

# New Definition of Density on Knapsack Cryptosystems

Noboru Kunihiro  
The University of Tokyo, Japan

# Knapsack Scheme (rough idea)

Public Key:

knapsack:  $\mathbf{a} = \{a_1, a_2, \dots, a_n\}$

Encryption: message  $m = (m_1, \dots, m_n)$

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

Decryption (or Attack):

Solve the equation to recover  $(m_1, \dots, m_n)$ .

Security?

# Subset Sum Problem

Input: knapsack  $\mathbf{a} = \{a_1, a_2, \dots, a_n\}$

$$C \left( = \sum_{i=1}^n m_i a_i \right), \quad k \left( = \sum_{i=1}^n m_i \right)$$

Hamming weight of subset

Output:  $(m_1, \dots, m_n)$

$$m_1, m_2, \dots, m_n \in \{0, 1\}$$

Subset sum problem is NP-hard.

So, the knapsack scheme seem to be difficult to break.

But...

## Many Knapsack Schemes were Broken.

Lagarias-Odlyzko introduced “density”:

$$d = \frac{n}{\log A}$$

, where  $A = \max\{a_i\}$  and  
 $n$  is a message length.

They proved that

if  $d < 0.6463$ , the knapsack scheme is broken  
by lattice attack.

→ low density attack.

Coster et al. improved the bound to 0.9408.

Many schemes were broken by low density attack.

# Shortest Vector Problem

A “lattice” is defined by a set of all integral linear combination of linearly independent vectors:  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ .

$$L(\vec{v}_1, \dots, \vec{v}_m) = \left\{ \sum_{i=1}^m x_i \vec{v}_i : x_i \in \mathbb{Z} \right\}$$

Shortest Vector Problem (SVP):  
find a shortest non-zero vector  $\mathbf{v}$  in  $L$ .

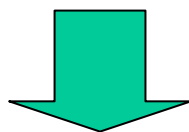
SVP is NP-hard under randomized reductions.

But, it is known that some lattice reduction algorithms solve SVP in practice if the dimension is moderate.

## Remarks on Lattice Attack:

In our presentation,

“a scheme is broken by lattice attack”



“if we can use the oracle which solves SVP,  
the knapsack scheme is broken.”

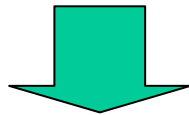
NOT totally broken.

If the dimension is high (300-400), SVP  
is not solvable in practice.

# How to Prevent Low Density Attack?

Some designers choose to **reduce the Hamming weight** of messages.

By reducing the Hamming weight, the message length will be long.



The density becomes larger.

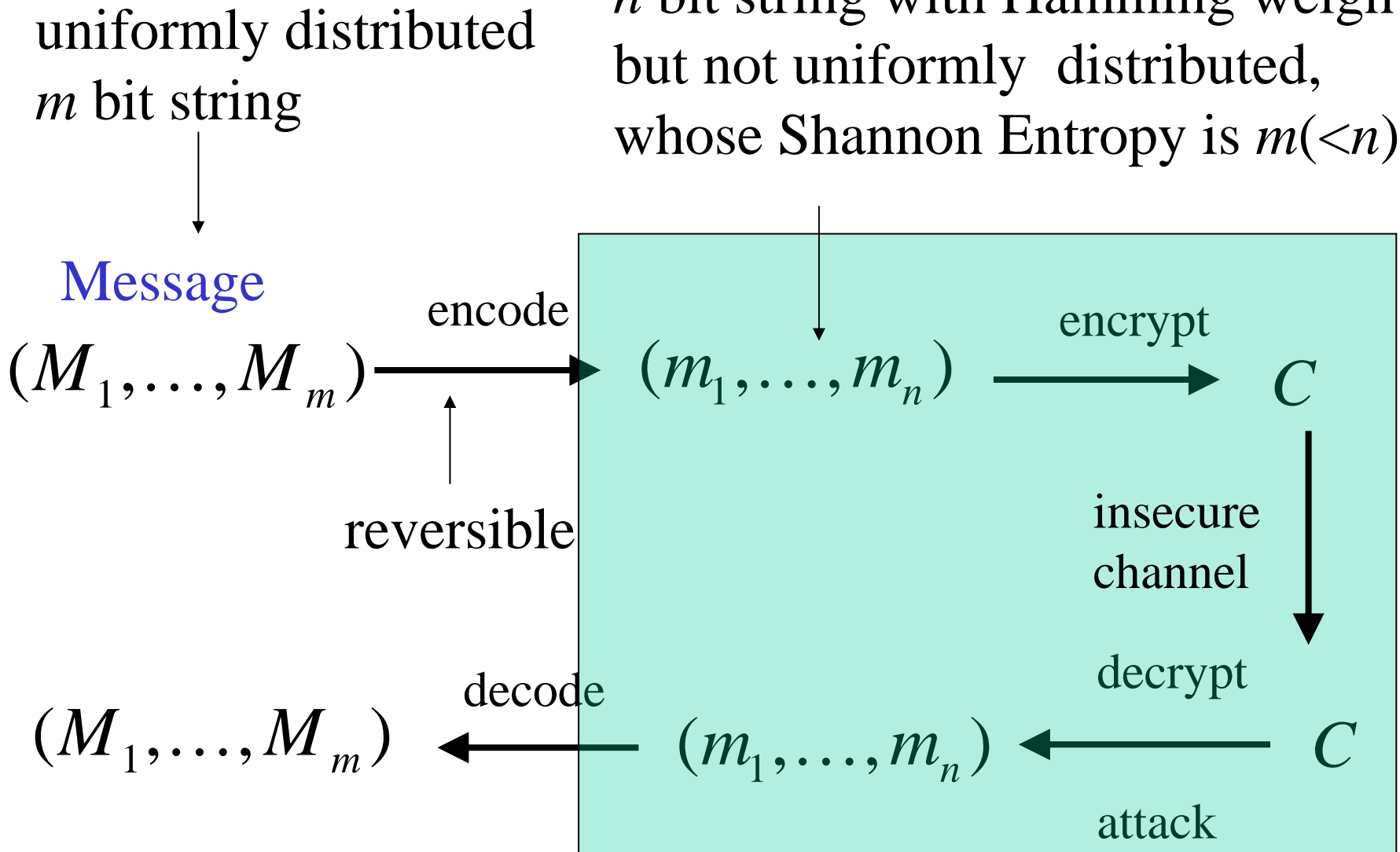
Remember:

$$d = \frac{n}{\log A}$$

Chor-Rivest proposed **low-weight** knapsack scheme. Okamoto-Tanaka-Uchiyama (OTU) also proposed another type of low-weight scheme.

# Low Weight Knapsack Cryptosystem

$n$  bit string with Hamming weight  $k$ , but not uniformly distributed, whose Shannon Entropy is  $m(<n)$

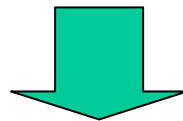




# Security of Low Weight Knapsack Scheme

By reducing the Hamming weight, densities of Chor-Rivest and OTU schemes are larger than 1.

Experimental results by Schnorr-Horner, Omura-Tanaka and Izu et al. show that low weight scheme can be broken by lattice attack even if the density is larger than 1.



Nguyen-Stern introduced another kind of density: pseudo-density.

They theoretically proved that if pseudo-density is low, low weight schemes are broken by lattice attack.

# Lattice Attack on Knapsack Cryptosystem

1. Construct a lattice from a knapsack  $a$  and a ciphertext  $C$ .
2. Obtain the shortest vector in the lattice  
(by using LLL etc.)

## Know Facts 1

density (or pseudo-density) is sufficiently low

→ Shortest vector correspond to real solution of subset sum problem, that is, message

## Know Facts 2

the dimension is **small**

→ we can obtain the “shortest vector”  
by LLL algorithm in practical time.

# Motivation of Our Research

•What is relation between usual density and pseudo-density?

- If the Hamming weight of message is **high**,  
we should use **usual density**.
- If the Hamming weight of message is **low**,  
we should use **pseudo-density**.
- If the Hamming weight is **not so low and not so high**,  
what should we use?

If we have “unified density”, we don’t have to bother which of density should we use.

So, we need unified density.

We must rewrite conditions for unified density.

# Our Contributions

1. introduce **new definition of density  $D$**  which naturally unifies two densities.
2. derive **conditions for our density** so that a knapsack scheme is broken by lattice attack ( $D < 0.8677$ ).
3. show that it is quite difficult to construct a low weight knapsack scheme which is supported by an argument of density.

# Two Variations of Definition of Density

(usual) density  $d = \frac{n}{\log A}$

Lagarias et al. proved that if  $d < 0.6463$ ,  
Coster et al. proved that if  $d < 0.9408$ ,  
a scheme is broken by lattice attack.

## pseudo-density

$$\kappa = \frac{k \log n}{\log A} \quad \text{for small } k$$

Nguyen-Stern proved that  
if  $\kappa$  is low, a scheme is broken by lattice attack.

## New Definition of Density

$$D = \frac{nH\left(\frac{k}{n}\right)}{\log A}, \text{ where } H(x) \text{ is an Entropy function:}$$
$$H(x) = -x \log x - (1-x) \log (1-x).$$

or, since  $m = nH\left(\frac{k}{n}\right)$

$$D \equiv \frac{m}{\log A}$$

## Remarks on Our Density

Remark1:

Lagarias-Odlyzko also remarked that their density is explained as

$$d = \frac{\text{message length}}{\text{ciphertext length}}$$

that is, so called, information ratio.

Remark2: our density:  $D = \frac{nH(k/n)}{\log A} = dH(k/n)$

Intuitively, normalization of the density by multiplying  $H(k/n)$ .

# Our Definition Unifies two Densities

Random message:

Suppose  $M_i$  is 0 with probability 1/2 and 1 with prob.1/2.

(1) Since  $k=n/2$  with overwhelming probability  
by the law of large numbers,  
 $H(k/n)=H(1/2)=1$ . So,  $D=d$ .

(2) (Information theoretic meaning)

True random string cannot be compressed any more.

So,  $n=m$  and  $D=d$ .



## Low Weight Case

Suppose  $k \ll n$ .

$$(1) \quad nH\left(\frac{k}{n}\right) = -n\left(\frac{k}{n}\log\frac{k}{n} + \left(1 - \frac{k}{n}\right)\log\left(1 - \frac{k}{n}\right)\right)$$

$$= k \log n - k \log k - (n - k) \log\left(1 - \frac{k}{n}\right)$$

$$\approx k \log n$$

So,  $D \approx K$

(2) (Information theoretic meaning)

**One easy encoding for string with low Hamming weight**

Bit position of 1 is represented by  $\log n$  bit.

•The number that bit is 1 is  $k$ .

•So, we can represent this sequence at most  $k \log n$ .

This encoding is effective only for small  $k$ .

# The Condition for Unique Decryptability

The necessary condition for unique decryption is

$$\binom{n}{k} \leq kA.$$

Then, 
$$D = \frac{m}{\log A} \leq 1 + \frac{\log k}{\log A} < 1 + \frac{2}{n}$$

By neglecting a small term, we have  $D \leq 1$ .

Remark1: This means that our density is normalization of  $d$ .

Remark2: Our densities of Chor-Rivest and OTU are less than 1.

## Condition for Success of Lattice Attack

We have to rewrite the success condition of lattice attack by using our density  $D$ .

Our analysis is based on Nguyen-Stern (Asiacrypt2005)

We will show that

If  $D < 0.8677$ , the scheme is broken by lattice attack.

More precisely,

if  $D < g_{\text{CJ}}(k/n)$ , the scheme is broken by lattice attack.

These condition is valid for both of random message case and low weight message case.

# Preliminaries of Analysis

Definition:  $N(n, k)$

is the number of integer points in the  $n$ -dimensional sphere of radius  $\sqrt{k}$  centered at the origin.

Theorem 4 in Nguyen-Stern2005

If a lattice is constructed as like Lagarias-Odlyzko, the probability that the shortest vector is not equal to  $\pm m'$  is less than

$$\left(1 + 2(1 + k)^{1/2}\right) \frac{N(n, k)}{A}$$

Remark:  $k$  is the Hamming weight of message.

## Evaluation of $N(n, k)$

Mazo-Odlyzko analyzed  $N(n, k)$  in details.

$$N(n, n/2) \leq 2^{1.54724 \cdots n}$$

$$N(n, n/4) \leq 2^{1.0628 \cdots n}$$

If  $k/n$  is constant,  $N(n, k)$  is exponential of  $n$ .

But, if  $k$  is extremely small, we need another evaluation.

Lemma1 in NS05

$$N(n, k) \leq 2^k \binom{n+k-1}{k}$$

## Precise Evaluation of $N(n,k)$ for small $k$

Nguyen-Stern transformed the inequality into

$$N(n, k) \leq \frac{2^k e^{k(k-1)/(2n)} n^k}{k!}$$

We will transform it into another style by using “inequality between the number of combination and Shannon Entropy”

$$\frac{1}{n+1} 2^{nH(k/n)} \leq \binom{n}{k} \leq 2^{nH(k/n)}$$

Roughly,  $\binom{n}{k} \approx 2^{nH(k/n)}$

## Precise Evaluation of $N(n,k)$ for small $k$ (cont.)

Then, we have

$$N(n, k) \leq 2^k 2^{(n+k)H(k/(n+k))} = 2^{k+(n+k)H(k/(n+k))}$$

Letting  $p=k/n$ , we have

$$\log N(n, k) \leq \underbrace{n(p + (1+p)H(p/(1+p)))}_{\text{depends on only } p} \equiv nf(p)$$

## Condition for Success of Lattice Attack

$$\begin{aligned}\log \Pr < \log \frac{N(n, k)}{A} &= nf(p) - n \frac{H(p)}{D} \\ &= n \left( f(p) - \frac{H(p)}{D} \right)\end{aligned}$$

If  $f(p) - H(p)/D$  is **negative**, the shortest vector corresponds to the message with high probability.

So, in this case, if we can solve SVP, we can recover the message with high probability.



## Condition for Success of Lattice Attack (cont.)

Then,  $f(p) - \frac{H(p)}{D} < 0$

$$D < \frac{H(p)}{f(p)} = \frac{H(p)}{p + (1+p)H(p/(1+p))} \equiv g_{LO}(p)$$

Hence, condition that knapsack scheme is secure to lattice attack is

$$\frac{H(p)}{p + (1+p)H(p/(1+p))} < D < 1$$

Interestingly, the condition depends on only  $p$ .

## Improved Bound based on Coster et al.

Nguyen-Stern 2005

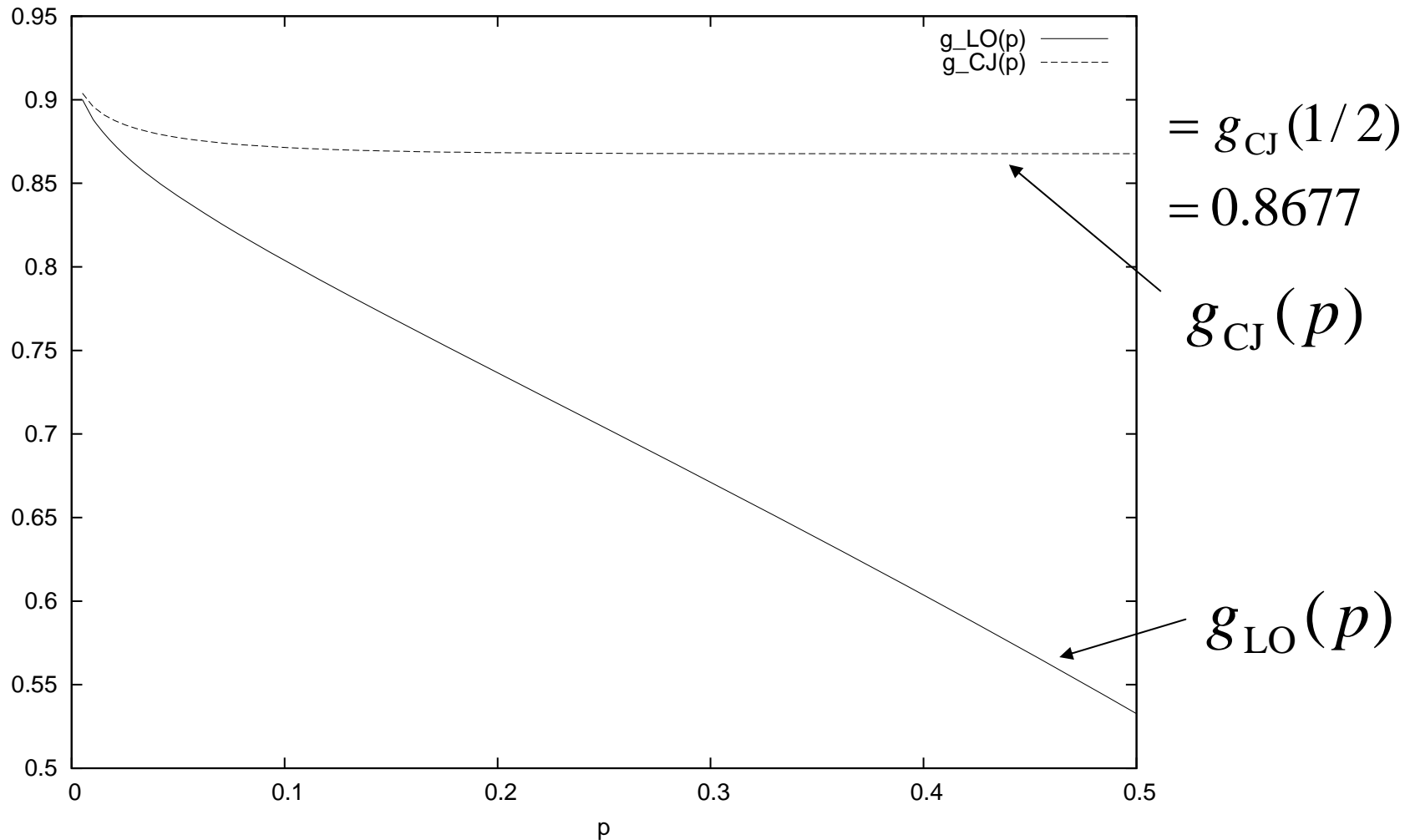
If a lattice is constructed like as Coster et al.,  
the probability that the shortest vector is not  $\pm m'$  is  
less than

$$\left(1 + 2(1 + k)^{1/2}\right) \frac{N(n, k - k^2 / n)}{A}$$

By the similar analysis, we have

$$g_{CJ}(p) \equiv \frac{H(p)}{p - p^2 + (1 + p - p^2)H(1/(1 + p - p^2))} < D \leq 1$$

# Critical Bounds for lattice Attack: $g_{LO}(p)$ and $g_{CJ}(p)$



Remark1: monotonously decreasing function

Remark2: if  $p \rightarrow 0$ ,  $g_{LO}(p), g_{CJ}(p) \rightarrow 1$

## Important two cases:

Case1:

As  $p \rightarrow 0$ ,  $g_{CJ}(p) \rightarrow 1$ .

Hence, it is impossible (or difficult) to construct low weight knapsack scheme which prevents lattice attack.

Case2: If  $p=1/2$ ,

$$g_{CJ}\left(\frac{1}{2}\right) = \frac{1}{1/4 + 5/4H(1/5)} = 0.8677$$

This value is smaller than Coster et al.'s bound: 0.9408.

The reason is why our analysis is based on Lemma1 in NS05, which is not so tight if  $k$  is not small.

# Simple Procedure for judging whether a knapsack scheme is broken by lattice attack

Step1: Calculate  $D = nH(k/n) / \log A$  by  $n$ ,  $k$  and  $A$ .

Step2: If  $D < 0.8677$ , the scheme is broken.

Step3: If  $D < g_{CJ}(k/n)$ , the scheme is broken.

Step4. If  $D < nH(k/n) / \log N(n, n(p-p^2))$ ,  
the scheme is broken.

Otherwise, the scheme is secure against lattice attack.


In Steps 1-3, we need not any complicated calculation.

The above procedure is valid for any values of Hamming weight not like usual density nor pseudo-density.

# Application to Chor-Rivest

cf. Vaudenay broke CR by not lattice attack.

$n$	197	211	243	256
$k$	24	24	24	25
$A$	182bit	185bit	190bit	200bit
$d$	1.08	1.14	1.28	1.28
$\kappa$	1.005	1.002	1.001	1
$D$	0.58	0.58	0.59	0.59
$g_{CJ}(p)$	0.87	0.87	0.87	0.87

critical  
bound of  
density 

In any parameters,  $d > 1$ , but  $D < g_{CJ}(p)$ .

So, CR scheme is broken by lattice attack.

## Conclusion

1. introduced **a new definition of density**, which naturally unifies the previous densities.
2. derived conditions for our density so that a knapsack scheme is broken by lattice attack.

$$D < \frac{H(p)}{p - p^2 + (1 + p - p^2)H(1/(1 + p - p^2))}$$

3. showed that if  $D < 1/(1/4 + 5/4H(1/5)) = 0.8677$ , the knapsack scheme is broken by lattice attack.
4. showed that it is quite difficult to construct a low weight knapsack scheme which is supported by an argument of density.