離散対数問題について

Sooh

2022.2.5

1 Introduction

暗号技術は我々の通信を秘匿することで機密性を高め、相手の認証をすることで真正性を担保する上で欠かせない技術となっている。特に公開鍵暗号は、秘密鍵を表す変数が与えられた場合には容易に解ける問題が、それ以外の変数を与えられると途端に計算困難な問題へと変わることで、安全性が担保されると同時に使いやすさも提供している技術である。具体的には素因数分解問題や離散対数問題及び楕円曲線上の離散対数問題が有名であるが、このうち離散対数問題についての解法に関するアルゴリズムやそれがどの程度の速さで暗号の解読が可能なのかについての議論を行う。

2 DL problem

離散対数問題 (Discrete Logarithm problem) がどのような問題であるかについて定式化しておく。

G を位数 n の有限巡回群とし、γ をこの群の生成元とする。また $\alpha \in G$ とする。この時、以下を満たす最小の x を求める問題を離散対数問題という。

$$\alpha = \gamma^x$$

ここで x が秘密鍵の役割を持つことに注意されたい。実際 γ 及び x が与えられている状況で α を求めることは容易だが、x の代わりに α が与えられると途端に難しくなることが分かるだろう。実用上の入力としては x が数百 bit となったりすることも、この問題をより難しくしている。
以下では特に p を素数として $G = (\mathbb{Z}/p\mathbb{Z})^\times$ の場合について考察していく。

3 Algorithm 1

まずは一番単純な方法を紹介しよう。それは x について全探索を行うことである。これはいかなる群についても試せる手法である。γ が生成元であることから $\gamma^x = \alpha$ なる x は $0 \leq x < n$ に必ず存在する。従って、0 から順に試していくことで $O(n)$ で求めることができる。しかし、このアルゴリズムでは G の位数が大きくなった途端に計算は止まらなくなることは明らかだろう。

4 Algorithm 2

ここで扱うアルゴリズムは Baby-Step Giant-Step Algorithm である。
まずは次のように x を m によって表す。

$$m = \lceil \sqrt{n} \rceil$$

$$x = mq + r \quad (0 \leq r < m)$$

m の決め方から $0 \leq q < m$ であることが分かる。元の式にこれを代入すると

$$\gamma^x = \gamma^{mq+r} = \alpha$$

従って、以下の式を満たす組 (r, q) を見つける問題に帰着される。

$$\gamma^r = \alpha\gamma^{-qm} = \alpha(\gamma^{-m})^q$$

Algorithm 1 では、$0 \leq r < m$ と $0 \leq q < m$ の 2 変数を動かしながら探索する手法だと解釈できる。Algorithm 2 ではメモリを使う、すなわち $\gamma^p \to p \quad (0 \leq p < m)$ のテーブルを作成しておくことで、q を 0 から順に $m-1$ まで動かした時に、毎回テーブルから該当する r が存在するかを調べることで計算量を落とす工夫をしている。なお、γ^{-m} は初めに計算する必要がある。

次に計算量について解析する。このアルゴリズムでは $\gamma^p (0 \leq p < m)$ のテーブルの前計算で $O(m)$ であり、q を全探索していくのでここでも $O(m)$ である。ただしこのテーブルは O(1) で検索できるとする (平衡二分探索木により少なくとも O(logn) で可能である)。さらに、γ^{-1} はフェルマーの小定理及び二分累乗法を用いることで O(logn) で計算できる。以上から全体としても $O(m) = O(\sqrt{n})$ の時間計算量しか掛からない。一方で、ここでは空間計算量も同様に $O(\sqrt{n})$ となってしまうことには気を付けたい。

5 Algorithm 3

次に紹介するアルゴリズムは Pollard ρ-Algorithm である。時間計算量自体は Algorithm 2 と変わらないものの、空間計算量が $O(1)$ で済むという利点がある。
具体的には以下のように動作させる。
群 G を互いに素な部分集合 G_1, G_2, G_3 に分割する。また、関数 f を

$$f : G \to G$$

$$f(\beta) = \begin{cases} \gamma\beta & (\beta \in G_1) \\ \gamma^2 & (\beta \in G_2) \\ \alpha\gamma & (\beta \in G_3) \end{cases}$$

のように定義する。
$x_0 \in \{0, 1, ..., n-1\}$ をランダムに選び、$\beta_0 = \gamma^{x_0}$ と置いて

$$\beta_{i+1} = f(\beta_i)$$

により更新していく。さて、関数 f の定義から任意の $i \in \mathbb{N}$ について

$$\beta_i = \alpha^{x_i}\gamma^{y_i}$$

と表せること及び有限群上での演算であるから、ある $k \neq 0$ が存在し

$$\begin{aligned} & \beta_i = \beta_{i+k} \\ \iff \quad & \alpha^{x_i}\gamma^{y_i} = \alpha^{x_{i+k}}\gamma^{y_{i+k}} \\ \iff \quad & \alpha^{x_i - x_{i+k}} = \gamma^{y_{i+k} - y_i} \end{aligned}$$

$\alpha = \gamma^x$ であったから、この式は次の式の解を求める問題へと帰着される。

$$x(x_i - x_{i+k}) \equiv y_{i+k} - y_i \pmod{n}$$

この合同方程式の解を求めるのは容易なので、この先の詳細は割愛する。

ここからは時間計算量の解析を行う。ρ-Algorithm は高速な素因数分解にも出てくる考え方であるが、「誕生日のパラドクス」と同じことを利用して高い確率で計算が終了すると考えられている。具体的には、異なる n 種類の物からランダムに $\sqrt{n}$ 個程度取ると、2 つ以上取られる物が存在する確率が $\frac{1}{2}$ 以上であることが言える。f の定め方より β_i は明らかにランダムではないが、それに近い挙動を示すことが知られている。従って、$O(\sqrt{n})$ であると考えられる。

この節の冒頭で空間計算量を $O(1)$ にできると言ったが、上のアルゴリズムのままでは (β_i, x_i, y_i) についての情報を溜めておく必要がありそうなので $O(\sqrt{n})$ かかりそうである。実際には $i = 2^k$ となる i についての情報 $(beta_i, x_i, y_i)$ のみを持っておき、$j = i+1, i+2, \ldots, 2i$ まで計算し、$\beta_i = \beta_j$ となったら合同方程式を解くことにし、そうでなければ i を $2i$ に更新して、その情報を代わりに保持することにすれば良い。これによって省メモリなアルゴリズムが完成した。

6 Algorithm 4

本節では Pohling-Hellman Algorithm について述べる。基本的なアイデアとしては、離散対数問題をより小さな問題に分割し計算した後に継ぎ合わせることで、全体としての計算量を落とそうというものである。以下 n が次のように素因数分解できるとする。

$$n = \prod_{p|n} p^{e(p)} \quad (p : prime\ number)$$

また、n_p, γ_p, α_p を以下のように定義する。

$$n_p = n/p^{e(p)}, \quad \gamma_p = \gamma^{n_p}, \quad \alpha_p = \alpha^{n_p}$$

6.1 Algorithm 4-1

まず、γ_p が巡回群の生成元となっていることを示す。

$$\gamma^n = \gamma^{n_p p^{e(p)}} = {\gamma_p}^{p^e(p)}$$

であるから、γ_p の位数が $p^{e(p)}$ であることが分かる。
また、α_p はこの群に属することも次のように確かめられる。

$$\begin{aligned}\alpha_p &= \alpha^{n_p} \\ &= (\gamma^x)^{n_p} \\ &= (\gamma^{n_p})^x \\ &= (\gamma_p)^x\end{aligned}$$

よって、離散対数問題の定義から n の素因数 p については

$$\alpha_p = {\gamma_p}^{x(p)}$$

という DL problem を考えることができる。そして、これらの問題はすでに紹介したアルゴリズムにより $O(\sqrt{p^{e(p)}})$ で計算できる。
詳しい証明は省略するが、元の問題の解 x は連立線形合同方程式

$$x \equiv x(p) \qquad (\mathrm{mod}\ p^{e(p)})$$

の解であるので、中国剰余定理のアルゴリズムにより容易に計算できる。
時間計算量のボトルネックは小さな DL problem を解くところであり、全体で $O(\sqrt{p^e(p)})$ となる。素因数分解が十分速く行えて、かつ複数の小問題に分けられる時にはこれだけでもかなり高速に問題が解けることになる。

6.2 Algorithm 4-2

4-1 では n を位数とする群についての離散対数問題を p^e が位数である巡回群に対する問題に分割して解くことができるアルゴリズムを紹介した。実は、分割された小問題がより高速なアルゴリズムにより解けることが知られている。
$n = p^e$ であるとする。いま $\alpha = \gamma^x$ を解きたい。x を p 進数にした時の各桁の値 $x_i (0 \leq i < e)$ を求める手法を以下に示した。

- x_0 について

$$\alpha = \gamma^x \iff \alpha^{p^{e-1}} = \gamma^{xp^{e-1}}$$

$$\begin{aligned}xp^{e-1} &= (x_0 + x_1 p + \cdots + x_{e-1}p^{e-1})p^{e-1} \\ &= x_0 p^{e-1} + p^e(expression\ of\ x)\end{aligned}$$

位数が p^e の群ゆえ $\gamma^{p^e} = 1$ に注意すると、

$$\alpha^{p^{e-1}} = (\gamma^{p^{e-1}})^{x_0}$$

さらに、$\gamma^{p^{e-1}}$ が位数 p の群の生成元となっていることから、結局 x_0 は位数が p の群に関する LD problem の解であることが分かる。

- x_i について
$x_0, \ldots, x_{i-1}$ まで決定しているとする。

$$\alpha = \gamma^x \iff \alpha\gamma^{-(x_0 + x_1 p + \cdots + x_{i-1}p^{i-1})} = \gamma^{(x_i p^i + \cdots + x_{e-1}p^{e-1})}$$

となるので、両辺を p^{e-i-1} 乗することで x_0 の場合と同じような式にすることができ、同様に解くことができる。

さて、時間計算量は結論から言ってしまうと $O(\sum_{p|n} e(p)(\sqrt{p}+logn))$ である。これは各素因数 p についての小問題を考えた時、各桁について解く際の累乗に $O(logn)$、DL problem を解くアルゴリズムの適用により $O(\sqrt{p})$ かかっていることから分かる。n の最大の素因数が小さい場合、Algorithm 2,3 で紹介したものよりも遥かに速く動作することが期待できる。

7 Algorithm 5

最後に LD problem を最も高速に解けるアルゴリズムである index calculus について少し触れる。具体的な手法自体の説明は省略するが、高速な素因数分解に基づくアルゴリズムであり、Algorithm 4 で紹介したものと同様に制約がより小さい問題を解いた結果を用いて、求める答えを導出する手法である。

時間計算量は $L_p[1/2, c+o(1)]$(c: 定数) である。ここで $L_N[a,b]$ は

$$L_N[a,b] = exp(b(logN)^a(loglogN)^{1-a})$$

と定義されている。すなわち、このアルゴリズムの時間計算量は準指数時間であり、Algorithm 1 から 4 で紹介したものが全て指数時間アルゴリズムであることを考えると、明らかに最速である。

8 Evaluation

ここまで LD problem を解くアルゴリズムについて見てきた。暗号解読にかかる時間が指数時間、準指数時間、多項式時間によってどのように異なっているかを表したグラフを以下に載せた。

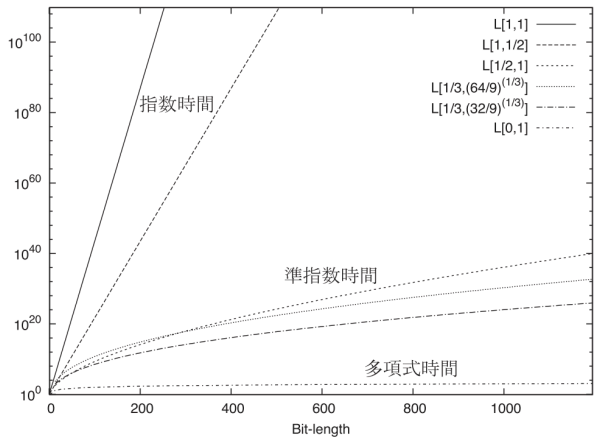

Figure 1: bit 長と時間の関係 (参考文献 2 より引用)

これを見ると、上で紹介した Algorithm4 までの指数時間アルゴリズムは bit 長が数百のオーダーである限りはかかる時間がとてつもなく長く、安全であるように思える。一方で、準指数アルゴリズムでは指数時間に比べて極めて速く DL Problem を解くことができるのもまた明らかである。実際に 2009 年時点で世界記録は 600bit 以上の暗号を解いたという記録もあり、実際に安全であると言えるのは 1000bit オーダーからなどとも言われている。

このように素因数分解の困難性に依拠した暗号の安全性も、より速いアルゴリズムが考案されると完全には信頼できるものではなくなってしまうのかもしれない。

References

Johannes A. Buchmann . INTRODUCTION TO CRYPTOGRAPHY 2nd Edition .P213-226.

林卓也 高木剛 . 離散対数問題解読世界記録更新への道 . 情報処理 Vol.51 No.9 P.1181-1183