RSA and Wiener's attack

TAMREF
tamref.yun@snu.ac.kr

CNT@SSHS — May 28, 2019

Contents

1 RSA cryptosystem

RSA는 소인수분해의 어려움에 기반하고 있는 암호체계입니다. 풀기 어려운 암호들 중에 가장 원리가 간단한 축에 속하기 때문에 이미 알고 있는 사람들도 많을 것이라고 생각하지만, 우선 간단히 review하고 넘어가도록 합시다.

(a) 큰 소수 p, q를 고릅니다. $N = pq$. 별다른 mention이 없으면 $p < q$입니다.

(b) $\varphi(N) = (p-1)(q-1)$을 계산합니다.

(c) $\gcd(e, \varphi(N)) = 1$인 e를 골라 공개하고, e의 mod $\varphi(N)$에 대한 inverse d를 계산합니다: $de = 1 \pmod{\varphi(N)}$입니다.

(d) 평문(message) m을 고릅니다. 당연히 $0 < m < N$이고, $\gcd(m, N) = 1$이어야 합니다.

(e) 암호문(ciphertext) $c = m^e \pmod{N}$를 계산하여 수신자에게 전달합니다. 이 때 N과 e는 공개하되 d는 수신자에게 **비밀스럽게** 전달합니다.

(f) 수신자는 **비밀스럽게** 전달받은 d를 가지고 $m = c^d \pmod{N}$을 계산합니다.

Question 1. RSA cryptosystem

(a) 이 암호 시스템에서 public하게 공개되어도 되는 정보는 N, e 뿐입니다. p, q, 또는 $\varphi(N)$ 중 하나가 노출되었을 때 공격자가 m을 복원해내는 방법을 제시하세요.

(b) (e)와 (f)를 실제로 보여 보세요.

(c) RSA에 대한 감각을 익히기 위해, $p = 53, q = 61, e = 17, m = 123$으로 두고 실제로 (a)..(f)를 simulate해 보세요.

TMI 1 일반적으로 e는 17, 65537과 같은 페르마 소수로 설정합니다. m^e를 계산할 때 $m, m^2, m^4, \cdots$를 이용할 텐데, $e = 2^k + 1$꼴이면 곱셈 횟수가 줄어들어 좋습니다. 또 $\gcd(e, \varphi(N)) = 1$을 테스트할 때 e가 소수라면 $p - 1 \neq 0 \pmod{e}$만 체크하면 되니, 여러모로 연산이 편리합니다. $2^k + 1$꼴의 소수는 $2^{2^i} + 1$ 꼴의 페르마 소수밖에 없다는 사실을 알아두세요!

TMI 2 일반적으로 $p < q < 2p$를 만족하도록 p, q를 잡습니다. Bertrand's postulate에 의해 $(p, 2p)$에는 무조건 소수가 존재함이 보장되어 있고, 굳이 $q > 2p$를 써봤자 Brute - force attack에 가장 취약한 값은 p이기 때문입니다. 암호화/복호화에 드는 계산 비용만 커지고 보안이 더 강해지지는 않는다는 것입니다.

2 RSA attacks

RSA cryptosystem을 지키는 문제인 소인수분해는 너무나 강력했습니다. 그래서 소인수분해를 통으로 푸는 대신, RSA cryptosystem의 조건에서 몇 가지 제약 조건을 추가해서 혹은 Brute-Force보다 훨씬 빠른 시간에 문제를 해결하는 bypass를 RSA attack이라고 부릅니다. 당연히 굉장히 많은 종류의 variation이 있지만, 오늘은 그 중에서 가장 방법론이 예쁜 *Wiener's attack*만을 알아보고자 합니다.

Claim 1. (Wiener) Secret exponent d가 $d \leq \frac{N^{1/4}}{3}$을 만족한다면, $\varphi(N)$은 다항 시간에 계산할 수 있다.

TMI 3 이런 'RSA attack'들은 일반적으로 아주 이해하기 어려운 계산과, 그 계산을 전부 소개하기에는 그다지 appealing하지 않은 결과를 갖습니다. 유독 Don Coppersmith는 이 분야에서 (어려운 계산과 함께) 흥미로운 결과를 많이 냈는데, 궁금하다면 Coppersmith's methods를 알아보기 바랍니다.

- public exponent e가 작다면 다항 시간에 N을 소인수분해할 수 있습니다. Wiener's attack에선 공격자가 d의 크기를 모르기 때문에 공격 성공 여부를 장담할 수 없는 것과 대조적입니다.
- 만약 p를 $O(N^{1/4})$의 오차로 estimate할 수 있다면 N을 다항 시간에 소인수분해할 수 있습니다. 그래서 $q-p=O(N^{1/4})$라면 $\sqrt{N}$이 p에 대한 estimator가 됩니다.

2.1 Continued fraction

어떤 수를 적당한 유리수로 근사하는 방법이 Decimal approximation만 있지는 않습니다. 예를 들어 $\sqrt{2}$를 근사하려 한다고 합시다.

- $\sqrt{2}=1$+0.XXX이므로, 어떤 $b>1$에 대해 $\sqrt{2}=1+\frac{1}{b}$로 표현할 수 있습니다. 구체적으로 $b=\sqrt{2}+1$입니다.
- $b=2$+0.XXX이므로, 어떤 $c>1$에 대해 $\sqrt{2}=1+\frac{1}{b}=1+\frac{1}{2+\frac{1}{c}}$로 나타낼 수 있습니다. 구체적으로 $c=b=\sqrt{2}+1$입니다.
- $b=2$+0.XXX이므로...

$$\sqrt{2}=1+\frac{1}{2+\frac{1}{2+\frac{1}{2+\frac{1}{\cdots}}}}$$

위와 같은 일련의 과정을 $\sqrt{2}$의 **연분수 전개(continued fraction)**라고 하고, $\sqrt{2}=[1;2,2,2,\cdots]$라고 씁니다. 일반적으로는 $\alpha=[\alpha_0;\alpha_1,\alpha_2,\cdots]$와 같이 쓰는데, $i>0$에 대해 α_i는 양의 정수여야 합니다. 이 때, 연분수 전개 중 앞의 몇 개 항만 계산한 것을 continued fraction $[\alpha_0;\alpha_1,\cdots]$의 convergent라고 합니다.

Question 2. Continued Fraction

(a) Convergent $[1;2]$, $[1;2,2]$, $[1;2,2,2]$를 계산하고 $\sqrt{2}$의 값과 비교하세요.

(b) α가 유리수라면 그 연분수 전개는 유한개의 항만을 가짐을 보이세요. 이는 **어떤 알고리즘**과 동일합니까? $\alpha = n + \frac{p}{q}$일 때 ($p < q$), 연분수 전개의 길이가 $O(\log q)$임을 보이세요.

(c) 제곱수가 아닌 d에 대해 $\sqrt{d}$의 연분수 전개는 주기성을 가짐을 보이세요.[a] 이 문제는 순전히 흥미를 위해 끼워넣은 것으로 해결할 필요는 없습니다.

[a] 신기하게도, 무리수 중에서 periodic continued fraction을 갖는 수는 $\sqrt{d}$꼴 뿐입니다. 반대로, $\sqrt{d}$의 decimal approximation에는 0..9의 모든 digit이 random하게 나타난다고 **강하게 추측되고 있습니다**.

다음의 사실을 증명 없이 받아들이기로 합니다:

Theorem 1. (Dirichlet, 1842) 기약분수 p/q와 x/y에 대해,

$$\left|\frac{x}{y} - \frac{p}{q}\right| < \frac{1}{2q^2}$$

가 성립한다면 p/q는 x/y의 연분수 전개의 convergent 중 하나이다.

2.2 Linking the Continued Fraction to RSA

어떤 적당한 자연수 t에 대해, $ed = 1 + t\varphi(N)$으로 나타낼 수 있습니다. 이 식을 변형하면 다음과 같은 식이 얻어집니다.

$$\frac{e}{\varphi(N)} - \frac{t}{d} = \frac{1}{d\varphi(N)} \tag{1}$$

만약 $d < \varphi(N)/2$라면, $\frac{e}{\varphi(N)}$을 연분수 전개하는 과정에서 Theorem 1에 의해서 $\frac{t}{d}$가 얻어질 것입니다. $\gcd(d,t) = 1$이기 때문에 d, t 모두 계산가능합니다.

그렇다면 $\frac{ed-1}{t} = \varphi(N)$이 계산되고, $\varphi(N) = N - (p+q) + 1$에서 $p+q$를 계산할 수 있을 것이고, 결과적으로 $N = pq$와 $p+q$를 모두 알고 있으니 p, q를 *얻을 수 있습니다...?*

함정은 바로 우리가 $\varphi(N)$을 **모른다**는 것입니다. 애초에 $\varphi(N)$을 알고 있다면 직접적으로 $p+q$를 얻을 수 있기 때문에 연분수 전개를 굳이 할 이유도 없습니다! 결국 저 식은 $\varphi(N)$을 계산하기 위해 $\varphi(N)$을 필요로 하는 셈이니, 활용가치가 없습니다.

하지만, 굳이 $\frac{t}{d}$를 근사하기 위해 정확히 $\frac{e}{\varphi(N)}$을 사용할 필요가 있을까요? $\varphi(N)$에서 어느 정도 벗어난 M을 쓰더라도, reasonable한 크기의 d에 대해 $\left|\frac{e}{M} - \frac{t}{d}\right| < \frac{1}{2d^2}$가 성립하기만 하면 됩니다! 그래서 우리는 $\varphi(N)$의 근사치로 무려 N을 사용합니다.

$p < q < 2p$를 가정할 때, N은 생각보다 괜찮은 근사치입니다.

$$N - \varphi(N) = p + q - 1 < 3p - 1 < 3\sqrt{N} - 1. \tag{2}$$

$$\begin{aligned}\left|\frac{e}{N}-\frac{t}{d}\right| &= \frac{|de-tN|}{Nd} \\ &= \frac{|1+t(\varphi(N)-N)|}{Nd} \\ &= \frac{3t\sqrt{N}}{Nd} = \frac{3t}{d\sqrt{N}}. \end{aligned} \tag{3}$$

이 때 $t \geq d$라면 $e > \varphi(N)$이 되어 모순입니다. 따라서 (3)의 오차는 $\frac{3}{\sqrt{N}}$으로 bound되고, $d < N^{0.25}/\sqrt{6}$이라면 이 값은 $\frac{1}{2d^2}$보다 작게 되어 Claim이 증명됩니다. □

Question 3. Wiener's attack

$N = 144369293$, $e = 42913357$이 주어져 있습니다.

(a) Wolframalpha를 이용하여 e/N의 continued fraction을 구하고, 0번째부터 7번째까지 convergent까지 그 값을 구하세요. 커맨드는 ContinuedFraction[x]입니다.

(b) 1번째 convergent부터 그 값을 t/d로 추정하여 $\varphi(N)$을 구하세요. 이차방정식 $x^2 - (N - \varphi(N) + 1)x + N = 0$을 풀어 p, q를 구하세요. d의 값은 얼마입니까?

Question 4. Post - Wiener Methodologies

(a) $p < q < 2p$의 조건에서, $N - \varphi(N) + 1 \in (2\sqrt{N}, \frac{3}{\sqrt{2}}\sqrt{N})$임을 보이세요. (Hint : $q = N/p$로 두고, p의 범위를 N에 대해 나타내세요.)

(b) $\varphi(N)$의 근사값으로 N대신 $N - 2\sqrt{N} + 1$이나 $N - \frac{3}{\sqrt{2}}\sqrt{N} + 1$을 사용하면 더 좋은 결과를 얻을 수 있을지도 모릅니다. 각각의 근사식을 어떤 조건에서 사용해야 할지 논하세요.

놀랍게도 이 성질을 이용해서 N대신 $N - a\sqrt{N} + 1$ 꼴을 사용하는 논문이 존재합니다(de Weger, 2002). 이외에도 e의 크기에 따라 d의 bound가 결정되는 기괴한 여러 방법들이 있습니다.