

ナップザック暗号における高密度化手法に関する考察

長尾 篤† 森井 昌克†

† 神戸大学大学院工学研究科
657-8501 兵庫県神戸市灘区六甲台町 1-1
a.nagao@stu.kobe-u.ac.jp
mmorii@kobe-u.ac.jp

あらまし ナップザック暗号の多くは 0-1 部分和问题を安全性の根拠とし、安全性の指標として密度を用いている。密度とは平文空間/暗号文空間で与えられる空間の比である。低密度攻撃に対抗するため、ナップザック暗号は高密度であることが望ましい。密度が 1 を超える暗号では平文を中間平文へと一度拡大する手法が取られる。本稿では中間平文への写像において多様性を持たせた場合に安全性が低下する恐れがあることを示す。ある平文が複数の中間平文を導く場合、0-1 部分和问题の解ではない擬似中間平文からの復号が起こり得る。我々は MHK 暗号に上記の問題があることを指摘し、LLL を利用することにより解読に成功した。同様の脆弱性は全ての高密度ナップザック暗号に存在し得る。ナップザック暗号において平文から中間平文を生成する手法は危険である。

Notes on High-Density Knapsack Cryptography

Atsushi Nagao† Masakatu Morii†

† Graduate School of Engineering, Kobe University,
1-1 Rokkodai, Nada, Kobe, 657-8501, Japan
a.nagao@stu.kobe-u.ac.jp
mmorii@kobe-u.ac.jp

Abstract Many knapsack cryptography are based on 0-1 subset sum problem, and use the density as the safety indicator. Density is defined as plaintext space / ciphertext space. The knapsack cryptography requires high-density to resist low-density attack. In the knapsack cryptography whose density is higher than 1, the plaintext is expanded to intermediary plaintext. In this paper, we show that the safety might decrease when one plaintext is represented to plural intermediary plaintext. The decoding from a pseudo intermediary plaintext may be possible, and the derivation of the pseudo intermediary plaintext is simpler than 0-1 subset sum problem. We break the MHK knapsack cryptography, which is one of the high-density knapsack cryptography, using LLL algorithm. All high-density knapsack cryptography may contain similar weakness. In the knapsack cryptography, the technique to generate intermediary plaintext is dangerous.

1 はじめに

ナップザック暗号とは安全性の根拠をナップザック問題に置く公開鍵暗号方式の総称である。特にナップザック問題の一部である0-1部分和问题の探索版を利用した暗号が多く提案されている。探索版0-1部分和问题とは n 個の自然数 $\{a_1, \dots, a_n\}$ と、その部分 C が与えられた際に

$$C = \sum_{i=1}^n a_i x_i \quad (1)$$

を満たす解 $\mathbf{x} = (x_1, \dots, x_n) \in \{0, 1\}^n$ を求める問題である。この探索版0-1部分和问题はNP困難であることが知られている。

1978年、初めてのナップザック暗号であるMH暗号がMerkleとHellmanによって提案された[1]。MH暗号は式(1)における \mathbf{x} を平文、 C を暗号文とし、公開鍵 \mathbf{a} によって暗号化を行う。MH暗号は秘密鍵の特性を元にShamirによって解読された[2]が、MH暗号を改良した暗号が数多く作られた。1985年、LagariasとOdlyzkoが探索版0-1部分和问题そのものを精度よく求解するアルゴリズムを提案した[3]。これを低密度攻撃と呼ぶ。密度とは公開鍵の次元とビット数により定義される部分和问题の安全性の指標である。Lagariasらの攻撃では密度が $d < 0.64$ である部分和问题を解読することができる。1992年、Costerらが低密度攻撃を改良し、 $d < 0.94$ である部分和问题に対して解読が可能であることを示した[4]。以後、ナップザック暗号は低密度攻撃への耐性を持たせるために高密度化が進んでいる。密度が1を超える部分和问题は、複数の解を持つような暗号文が存在する。 C を暗号文としたとき、高密度な部分和问题ではその解 \mathbf{x} を直接平文として暗号化すると一意の復号ができなくなる場合がある。従って平文 m を中間平文 \mathbf{x} へと写像する拡大変換を経た後、暗号文 C を作る手法をとることが多い。

本稿では特に拡大変換に多様性を持たせた場合の問題点を考察する。すなわち、複数の中間平文 $\mathbf{x} \in \{0, 1\}^n$ を単一の平文 m へと逆変換できる場合である。拡大変換を適切に設計していない場合、本来の中間平文に属さない擬似中間

平文 $\mathbf{y} \notin \{0, 1\}^n$ から平文 m を解読できる可能性がある。2012年に村上らの提案したMHK暗号[5]にはこの脆弱性が存在する。MHK暗号は暗号文 C を公開鍵 \mathbf{a} から成る探索版整数部分和问题とした場合の解 $\mathbf{y} \in \mathbb{Z}^n$ の一部から平文 m を復号することができる。この擬似中間平文 \mathbf{y} は格子簡約により容易に求めることが可能であり、MHK暗号で推奨されるパラメータの全てにおいて完全に平文解読を行うことができる。すなわち、MHK暗号はその設計において正しく0-1部分和问题を安全性の根拠として利用できていない。このような擬似中間平文は今後提案されるナップザック暗号でも十分発生し得る。中間平文を設ける際には十分に注意を払い、擬似中間平文の発生を防がなければならない。

2 ナップザック暗号

本章ではナップザック暗号についての概論を述べる。ナップザック暗号とは安全性の根拠をナップザック問題に置く公開鍵暗号方式の総称である。その多くはナップザック問題の一部である部分和问题を利用している。ナップザック暗号に対する攻撃はそれぞれの暗号に対する個別攻撃と部分和问题そのものへの攻撃に大別される。

2.1 MH暗号

MH暗号は1978年にMerkleとHellmanによって提案された最も古いナップザック暗号である。既に解読された暗号ではあるものの、多くのナップザック暗号はMH暗号と基本的な構造が共通している。最も単純なナップザック暗号の一つとしてここで取り上げる。一般にナップザック暗号は部分和问题を形成する二種類の数列とその相互変換によって構成される。すなわち、暗号化において公開鍵を用いて難しい部分和问题を作り、復号では秘密鍵によって容易な部分和问题へと変換し解を得る。MH暗号は超増加数列を利用することにより容易な部分和问题を作成した。

2.1.1 鍵生成

正の整数を要素とする n 次元ベクトル $\mathbf{b} = (b_1, \dots, b_n) \in \mathbb{N}^n$ を

$$b_i > \sum_{j=1}^{i-1} b_j \quad (i = 2, 3, \dots, n) \quad (2)$$

を満たすように定める. このベクトル \mathbf{b} を超増加数列と呼ぶ. 法 M を

$$M > \sum_{j=1}^n b_j \quad (M \in \mathbb{N}) \quad (3)$$

を満たすようにランダムに選択する. 乗数 w を

$$\gcd(w, M) = 1 \quad (1 < w < M) \quad (4)$$

を満たすように定める. n 次元ベクトル $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{N}^n$ を

$$a_i = wb_i \pmod{M} \quad (i = 2, 3, \dots, n) \quad (5)$$

とする. ここで \mathbf{b}, w, M が秘密鍵であり, \mathbf{a} が公開鍵である.

2.1.2 暗号化

平文 \mathbf{m} を n ビット列とする. すなわち

$$\mathbf{m} = (m_1, \dots, m_n) \in \{0, 1\}^n. \quad (6)$$

暗号文 $C \in \mathbb{N}$ は

$$C = \sum_{i=1}^n m_i a_i \quad (7)$$

で計算される.

2.1.3 復号

まず剰余変換により中間暗号文 C' を得る. すなわち,

$$C' = w^{-1}C \pmod{M} \quad (8)$$

$$\Leftrightarrow C' = \sum_{i=1}^n m_i b_i. \quad (9)$$

次に逐次的に平文 \mathbf{m} を復号する. 超増加数列の性質より

$$m_n = \begin{cases} 0 & (C' < \sum_{i=1}^{n-1} b_i) \\ 1 & (C' \geq \sum_{i=1}^{n-1} b_i) \end{cases} \quad (10)$$

が復号される. $C' \leftarrow C' - m_n b_n$, $n \leftarrow n - 1$ として式 (10) を繰り返し, 全ての \mathbf{m} を得る.

2.2 低密度攻撃

本節では Lagarias と Odlyzko の提案した低密度攻撃について述べる. 低密度攻撃は密度の低い部分和问题を効率的に求解することでナップザック暗号を解読する攻撃手法である. ここで密度とは

$$d := \frac{n}{\log_2 \max_i a_i} \quad (11)$$

で定義される値である. 攻撃手法を説明する. まず n 次元ベクトル $\mathbf{a} \in \mathbb{Z}^n$ と平文 $\mathbf{m} \in \{0, 1\}^n$ から式 (7) を用いて暗号文 C が得られているとする. 公開鍵 \mathbf{a} と暗号文 C を用いて, $(n+1, n+1)$ 行列 \mathbf{A} を以下のように定める:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 & -\lambda a_1 \\ 0 & 1 & 0 & \cdots & 0 & -\lambda a_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 & -\lambda a_{n-1} \\ 0 & \cdots & \cdots & 0 & 1 & -\lambda a_n \\ 0 & \cdots & \cdots & \cdots & 0 & \lambda C \end{pmatrix} \quad (12)$$

ここで $\lambda \in \mathbb{Z}$ は効率化のためのパラメータであり, ある程度の大きさを持つ整数である. このとき行列 \mathbf{A} における $(n+1)$ 個の行ベクトルから成る格子 $\mathcal{L}(\mathbf{A})$ に対し格子簡約を行いユークリッドノルムに基づく簡約基底を求める. 簡約基底に以下のような基底 \mathbf{v} が含まれる場合がある.

$$\mathbf{v} = (m_1, m_2, \dots, m_n, 0). \quad (13)$$

格子簡約とは基底をより簡単な基底へと変換するアルゴリズムである. 基底 \mathbf{v} はそのユーク

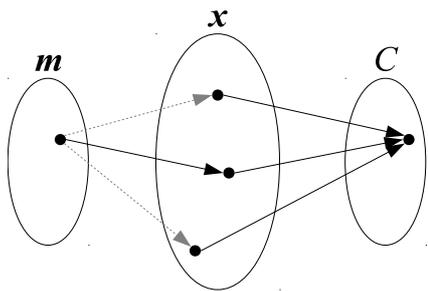


図 1: 高密度化手法の概念図. 高密度な部分和问题は暗号文 C の空間が狭く, 複数の解が存在する. 平文 m を一度中間平文 x へと拡大させる.

リッドノルムが最大でも \sqrt{n} であるため, 簡約基底に含まれる可能性が高い. 格子簡約を近似的に行うアルゴリズムとして LLL アルゴリズム [6] や BKZ アルゴリズム [7] が知られている. Lagarias と Odlyzko は上記の操作により密度 $d < 0.64$ の部分和问题が解読可能であることを示した. 1992 年に Coster らが格子を改良し, $d < 0.94$ の部分和问题にも適用できることを示した. 現在は少なくとも密度が 0.94 を下回るナップザック暗号は安全ではないことが示されている.

2.3 高密度化

低密度攻撃に対する耐性を得るため, ナップザック暗号は高密度であることが求められる. 密度とは式 (11) で与えられる実数値であるが, これはナップザック暗号における暗号化率と捉えることもできる. すなわち, 密度 d とは

$$d = \frac{\text{平文のビット長}}{\text{暗号文の平均ビット長}} \quad (14)$$

と考えることができる. したがって密度 d が 1 を超える場合, 平文空間より暗号文空間の方が小さくなり必然的に暗号文の重複が発生する (図 1 参照). すなわち, 暗号文 C と公開鍵 a から成る探索版 0-1 部分和问题が複数の解を取り得る. そのため密度が 1 を超えるナップザック暗号では事前に平文の拡大変換を行うことで複数解を持ちつつも平文と暗号文の対応関係を維持

している. つまり平文のサイズを n' ($n' < n$) とし, 平文 $m \in \{0,1\}^{n'}$ を中間平文 $x \in \{0,1\}^n$ へと写像する. この手法は様々であるが, 主に二種類に大別することができる.

1. 平文 m から中間平文 x が一意に定まる. (図 1 の濃い実線のみ)
2. 平文 m から多様性のある中間平文 x を生成する. (図 1 の薄い点線を含む)

1. の手法では, 部分和问题の複数解のうち一つが正しい中間平文 x であり, その他の解からは平文を復号できない. この手法をとるナップザック暗号の例として笠原らの提案した $(u | u+v)\Sigma$ PKC [8] がある. この暗号は平文 m を元にしたノイズ系列を追加した中間平文 x を生成することで密度を高めていた. しかし少量の平文 m を推定することで多量のノイズを除去できてしまい, 低密度攻撃が成功する規模まで容易に密度を下げる事ができた [9].

2. の手法ではある平文 m が複数の中間平文 $\{x, x', x'', \dots\}$ を取り得る. したがってこれらの複数解が全て同一の平文 m を導くような逆変換が必要となる. この手法に対する攻撃方法の考察を 4 章で与える.

3 MHK 暗号 [5]

多様性のある中間平文を用いるナップザック暗号の一つに MHK 暗号がある. MHK 暗号は村上らによって SCIS2012 で提案され, SITA2012 では MHK 暗号を用いて鍵を配送するアルゴリズムが提案された [10]. MHK 暗号は完全な乱数列を秘密鍵として用いることで, 秘密鍵の特性に基づく攻撃を防いでいる.

3.1 鍵生成

パラメータとして正整数 $N \in \mathbb{Z}$ と $t \in \mathbb{N}$, および $n \in \mathbb{N}$ を定める. ただし t は $2^t < N$ を満たすようにする. ここで t は平文のビット長であり, 文献 [10] では, $t = 128$ の場合に $224 < \log_2 N < 256$, $256 < n < 384$ が推奨されている.

まず N 以下の正整数を持つ一様乱数ベクトル $\mathbf{s} = (s_1, s_2, \dots, s_n)$ を生成する. 次に正の素数 P を

$$P > \sum_{i=1}^n s_i > \frac{P}{2} \quad (15)$$

を満たす範囲でランダムに生成する. 正の整数 e を

$$P > e > \frac{P}{2} \quad (16)$$

を満たす範囲でランダムに生成する. ベクトル $\mathbf{a} = (a_1, a_2, \dots, a_n)$ を

$$a_i = es_i \bmod P \quad (17)$$

とする. 正整数 Z を $Z = 2^t$ とし, ベクトル $\mathbf{b} = (b_1, b_2, \dots, b_n)$ を

$$b_i = s_i \bmod Z \quad (18)$$

とする. \mathbf{s}, P, e を秘密鍵とし, $\mathbf{a}, \mathbf{b}, Z$ を公開鍵とする.

3.2 暗号化

平文を $m \in \{0, 1, \dots, Z-1\}$ とする. 中間平文 $\mathbf{x} \in \{0, 1\}^n$ を

$$m = \sum_{i=1}^n x_i b_i \bmod Z \quad (19)$$

を満たすようにランダムに生成する. 暗号文 C を

$$C = \sum_{i=1}^n x_i a_i \quad (20)$$

により計算する.

MHK 暗号を鍵交換として用いる場合は, 平文 m を交換する鍵とする. 中間平文 \mathbf{x} を任意に取り, そこから得られる平文 m を共有することで, 平文を自由にとることができなくなるが中間平文の計算が容易となる.

3.3 復号

平文 m において

$$m = \sum_{i=1}^n x_i b_i \bmod Z \quad (21)$$

$$= \sum_{i=1}^n x_i s_i \bmod Z \quad (22)$$

が成立する. また

$$C = \sum_{i=1}^n x_i a_i \quad (23)$$

$$= \sum_{i=1}^n x_i (es_i \bmod P) \quad (24)$$

より,

$$m = (e^{-1}C \bmod P) \bmod Z \quad (25)$$

として復号できる.

4 中間平文の多様性に対する攻撃

高密度ナップザック暗号の設計手法の一つとして, 平文 m を多様性を持つ中間平文 \mathbf{x} へと拡大変換して暗号文 C を生成する手法がある. 密度が 1 を超える探索版 0-1 部分和问题は複数の解を取り得るが, その全てが同一の平文 m から変換した中間平文となるように設計することで平文の一意性を保証する. しかし中間平文に多様性を認めることで, 本来の中間平文における条件 $\mathbf{x} \in \{0, 1\}^n$ から外れる擬似中間平文 $\mathbf{y} \notin \{0, 1\}^n$ からの復号が行える場合がある.

4.1 MHK 暗号を用いた検討

我々は MHK 暗号が擬似中間平文からの復号が行えることを発見した. MHK 暗号は 0-1 部分和问题から外れる, 整数部分和问题の解の一部を擬似中間平文として用いることで解読が可能である. この擬似中間平文は LLL アルゴリズムを 1 度実行することで容易に求めることが可能であり, MHK 暗号を効率的に解読することができる.

MHK 暗号における中間平文は式 (19) から生成される 2 進ベクトル $\mathbf{x} \in \{0, 1\}^n$ である。ここで、擬似中間平文として $\mathbf{y} \in \mathbb{Z}^n$ を考える。正規の平文と暗号文をそれぞれ m, C とし、 $\mathbf{y} = (y_1, \dots, y_n)$ が

$$C = \sum_{i=1}^n y_i a_i \quad (\forall y_i \in \mathbb{Z}) \quad (26)$$

を満たすとする。ここで式 (25) より、擬似中間平文 \mathbf{y} が

$$0 \leq \sum_{i=1}^n y_i s_i \leq P \quad (27)$$

を満たすとき、同時に

$$m = \sum_{i=1}^n y_i b_i \pmod{Z} \quad (28)$$

も成立する。 \mathbf{b}, Z は公開鍵であるので、式 (26, 27) を満たす $\mathbf{y} \in \mathbb{Z}^n$ を導出できれば MHK 暗号は解読できる。すなわち、0-1 部分和问题を解かずに暗号の解読が可能である。ただし式 (27) において \mathbf{s} は秘密鍵であるので、秘密鍵を持たない攻撃者がこの条件を厳密に判定することは困難である。従って、攻撃者は近似的な条件としてユークリッド距離が短い \mathbf{y} を求めることで解読を行う。

擬似中間平文 \mathbf{y} は公開鍵 \mathbf{a} と暗号文 C から成る探索版整数部分和问题の解の一部である。探索版整数部分和问题は低密度攻撃と同一の手法で解を求めることができる。式 (12) で表される行列を用いても良いが、期待する基底を判別し難い。下の行列 \mathbf{A}' を用いると基底の確認が容易である。

$$\mathbf{A}' = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 & -\lambda a_1 & 0 \\ 0 & 1 & 0 & \cdots & 0 & -\lambda a_2 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 & -\lambda a_{n-1} & 0 \\ 0 & \cdots & \cdots & 0 & 1 & -\lambda a_n & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \lambda C & 1 \end{pmatrix} \quad (29)$$

このとき期待される基底は

$$\mathbf{v}' = (y_1, y_2, \dots, y_n, 0, 1) \quad (30)$$

である。本来の 0-1 部分和问题では期待される基底は式 (13) で表される。基底 \mathbf{v} は 0-1 部分和问题の解における条件として $\forall m_i \in \{0, 1\}$ を満たす必要があり、高密度な問題に対しては導出が困難であった。しかし整数部分和问题では式 (27) を満たせば良く、各成分 y_i に個別の条件は無い。格子簡約によりユークリッド距離の短い基底が導出されるため、問題の密度に依存せず高確率で解を求めることができる。

5 計算機実験

MHK 暗号に対する解読実験を行う。文献 [10] で推奨されているパラメータは

$$\begin{cases} 1.75t \leq \log_2 N \leq 2t \\ 2t \leq n \leq 3t \end{cases} \quad (31)$$

であり、これを攻撃の基準とする。解読には式 (29) で示す行列 \mathbf{A}' を用い、 $\lambda = 1000$ とする。また格子簡約アルゴリズムとして LLL アルゴリズムを採用し、NTL ライブラリ [11] に含まれる LLL_XD 関数を利用する。LLL は BKZ よりも解の精度が劣るが、速度面で有利である。本稿で提案する攻撃は厳密な簡約基底を必要としないため LLL アルゴリズムを用いた。まず $t = 128$ の場合に対して、推奨パラメータの範囲を網羅的に攻撃する。このとき

$$224 \leq \log_2 N \leq 256 \quad (32)$$

$$256 \leq n \leq 384 \quad (33)$$

である。パラメータをそれぞれ 8 間隔で変更し、合計 85 種類に対して解読実験を行う。全て異なる鍵および平文に対してそれぞれ 256 回試行し、その全てにおいて完全に平文が解読できた。なお、解読時間も同時に計測した。使用した CPU は Intel Core i7 2600 (1 コアのみ使用) である。最も大きなパラメータである $t = 128, \log_2 N = 256, n = 384$ に対しても平均 11.1 秒で実行できた。

次に $t = 256$ の場合にも同様に攻撃する。計算時間の都合により、パラメータの間隔を 32 に

変更し,

$$448 \leq \log_2 N \leq 512 \quad (34)$$

$$512 \leq n \leq 768 \quad (35)$$

の合計 27 種類のパラメータに攻撃した. この場合においても確率 1 で平文が解読可能であることを確認した. また, 解読時間は $t = 128, \log_2 N = 256, n = 384$ に対して平均 387 秒であった.

以上により, 我々は MHK 暗号に対して安全なパラメータを発見できなかった. したがって MHK 暗号を安全に運用することは難しいと思われる.

6 まとめ

本稿では, ナップザック暗号の高密度化において平文から中間平文を多様性を持って生成する手法の問題点を指摘した. 中間平文に多様性を持たせることで, 想定していない擬似中間平文からの復号が行える場合がある. MHK 暗号にはこの脆弱性が存在し, LLL アルゴリズムを用いることで暗号のパラメータに依らず効率的に擬似中間平文を解読することができた. 中間平文を利用するすべてのナップザック暗号には同様の脆弱性が存在し得る. 本稿で提案した擬似中間平文を導く攻撃は他のナップザック暗号にも有効であると考えられる. ナップザック暗号において平文を拡大変換し中間平文を生成する手法は用いるべきではない.

謝辞

本研究の一部は, 科研費 (基盤研究 (C) 課題番号 23560455) の助成を受けたものである. 本研究の一部は (独) 情報通信研究機構委託研究

「組織間機密通信のための公開鍵システムの研究開発」の一環である.

参考文献

- [1] R. Merkle and M. Hellman. Hiding information and signatures in trapdoor knap-

sacks. *Information Theory, IEEE Transactions on*, Vol. 24, No. 5, pp. 525–530, 1978.

- [2] A. Shamir. A polynomial-time algorithm for breaking the basic merkle-hellman cryptosystem. *Information Theory, IEEE Transactions on*, Vol. 30, No. 5, pp. 699–704, 1984.
- [3] J. C. Lagarias and A. M. Odlyzko. Solving low-density subset sum problems. *J. ACM*, Vol. 32, pp. 229–246, January 1985.
- [4] M.J. Coster, A. Joux, B.A. LaMacchia, A.M. Odlyzko, C.P. Schnorr, and J. Stern. Improved low-density subset sum algorithms. *computational complexity*, Vol. 2, No. 2, pp. 111–128, 1992.
- [5] 村上恭通, 濱正真佑, 笠原正雄. 秘密鍵に乱数列を用いるナップザック暗号. *SCIS2012*, Vol. 2012, , 2012.
- [6] A.K. Lenstra, H.W. Lenstra, and L. Lovász. Factoring polynomials with rational coefficients. *Mathematische Annalen*, Vol. 261, No. 4, pp. 515–534, 1982.
- [7] C.P. Schnorr and M. Euchner. Lattice basis reduction: Improved practical algorithms and solving subset sum problems. *Mathematical programming*, Vol. 66, No. 1, pp. 181–199, 1994.
- [8] M. Kasahara. $(u | u + v)\sigma$ PKC along with challenge problems of small key-size. *SCIC 2011*, Jan 2011. 3A3-5.pdf.
- [9] 長尾篤, 藤堂洋介, 森井昌克. $(u | u + v)\Sigma$ PKC に対する格子攻撃. 電子情報通信学会技術研究報告. LOIS, ライフインテリジェンスとオフィス情報システム, Vol. 111, No. 286, pp. 7–12, nov 2011.

- [10] 村上恭通, 濱正真佑, 笠原正雄. MHK ナップザック暗号を応用した公開鍵配送方式. *SITA2012*, Vol. 2012, , 2012.
- [11] Shoup. NTL: A library for doing number theory. available at <http://www.shoup.net/ntl/>.