras trocadas que é previamente combinada entre Beto e Alice é chamada CHAVE de ciframento. A frase que Beto envia para Alice chama-se TEXTO CIFRADO. O ato de trocar as letras que Beto faz chama-se CIFRAR, e o ato de destrocar que Alice faz, DECIFRAR.

Entretanto, como Carlos é um rapaz persistente e está "louco" para decifrar mesmo sem conhecer a chave, ele vai estudar matemática e língua portuguêsa, e depois de muito tempo de pesquisa ele descobre um "calcanhar de Aquiles", um ponto fraco, no ciframento que talvez Beto não conheça. Carlos fica sabendo que alguém estudioso contou todas as ocorrências de cada letra em livros e mais livros armazenados em computador, e concluiu que cada letra aparece nesses livros com uma frequência mais ou menos FIXA. Por exemplo, ele sabe que as vogais aparecem com mais frequência (lembra do Jogo da Forca em que você tem que adivinhar uma palavra secreta tentando adivinhar as suas letras? Você começa com as vogais!). A vogal "a" é a mais frequente, aparece 13.5% das vezes, seguida pela "e" com 12.5%, "i" com 6.0%, "o" com 5.5%, e "u" com 4.5%.

E como Carlos poderia usar esse conhecimento sobre as letras em português para decifrar? Ele raciocina da seguinte forma: se uma letra, digamos "y", substitui a letra "x" original (que Carlos não conhece), então a letra "y" aparece no texto cifrado um número de vezes IGUAL à letra "x" no texto original. Então basta contar quantas vezes "y" aparece no texto cifrado e a porcentagem de vezes que "y" ocorre deve ser PRÓXIMA da porcentagem da letra "x" em português. E aí Carlos fica sabendo que "y" está substituindo "x", ou seja, ele consegue calcular parte da chave. Se ele repetir esse raciocínio para cada letra que ocorre no texto cifrado, ele acaba calculando a chave toda! Um leitor que conhece um pouco de matemática pode observar que quanto mais texto cifrado Carlos consiga obter, mais preciso vai ser esse cálculo, pois mais próxima será a porcentagem entre "y" e a letra "x" correta.

Carlos, ansioso em aplicar o novo conhecimento, conta quantas vezes cada letra aparece no texto cifrado que havia capturado, e obtém a seguinte contagem, em ordem decrescente:
"g", 9 vezes, "x", 7 vezes, "w", 4 vezes, "c", 4 vezes, "n", 4 vezes, "d", 3 vezes, "f" 3 vezes, "z", 2 vezes, "k", 2 vezes, "r", 2 vezes, "t", 2 vezes, e 1 vez as letras "b", "h", "q", "l", "m", e "s".

Os campeões são "g", "x" e "w", que provavelmente estão substituindo vogais. Carlos agora pode tentar decifrar, por exemplo, destrocando "g", "x", "w", por "a" ou "e" ou "i" ou "o", pois estas são as vogais mais frequentes e

ver se o resultado é algo legível. Se Carlos tivesse muito mais texto cifrado, ele conseguiria com precisão deduzir que "g" deve ser destrocada por "a", "x" por "i", e "w" por "e".

Após destrocar "g" por "a" resulta:
"A mxnha fwlxcxdadw fxca xnfxnxta qkandz rwcwbz ska carta".

Não ajuda muito, ainda é muito ilegível. Após destrocar "x" por "i", resulta: "A minha fwlicidadw fica infinita qkandz rwcwbz ska carta", e após "w" por "e": "A minha felicidade fica infinita qkandz recebz ska carta".

Nesta última já se percebe algumas palavras totalmente legíveis. Mas outras ainda são parcialmente ilegíveis. Mas já dá para deduzir que "k" deve ser destrocada por "u", e "z" por "o"

Por outro lado, se Beto e Alice tiverem o cuidado de trocarem a chave a cada sete dias, por exemplo, todos os textos cifrados que Carlos tiver capturado, e todos os seus cálculos de frequência feitas durante uma semana não vão mais valer na semana seguinte. Por quê? A resposta não é tão difícil assim, pense um pouco.

Há um detalhe que deixei de lado até agora que é o seguinte: se por acaso a carta de Beto contiver uma palavra com a letra "g" como por exemplo "gostar", o texto cifrado conteria "gzstgr". Alice então ficaria confusa pois ao destrocar as letras de acordo com a chave ela obteria "aostar", ou seja, Alice não está sabendo que o primeiro "g" NÃO deve ser destrocado por "a". A mesma confusão ocorreria se a carta contiver qualquer palavra com "w", "x", "z", ou "k" pois elas ocorrem na chave. Uma solução para evitar tal confusão é trocar as consoantes na chave por vogais. Por exemplo trocar toda letra "g" por "e", "w" por "a", "z" por "i", "k" por "u", e "x" por "o". E então o texto cifrado de "gostar" seria "ezstgr" e Alice destrocaria corretamente "e" por "g".

O leitor atento pode perceber que mesmo com essa nova chave Carlos consegue decifrar através da contagem das letras no texto cifrado, apesar de ser mais difícil. Veja o caso mais geral descrito a seguir.

A chave mais geral para este ciframento é a de trocar qualquer letra, vogal ou consoante, por outra letra. Como são no total 26 letras, Alice e Beto combinam uma chave de 26 trocas. Por exemplo, a chave pode ser estendida para trocar cada consoante à esquerda do sinal = pela letra à direita como a seguir:
"b"="d" "c="m" "d"="r" "f"= "j" "g"="u" "h"="s" "j"="h" "k"="c"
"l"="q" "m"="v" "n"="f" "p"="o" "q"="b" "r"="p" "s"="e" "t"="y"
"v"="a" "w"="i" "x"="t" "y"="l" "z"="n". Só para lembrar, as vogais

são: "a"= "g" "e"= "w" "i"= "x" "o"= "z" "u"= "k". E então aquela primeira frase fica cifrada e mais ininteligível: "G vxfsg jwqxmxrgrw jxmg xfjxfxyg bkgfrz pwmwdz ekg mgpyg"

Portanto se por acaso Carlos, o intruso, capturar o texto cifrado envolvendo a troca das 26 letras, ele fica bem confuso e torna-se mais difícil decifrar SEM conhecer a chave usada por Beto e Alice.

Entretanto, Carlos pode estudar um pouco mais para saber a ocorrência usual das consoantes na língua portuguesa, além das vogais que ele já sabia. Entre as consoantes, a campeã é "p" com 11.5%, seguida por "t" com 9.0%, "s" com 8.0%, "d" com 5.5%, "n" com 4.5%,"c" com 4.%,"v" com 4.0%,"q" com 3.0%,etc..

Note que as porcentagens são agora menores e mais próximas uma da outra. Por isso se Carlos tentar usar o mesmo método usado antes para decifrar, ele terá mais possibilidade de destrocar erradamente (por quê?).

Se Carlos ficar desencorajado por tal possibilidade de erros e tentar achar um outro método? Ele pode aplicar TODAS as chaves sobre 26 letras até descobrir a chave correta, por tentativa e erro (como Carlos percebe que não chegou à chave correta, se ele não conhece a frase original?). Mas o total de chaves possíveis é 26! (i.e., fatorial de 26) que é igual ao produto 25x24x23x...3x2, valor em torno de 4 seguido por 26 zeros (por quê fatorial de 26?). Esse número é tão grande que mesmo com a ajuda de um computador Carlos não conseguiria decifrar rapidamente, a não ser que tenha muita sorte e acerte depois de algumas dezenas de tentativas - mas isso é mais difícil que acertar na loteria!

O leitor mais interessado poderá ler na minha página WWW sobre outros tipos de ciframento que são de fato usados para proteger informações como senhas em cartões magnéticos de banco e cartões de crédito:
http://www.ime.usp.br/~rt.

FIM - FIM