Implementação de Gerador de Chaves RSA em Python

---

1 Introdução

Pretende-se neste relatório descrever a Implementação de uma função em python, `genRSAkey(l)` que receba como argumento um tamanho de chave ℓ e retorne um tuplo da forma (n, p, q, e, d), respeitando as condições descritas a seguir:

1. $n = pq$ sendo p e q primos e por forma a que $log_2(n) \geq \ell$ (ou seja que n tenha na sua representação binária pelo menos ℓ bits);
2. e inteiro coprimo com $(p-1)(q-1)$, ou seja, o máximo divisor comum entre e e $(p-1)(q-1)$ seja 1;
3. d, por forma a que $ed \equiv 1(mod(p-1)(q-1))$

2 Geração de candidatos a primos p e q

Usando o módulo `random` do Python, a função `genRand(l)` gera números ímpares no intervalo $2^{w-1}+1$ e $2^w - 1$ em que $w = \lfloor \frac{2}{\ell} \rfloor$, sendo que todos os números passíveis de serem gerados por `random.randrange` no intervalo descrito acima são depois **OR**-ed com 1 no seu bit menos significativo, aumentando assim a eficiência do gerador já que isto elimina a necessidade de gerar números candidatos a primos e depois testar num ciclo `while` a divisibilidade destes por 2, visto que qualquer primo > 2 é necessariamente ímpar.

2.1 Crivo de Eratóstenes

Depois de gerado o número p (candidato a primo), verifcar-se-à, numa primeira fase, contra um crivo de Eratóstenes, gerado apenas uma vez aquando a chamada de `genRSAkey(l)` e guardado em `small_primes`, contendo os 1000 primeiros números primos. Se algum destes é factor de p, então sabemos que p não é primo, sendo assim necessário invocar `genRand(l)` novamente e repetir o teste.

2.2 Teste de Miller-Rabin

No caso em que o número p gerado passa pelo processo de *sieving* supra descrito sem serem encontrados factores primos, prosseguimos para o teste de primalidade de Miller-Rabin (probabilístico). É de salientar que este último domina a complexidade temporal de `genRSAkey(l)`. Usar *sieving* é eficaz em minimizar o número de vezes em que é gerado um número composto p e este tem que ser sujeito ao teste de Miller-Rabin, ou seja: se p tem factores primos relativamente pequenos, detectamos desde logo que este é composto, gerando assim um novo p cuja primalidade volta a ser testada pelo mesmo processo, minimizando assim o número de iterações "desnecessárias" do teste de M-R. Formalmente, dada uma constante B, só serão sujeitos ao teste de M-R os canidatos a primos que se revelam B-*smooth* **TODO: INCLUIR FONTE SOBRE B-SMOOTH**.

O valor de `max_rounds` em `genRSAkey(l)`, corresponde ao número máximo de testemunhas aleatórias a geradas no intervalo $[2, n-2]$ com relação à primalidade de um inteiro n ímpar tal que $n > 3$. Já que o teste de M-R retorna `False` quando n é composto, porque foi gerado um a que é testemunha da existência de factores primos de n, então podemos dizer que se n é um número composto, quanto mais iterações do teste de M-R forem feitas para esse n, maior é a probabilidade de encontrar uma testemunha a (gerada aleatóriamente) que respeite esses critérios e determine que n não é primo.

Baseando-nos no parágrafo anterior, podemos concluir que: se n é um número ímpar composto e ao fim de `max_rounds` o teste M-R não retornou `False`, então n é um "falso primo".

Sejam as variáveis aleatórias:

- $Y_k \leftarrow n$ é declarado primo depois de k iterações do teste de M-R;
- $X \leftarrow n$ é um número composto (sendo $\overline{X} \leftarrow n$ é um número primo).

Por **TODO: INCLUIR FONTE** é possivel mostrar que pelo menos $\frac{3}{4}$ das testemunhas aleatórias em $a \in [2, n-2]$ atestam que n é composto, que é o mesmo que dizer que no máximo $\frac{1}{4}$ das testemunhas aleatórias no mesmo intervalo, declaram que este número é primo. Com isto podemos dizer que $Pr[Y_k|X] \leq \frac{1}{4}$ para uma iteração do teste de M-R, ou $k = 1$. Sendo que para todas as k iterações do teste de M-R, é independente a escolha de a, mostrou-se em **TODO: INCLUIR FONTE** que a probabilidade de erro deste teste pode ser descrita em função de k como $Pr[X|Y_k] \leq (\frac{1}{4})^k$ e que $Pr[Y_k|X] = Pr[X|Y_k]$ **TODO: CLARIFICAR ISTO**.

Em vista dos resultados obtidos escolheu-se um $k = 100$, coincidente com `max_rounds` na função de geração de tuplos para as chaves RSA, o que admite uma porbabilidade de erro igual ou inferior a $\frac{1}{4^{100}}$, algo admissível na geração dos factores primos de n.

2.3 Noção de *strong primes* em RSA

TODO: FALAR DAQUELA COISA DA SAFEDIST

3 Escolha do expoente público

4 Cálculo de d

5 Tempos de execução e conclusões finais