

Cryptography Final Report

PRIMES is in P

臺灣大學資訊工程學系 B04902012 劉瀚聲

在得知這學期的密碼學報告可以自選相關主題時，腦海中浮現的第一選擇，便是這篇「PRIMES is in P」。質數在大多數的密碼系統中，往往是相當重要的一環。身為一個資工系學生，目前的研究領域又是關於演算法與複雜度，這篇論文一直在我的 Wishing List 的前幾位。PRIMES 屬於 coNP 相當顯然，但 PRIMES 在 NP 內並不直觀。可以想像，這篇論文的在 2002 年的發表震驚了多少猜測 PRIMES 屬於 NP-Complete 或 coNP-Complete 的學者。

一些教授和網路上大多數的網友對這篇論文的評價都是「直白易懂」、「半小時內可以讀完」之類。但我親自讀之後，並不覺得它有傳聞中的那麼簡單。除了某些引理用到了一些我不曾學過的定義和性質（如分圓多項式等），相對影響較大的，是裡面某些推導或敘述並沒有給出詳細的原因，試著把敘述的正確性證明一次，卻發現原因並不顯然等諸如此類的情況。

0 Notation

1 Overview

Algorithm 1.1 判定質數的算法

```
0 Input: integer  $n > 1$ .
1 If  $n = a^b$  for some  $a \in \mathbb{N}$  and  $b > 1$ , output COMPOSITE.
2 Find the smallest  $r$  such that  $o_r(n) > \lg^2 n$ .
3 If  $1 < (a, n) < n$  for some  $a \leq r$ , output COMPOSITE.
4 If  $n \leq r$ , output PRIME.
5 For  $a$  from 1 to  $\lfloor \sqrt{\phi(r)} \lg n \rfloor$  do:
    If  $(X + a)^n \not\equiv X^n + a \pmod{X^r - 1, n}$ , output COMPOSITE.
6 Output PRIME.
```

Theorem 1.2 當算法 1.1 輸出 COMPOSITE 時， n 為合數

第 2 節為定理 1.2 之證明。

Theorem 1.3 當算法 1.1 輸出 PRIME 時， n 為質數

第 3 節為定理 1.3 之證明。

Theorem 1.4 算法 1.1 的時間複雜度為 $O(\lg^{12} n)$

第 4 節為定理 1.4 之證明。

定理 1.2 及定理 1.3 保證了算法的正確性。而由於輸入規模為 $\lg n$ ，定理 1.4 保證了算法的運行時間為多項式時間。故算法 1.1 是一個 PRIME 的多項式時間算法。

2 Correctness When Output COMPOSITE

當算法在第 1 步或第 3 步輸出 COMPOSITE 時， n 是合數，因為第 1 步的 a 和第 3 步的 (a, n) 會是一個 n 的非平凡因數。

Lemma 2.1 $a \in \mathbb{Z}$ ， $n \in \mathbb{N}$ ， $n \geq 2$ ， $(a, n) = 1$ ，那麼 n 為質數若且唯若

$$(X + a)^n \equiv X^n + a \pmod{n}$$

當算法在第 5 步輸出 COMPOSITE 時，表示 $(X + a)^n \not\equiv X^n + a \pmod{X^r - 1, n}$ ，因此 $(X + a)^n \not\equiv X^n + a \pmod{n}$ 。根據引理 2.1， n 為合數。

2.1 Proof to Lemma 2.1

根據二項式定理， $\forall 0 < i < n$ ， $(X + a)^n$ 中 X^i 的係數為 $C_i^n a^{n-i}$ ，而 X^n 的係數顯然同餘。

(1) n 為質數

由於 $\forall 0 < i < n$ ， $C_i^n = \frac{n!}{i!(n-i)!}$ ， $n \mid (n!)$ ， $n \nmid (i!(n-i)!)$ ，故 $C_i^n a^{n-i} \equiv 0 \pmod{n}$ 。根據費瑪小定理， $a^n \equiv a \pmod{n}$ 。故 $(X + a)^n \equiv X^n + a \pmod{n}$ 。

(2) n 為合數

對於 n 的任何一個質因數 q ，若 $q^k \mid n$ 但 $q^{k+1} \nmid n$ ，那麼由於 $C_q^n = \frac{n(n-1)\dots(n-q+1)}{q(q-1)(q-2)\dots(1)} \cdot \frac{n}{q}$ ，而 $q^k \nmid \frac{n}{q}$ ，且分子及分母中之其他項皆與 q 互質，故 $n \nmid C_q^n$ 。亦即， $C_q^n a^{n-q} \not\equiv 0 \pmod{n}$ 。

3 Correctness When Output PRIME

當算法在第 4 步輸出 PRIME 時， n 為質數。因為第 3 步和 $n \leq r$ 保證了 $\forall a < n, (a, n) = 1$ 。以下證明算法在第 6 步輸出 PRIME 的正確性。

由於 $o_r(n) > 1$ ， n 必定有質因數 p 滿足 $o_r(p) > 1$ 。又因為通過了第 3 步和第 4 步的檢驗，所以 $(r, n) = 1$ ，且 $p > r$ ，故 $p, n \in \mathbb{Z}_r^*$ 。另外，令 $l = \lfloor \sqrt{\phi(r)} \lg n \rfloor$ 。在第 5 步中，算法檢驗了 l 個等式。由於第 5 步沒有輸出 COMPOSITE，因此對於所有的 $0 \leq a \leq l$ ，都有：

$$(X + a)^n \equiv X^n + a \pmod{X^r - 1, n}$$

由於 p 是 n 的因數，故：

$$(X + a)^n \equiv X^n + a \pmod{X^r - 1, n}$$

Lemma 3.1 $(X + a)^{\frac{n}{p}} \equiv X^{\frac{n}{p}} + a \pmod{X^r - 1, p}$ 。

對於多項式函數 f 和自然數 m ，定義 m 對於 f 是冪同構的，如果 $f(X)^m \equiv f(X^m) \pmod{X^r - 1, p}$ 。由上面的敘述可知，對所有的 $0 \leq a \leq l$ ， n 、 p 、 $\frac{n}{p}$ 對於 $(X + a)$ 都是冪同構的。

Lemma 3.2 如果 m_1 、 m_2 對於 $f(X)$ 都是冪同構的，那麼 $m_1 m_2$ 對於 $f(X)$ 也是冪同構的。

Lemma 3.3 如果 m 對於 $f_1(X)$ 、 $f_2(X)$ 都是冪同構的，那麼 m 對於 $f_1(X)f_2(X)$ 也是冪同構的。

接下來考慮以下幾個集合：令 $I = \{p^i(\frac{n}{p})^j \mid i, j \geq 0\}$ ， $P = \{\prod_{a=0}^l (X + a)^{e_a} \mid e_a \geq 0\}$ 。根據引理 3.2 和 3.3， I 中的每一個元素對於 P 中的每一個元素都是冪同構的。由於 n 、 p 、 $\frac{n}{p}$ 都分別和 r 互質，故 I 模 r 所形成的集合會在 \mathbb{Z}_r^* 內，而且是一個

群。令 G 為那個群，亦即， $G = I/r\mathbb{Z} = \{p^i(\frac{n}{p})^j \bmod r | i, j \geq 0\}$ ，並令 $t = |G|$ 。由於 $o_r(n) > \lg^2 n$ ，故 $t > \lg^2 n$ 。令 $Q_r(X)$ 為 r 次分圓多項式，根據分圓多項式的性質， $Q_r(X) \mid (X^r - 1)$ ，且在 \mathbb{F}_p 上， $Q_r(X)$ 是若干個 $o_r(p)$ 次不可約多項式的乘積。令 $h(X)$ 是其中一個這樣的多項式。由於 $o_r(p) > 1$ ，故 $\deg(h(X)) > 1$ 。令 \mathcal{G} 為 P 模 $h(X)$ 再把係數模 p 得到的集合。由於 $h(X)$ 不可約，故 \mathcal{G} 是一個乘法群。 \mathcal{G} 可以看作是由 X 、 $(X+1)$ 、 $(X+2)$ 、.....、 $(X+l)$ 所生成的、在 $\mathbb{F}_p/h(X)$ 上的乘法群。

Lemma 3.4 $|\mathcal{G}| \geq \binom{t+l}{t-1}$ 。

Lemma 3.5 若 n 不是 p 的冪次，則 $|\mathcal{G}| \leq n^{\sqrt{t}}$ 。

根據引理 3.4，

4 Time Complexity of Algorithm

第 1 步枚舉 b 自 1 至 $\lg n$ ，二分搜尋對應的 a ，檢驗 a^b 與 n 的關係。時間複雜度 $O(\lg^3 n)$ 。

第 2 步自 1 開始枚舉 r ，檢驗是否 $\forall 1 \leq i \leq \lg^2 n, n^i \not\equiv 1 \pmod{r}$ 。時間複雜度 $O(r \lg n)$ 。

第 3 步枚舉 a 自 1 至 r ，計算 (a, n) 。時間複雜度 $O(r \lg n)$ 。

第 4 步時間複雜度 $O(1)$ 。

第 5 步每次迭代可用快速冪在 $\lg n$ 次多項式乘法內算出 $(X+a)^n \pmod{X^r - 1, n}$ 的值，而多項式的次數不超過 r ，故時間複雜度為 $O((\sqrt{\phi(r)} \lg n)(r^2 \lg n)) = O(r^{\frac{5}{2}} \lg^2 n)$ 。

第 6 步時間複雜度 $O(1)$ 。

故整體時間瓶頸為第 5 步，複雜度為 $O(r^{\frac{5}{2}} \lg^2 n)$ 。故只須證明 r 的上界。

令 $B = \lceil \lg^5 n \rceil$ ， $S = n^{\lfloor \lg B \rfloor} \prod_{i=1}^{\lfloor \lg^2 n \rfloor} (n^i - 1)$ 。

$$S < n^{\lfloor \lg B \rfloor} \prod_{i=1}^{\lfloor \lg^2 n \rfloor} (n^i) = n^{\lfloor \lg B \rfloor + \frac{1}{2} \lg^2 n (\lg^2 n - 1)} \leq n^{\lg^4 n} \leq 2^B。$$

考慮 $R = \min\{R' | R' \nmid S\}$ ，由於 $\forall i \in [1, \lfloor \lg^2 n \rfloor], R \nmid (n^i - 1)$ ，故 $o_R(n) > \lg^2 n$ ，亦即 R 是 r 的一個上界。

Lemma 4.1 (*Nair [1]*) 令 $LCM(m)$ 表示前 m 個自然數的最小公倍數，那麼對於 $m \geq 7$ ， $LCM(m) \geq 2^m$ 。

假設 $R > B$ ，那麼 $\forall i \leq B, i \mid S$ ，亦即 $S \mid LCM(B)$ ，但根據引理 4.1， $LCM(B) \geq 2^B > S$ ，矛盾，故 $R \leq B$ 。從而， $r \leq \lg^5 n$ ，定理得證。

5 References

[1] M. Nair. On Chebyshev-type inequalities for primes. *Amer. Math. Monthly* 89:126–129, 1982.