情報処理学会研究報告
IPSJ SIG Technical Report

Vol.2009-CSEC-47 No.6
2009/12/18

新しい高密度ナップザック暗号

林　彬[†1]

Merkle-Hellman のナップザック暗号およびその多くの改良提案方式は，既に解読されている．Lagarias-Odlyzko による強力な解読法（LO 法）があるためである．この解読法に耐性を持たせるため，密度が高いナップザックを鍵とする暗号方式を考案した．この暗号を説明し，さらに LO 法に十分な耐性を有することを計算機による解読実験で確かめた．

A New Knapsack Cryptosystem of High Density

Akira Hayashi[†1]

The knapsack cryptosystem by Merkle and Hellman and most of its variants have already been broken. The method by Lagarias and Odlyzko, known as the low density attack, is a powerful tool for attack. We invented a knapsack system with a high density knapsack. Some computer experiments show that it is invulnerable to the LO method.

1. はじめに

ナップザック暗号は Merkle-Hellman[1] が考案した最初期の公開鍵暗号[2] の一つである．同時期に発明された RSA 暗号[3] が現在も使われているのに対し，この暗号（以下 MH 暗号と称する）は実用されていない．強力な解読法が存在するためである．MH 暗号は次のような特徴を有する．

(1)　高速である．暗号化が整数の加算，復号が比較と減算だけからなるからである．

(2)　既に解読法[4),5)] が存在し安全ではなく，現在実用されていない．

†1 金沢工業大学
Kanazawa Institute of Technology

上述のように MH 暗号は解読されたが，その後部分和問題に基づく類似の改良方式が多数提案された．本報告においては，MH 暗号類似の方式を一括してナップザック暗号と呼ぶことにする．各提案は種々の工夫にも拘わらずそのほとんどが解読されている．

しかしながらナップザック暗号は未だ研究が続けられている．その理由は次のように考えられる．

(1)　RSA 暗号，楕円曲線暗号などに比し格段に高速である．

(2)　量子計算機が出現すると RSA 暗号や離散対数型暗号は危うくなる．これに対して，ナップザック暗号は組合せ問題を利用しており，量子計算によってさらに危うくなることはない．

本報告は新しいナップザック暗号を提案し，その安全性を検討するものである．まず MH 暗号を説明し，ついで本提案方式のアルゴリズムを説明する．さらに安全性についての計算機実験結果を述べる．

2. MH 暗号概説

ナップザック暗号はナップザック問題の解法困難性を利用する．公開鍵暗号を構成するためには，解読者には解法困難であるが，正当な受信者には容易に解ける仕掛けを組み込む必要がある．Merkle-Hellman は，超増加数列とモジュラ乗算変換によってこれを実現した[1)]．本節でその仕組みを簡単に記述する．

2.1 鍵の設定

[秘密鍵]

自然数を要素とする超増加ベクトル $b = (b_1, b_2, \ldots, b_n) \in \mathbb{N}^n$（$\mathbb{N}$ は自然数の全体）を定める．ここで b が超増加 (super increasing) であるとは，b が次を満たすときをいう．

$$b_i > \sum_{j=1}^{i-1} b_j, \quad i = 2, 3, \ldots, n. \tag{1}$$

法 M を次の式を満たすランダムな整数とする：

$$M > \sum_{j=1}^{n} b_j. \tag{2}$$

M を余り大きくするとナップザックの密度が低くなり，結果として解読され易くなるので，たとえば $M < 2\sum_{j=1}^{n} b_j$ とする．次を満たす乗数 w をランダムに選び，さらに M を法とする w の逆元 w^{-1} を求める：

$$\gcd(w, M) = 1, \quad 1 < w < M \tag{3}$$

$$ww^{-1} \equiv 1 \pmod{M} \tag{4}$$

[公開鍵]

秘密鍵 b を w と M を用いてモジュラ乗算変換し，$a = (a_1, a_2, \ldots, a_n) \in \mathbb{N}^n$ を作る：

$$a_i = wb_i \bmod M, \quad i = 1, \ldots, n. \tag{5}$$

ここで $x \bmod M$ は x を M で割った剰余 $r,\ 0 \le r < M$, を表す．a を公開鍵とする．

2.2 暗 号 化

平文 m は n ビットであるとする：

$$m = (m_1, \ldots, m_n) \in \{0,1\}^n. \tag{6}$$

この平文に対する暗号文 C は次の式で与えられる．

$$C = \sum_{i=1}^{n} m_i a_i \tag{7}$$

2.3 復 号

まず暗号文 C を w^{-1} と法 M を用いるモジュラ乗算変換により，C' に変換する． この逆変換によって，C と a から m を求めるという難しいナップザック問題は，C' と b から m を求めるという易しいナップザック問題になる． こうして復号は次のようにして行われる．得られる x が平文 m と一致する．

$C \leftarrow w^{-1}C \bmod M,\ x \leftarrow 0^n$
for $i = n$ downto 1 {
  if $C \ge b_i$
    then $\{x_i \leftarrow 1,\ C \leftarrow C - b_i\}$
}

3. 提 案 方 式

提案する方式の特徴は秘密鍵生成法と，それに伴う暗号化の仕組みにある．秘密鍵が超増加でなく「緩増加」(slowly increasing) ⋆1になるため，高密度が得られる．そしてその結果として，ナップザック暗号の解読の強力な方法である Lagarias-Odlyzko 法[5] に耐性を有することになる．

⋆1 緩増加という専門用語があるわけではない．ここでの勝手な言い回しである．

3.1 鍵 の 設 定

[秘密鍵]

2 つの自然数 n と $h(\le n)$ を定める．n は MH 暗号と同じく平文のビット数であり，h は本提案方式に特有のパラメータである．$h = 1$ の場合には MH 暗号になる．ここでは $n = kh$ (k はある整数）とする⋆2.

ベクトル $b = (b_1, b_2, \ldots, b_n) \in \mathbb{Z}^n$ は次の不等式を満たすものとする．

$$b_{i+1} > \sum_{j=0}^{j_1} b_{i-jh}, \quad i = 1, 2, \ldots, n-1. \tag{8}$$

ここで $j_1 = \lfloor (i-1)/h \rfloor$ は $(i-1)/h$ を越えない最大の整数である⋆3.

さらに法 M を，h 個ごとにとった b の要素の総和の最大値より大きな整数とする⋆4.

$$M > b_n + b_{n-h} + b_{n-2h} + \cdots + b_h \tag{9}$$

また M と互いに素な整数 $w,\ 1 < w < M$ および w の法 M に関する逆元 $w^{-1},\ 1 < w^{-1} < M,\ w^{-1}w \equiv 1 \pmod{M}$ を定める．

以上のように定めた b, M, w, w^{-1} を秘密鍵とする．

[公開鍵]

秘密鍵 b から公開鍵 $a = (a_1, a_2, \ldots, a_n)$ への変換は，MH 暗号の場合とまったく同じである：

$$a_i = wb_i \bmod M, \quad i = 1, \ldots, n.$$

以上のように定めた a と h を公開鍵とする．

3.2 暗 号 化

平文 $m = (m_1, \ldots, m_n) \in \{0,1\}^n$ を次のように，h 個の暗号文 $C_1, \ldots, C_h$ に変換する⋆5.

$C_1 = \cdots = C_h = 0,\ j = 1$
for $i = n$ downto 1 {
  if $m_i = 1$
    then $\{C_j = C_j + a_i, \quad j = j \bmod h + 1\}$
}

⋆2 必ずしも h は n の約数である必要はない．説明の簡単のための仮定である．
⋆3 これは $1 \le i - j_1 h \le h$ を意味する．
⋆4 不等式 (9) が成立すれば任意の h 個ごとの b_i の総和よりも M は大きくなる．
⋆5 より一般には，暗号化に際しては秘密鍵を生成するときのパラメータ h 以上の $H \ge h$ を用いて，H 個の暗号文にしてもよい．

Vol.2009-CSEC-47 No.6
2009/12/18

これはつまり，a_i が C_j に加算されるのは，m_i が m_n から始まって $j+th$ 番目の 1 であるときである（t はある整数），ということである．

3.3 復　　号

先ず M と w^{-1} により h 個の暗号文 C_j をモジュラ乗算変換する．これは通常の MH ナップザック暗号の復号と同じく，公開鍵 a の要素の部分和を秘密鍵 b の要素の部分和に変換するためである．

$$C'_j = w^{-1}C_j \bmod M$$

以下簡単のため C'_j を改めて C_j と書く．復号は以下のように進む．得られる x が平文 m と一致する．

$C_j \leftarrow w^{-1}C_j \bmod M,\ j=1,\ldots,h,\ x \leftarrow 0^n$
$j=1$
for $i=n$ downto 1 {
　if $C_j \geq b_i$
　　then $\{x_i=1,\quad C_j=C_j-b_i,\quad j=j \bmod h+1\}$
}

これはつまり，$x_i=1$ となるのは，C_j が b_i 以上のときである．ここで j は 1 に始まるが，$x_i=1$ としたとき j を 1 だけ増す（ただし h の次は 1），ということである．

3.4 数　値　例

小さな数値例を示す．$n=9, h=3$ とする．まず秘密鍵 b を**表 1** の b_i の欄のように定める．これは式 (8) を満たしている．式 (9) を満たすように $M=154$ とする．ランダムに $w=87$ とすれば，$w^{-1}=131$ となる．

表 1 の a_i 欄は w, M による b_i のモジュラ乗算変換の結果である．以上で鍵の設定が終わる．なお密度は $d=9/\log_2 133=1.27$ となる．

暗号化の例として，平文を $m=111111010$ とする．表 1 の m_i 欄がこれを示している．暗号化過程は $i=9$ から始まる．j は暗号文の添え字であり，初期値は $j=1$ である．j_+ は更新後の j の値であり，$m_i=1$ のとき $j_+=j+1$，$m_i=0$ のとき $j_+=j$ となっていることが表 1 から読み取れる．$m_i=1$ のとき，C_j に a_i を加算する．その結果が C_1, C_2, C_3 の欄である．$C_1=130, C_2=172, C_3=233$ となる．

表 2 は復号過程を示す．まず C_j を逆モジュラ乗算変換したものを C'_j とする．最下行の $90, 48, 31$ がこれである．$i=9$，$j=1$ を初期値とし，C'_j を b_i と比較する．$C'_j \geq b_i$ なら $x_i=1$，$C'_j=C'_j-b_i$，$j_+=j+1$，$i=i-1$ とする．$C'_j < b_i$ なら $x_i=0$，$j_+=j$，$i=i-1$ とする．得られた x は m と一致し正しく復号されていることが表 2 から読み取れる．

表 1　提案方式の数値例
Table 1　A numerical example of the proposed scheme.

i	b_i	a_i	m_i	j	j_+	C_1	C_2	C_3
1	2	20	1	1	2	130		
2	10	100	1	3	1			233
3	15	73	1	2	3		172	
4	17	93	1	1	2	110		
5	21	133	1	3	1			133
6	33	99	1	2	3		99	
7	50	38	0	2	2			
8	71	17	1	1	2	17		
9	103	29	0	1	1			
						130	172	233

表 2　提案方式の数値例 (復号)
Table 2　A numerical example of the proposed scheme(deciphering).

i	b_i	C'_1	C'_2	C'_3	x_i	j	j_+
1	2	0			1	1	2
2	10			0	1	3	1
3	15		0		1	2	3
4	17	2			1	1	2
5	21			10	1	3	1
6	33		15		1	2	3
7	50				0	2	2
8	71	19			1	1	2
9	103				0	1	1
		90	48	31			

4. Lagarias-Odlyzko の攻撃法

Lagarias-Odlyzko の攻撃[5](LO 法) は，ナップザック暗号に対する強力な攻撃法で，低密度攻撃とも呼ばれる．密度 d が $d<0.6463$ のナップザック暗号は多項式時間で解読されることが証明されている[5]．ただし整数格子の中の最短ベクトルを多項式時間で見出すことが

できることが仮定されている[⋆1]．

臨界密度はその後 0.6463 から 0.9408[7] にまで改良されている．その結果密度が 1 以下である MH 暗号はほとんど解読されることになった．

4.1 MH 暗号への LO 法による攻撃

公開鍵 a と盗聴した暗号文 C を用いて次のような行列を定義する．

$$A_{\mathrm{LO}} = \begin{pmatrix} 1 & 0 & \cdots & 0 & a_1 \\ 0 & 1 & \cdots & 0 & a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a_n \\ 0 & 0 & \cdots & 0 & C \end{pmatrix} \tag{10}$$

この行列の $n+1$ 個の行をベクトルとみなし，これらのベクトルの整数係数線形結合の全体 L を，A_{LO} が生成する格子といい，$L = L(A_{\mathrm{LO}})$ と表す．A_{LO} を基底行列という．

$$L = \{(y_1, \cdots, y_n, y_{n+1}) \mid y_{n+1} = \sum_{i=1}^{n} a_i y_i - Cy, y_i, y \in \mathbb{Z}\} \tag{11}$$

このとき A_{LO} の $n+1$ 個の行ベクトルは線形独立なので，L の階数は $n+1$ である．LO 法の原理は，平文に対応するベクトルが L の比較的短いベクトルであり，したがって L の短いベクトルを探索すれば平文が見出されるであろうというものである．実際 $i = 1, \ldots, n$ に対して $y_i = m_i$ で，$y = 1$ のとき $y_{n+1} = 0$ であるから

$$v = (m_1, m_2, \cdots, m_n, 0) \tag{12}$$

は L のベクトルであり，その長さは[⋆2]

$$||v|| = \sqrt{m_1^2 + m_2^2 + \cdots + m_n^2} \leq \sqrt{n} \tag{13}$$

となり短い[⋆3]．一方 a_i たちは大きいので，A_{LO} の各行ベクトルは長い．

格子基底の各ベクトルをできるだけ短くすることを，格子基底縮小（または簡約）という[⋆4]．そしてそのための格子基底縮小アルゴリズムのひとつに LLL アルゴリズム[8] がある．

4.2 ナップザックベクトルの密度と解読率の関係

LO 法による攻撃は極めて有効であるが，その解読成功率は，次式で定義されるナップザックベクトルの密度 d と密接に関係する．

$$d = \frac{n}{\max \log_2 a_i} \tag{14}$$

Lagarias-Odlyzko[5] や Coster ら[7] は，d がある値 d_c 以下であれば，平文に対応するベクトルが格子の最短ベクトルである確率が，n が無限大になるとき漸近的に 1 になることを示した．さらに LLL アルゴリズム[8] を最短ベクトルを探索する近似法として用いて，上のことを実験的に検証した[5]．通常のナップザック暗号（MH 暗号）では必然的に $d < 1$ となり，したがって LO 法が解読の有効な道具である．

さてわれわれの新方式の場合はどうか．本方式の秘密鍵はその設定法から分かるように，MH 方式に比べ b は緩やかにしか増加しない数列である．この増加の仕方を調べるために式 (8) の不等式をすべて等式に置き換えて考える．詳細は付録に譲るが，b の要素を公比 r の等比数列とみなすとき，r は次の方程式の解である．

$$r^h - r^{h-1} - 1 = 0 \tag{15}$$

式 (14) の d はこの r を用いて

$$d = \frac{1}{\log_2 r} \tag{16}$$

と表すことができる．これを使うと h と d の関係を求めることができる．**表 3** において d_h は理論値（式 (16)），d は実験値である．ただし，$n = 150$，鍵を 100 個，秘密鍵が満たす不等式 (8) において加える乱数の最大幅を 10 とした．

表 3 h と密度 d の関係（理論限界と実験値）
Table 3 Relationship between h and density d (theoretical and empirical values).

h	d_h（理論）	d（実験）
1	1.00	0.98
2	1.44	1.40
3	1.81	1.73
4	2.15	2.04
5	2.46	2.30
10	3.84	
20	6.17	

4.3 解読実験結果

次元 n と暗号文数 h の種々の値に対して，LO 法による解読実験を行った結果の一部を

⋆1 実際にはこのようなアルゴリズムは存在しない．実用的近似アルゴリズムとして LLL アルゴリズム[8] がある．
⋆2 ここでは 2 乗ノルムとする．
⋆3 最後の不等式は各 m_i が 0 あるいは 1 であることによる．
⋆4 何をもって基底が小さいとするかには種々の考えがある．

述べる．使う格子基底は式 (10) の A_{LO} である．ただし，$C = C_1 + \cdots + C_h$ とした[⋆1]．また試行回数は，鍵と平文の組合せで n の値に応じて 1 万あるいは 10 万とした．

$h = 1$ の場合（MH 暗号）では，$n = 20, 40, 60, 80, 100$ に対し，解読率はそれぞれ，$62, 13, 2.5, 0.70, 0.11(\%)$ であった．これに対し $h = 3$ の場合は，$n = 20, 30, 40$ に対し，解読率はそれぞれ，$4.5, 0.18, 0.03(\%)$ であり，$n = 50$ 以上においては解読はすべて失敗した．

5. おわりに

Merkle-Hellman のナップザック暗号系では，公開鍵ナップザックの密度が 1 以下であるが，密度が 1 を越えるような設計法を示した．これはナップザックの要素が h 個ごとに超増加になるように設定することで可能となった．

本暗号系に対し Lagarias-Odlyzko の方法（低密度攻撃）を適用し，解読実験を行った．その結果 h の増加とともに，解読成功率は急激に減少することを明らかにした．これは安全性の証明にはならないが，少なくとも LO 法に耐性を有することが予想される．なお MH 暗号に対する攻撃法として Shamir の方法[4] もあるが，これが適用できるかどうかは不明である．

今後別の格子基底を用いること（たとえば個別暗号文 C_j を用いる），解読の試行回数をさらに増加するなどの実験の徹底が課題としてある．また解読困難性の理論的裏づけが望まれる．

謝辞 本報告の一部は総務省の Scope の支援によるものである．また研究グループ各位には日頃熱心に討論いただいている．併せ謝意を表する．

参考文献

1) R. C. Merkle and M. E. Hellman, “Hiding information and signatures in trapdoor knapsacks”, IEEE Trans. Inform. Theory, vol.IT-24, no.5, pp.525-530, 1978.
2) W. Diffie and M. E. Hellman : “New Directions in Cryptography”, IEEE Trans. Inform. Theory, vol. IT–22, No. 6, pp.644–654, 1976.
3) R. Rivest, A. Shamir and L. Adleman: “A method for obtaining digital signatures and public-key cryptosystems”, Comm. ACM, vol.21, No. 2, pp.120-126, 1978.
4) A.Shamir, “A polynomial time algorithm for breaking the basic Merkle-Hellman cryptosystems”, IEEE Trans. Inform. Theory, vol. IT-30, no. 5, pp.699-704, 1984.
5) J. C. Lagarias and A. M. Odlyzko, “Solving low density sum problems”, J. ACM, vol.32, pp.229-246, 1985.
6) A. M. Frieze, “On the Lagarias–Odlyzko algorithm for the subset sum problem”, SIAM J. Comput., vol. 15, pp.536–539, 1986.
7) M. J. Coster, A. Joux, B. A. LaMacchia, A. M. Odlyzko, C. -P. Schnorr, and J. Stern, “Improved low-density subset sum algorithms”, Comput. Complexity, 2:111–128, 1992.
8) A. K. Lenstra, H. W. Lenstra, Jr. and L. Lovász, “Factoring Polynomials with Rational Coefficients”, Mathematische Annalen, 261, 515-534, 1982.

⋆1 実際には，第 $n+1$ 列要素を全て n 倍した．これは解読率増加に効果がある[6]

付　　録

A.1 式 (15) の導出

式 (8) における不等号を等号で置き換える：

$$b_{i+1} = \sum_{j=0}^{j_1} b_{i-jh}, \quad i = 1, 2, \ldots, n-1. \tag{17}$$

さらに b を公比 r の等比数列とみなし，

$$b_{i+1} = b_0 r^i \tag{18}$$

と置く．そして式 (17) に式 (18) を代入し b_0 で除して

$$r^i = r^{i-1} + r^{i-h-1} + \cdots + r^{i_0} \tag{19}$$

を得る．ここで i_0, $1 \le i_0 \le h$ はある整数 j_1 について $i_0 = i - 1 - j_1 h$ と書くことができる．したがって

$$\begin{aligned} r^i = r^{i_0+j_1h+1} &= r^{i_0}\{(r^h)^{j_1} + (r^h)^{j_1-1} + \cdots + 1\} \\ &= r^{i_0}\frac{(r^h)^{j_1+1} - 1}{r^h - 1} \end{aligned}$$

となる．ここで r_0 で割って整理すると

$$r^{j_1h+h+1} - r^{j_1h+h} - r^{j_1h+1} + 1 = 0$$

となる．両辺をさらに r^{j_1h+1} で割り，n が十分大きいとき $j_1 h$ も十分大きいので，$r^{-(j_1h+1)}$ の項を無視すれば式 (15) となる．

次に M を式 (9) の右辺で見積もる．

$$\begin{aligned} M &> b_h + b_{2h} + \cdots + b_n \\ &= b_0 r^{h-1}(1 + r^h + r^{2h} + \cdots + r^{n-h}) \\ &= b_0 r^{h-1}\frac{r^n - 1}{r^h - 1} \end{aligned}$$

Vol.2009-CSEC-47 No.6
2009/12/18

ここで式 (15) により $r^h - 1 = r^{h-1}$ であり，これを用いると

$$= b_0(r^n - 1)$$

式 (16) における分母に $\log_2 M$ を代用し，n が十分大きいことを使うと式 (15) が得られる．