

# A Knapsack Cryptosystem Secure Against Attacks Using Basis Reduction and Integer Programming

Bala Krishnamoorthy

William Webb

Nathan Moyer

Washington State University

ISMP 2006

August 2, 2006

# Public Key Cryptosystems

---

- Diffie and Hellman (1976)
- Bob wants to receive a message from Alice
  - a 0–1  $n$ -vector
- Bob keeps a public key
- Alice encrypts message and transmits
- Bob decrypts message using his private key
- hard to decrypt without private key

# Public Key Cryptosystems

---

- Eve (eavesdropper) has to solve a “hard” problem to intercept
- one-way functions
  - given  $x$ , easy to find  $y = f(x)$
  - hard to find  $x = f^{-1}(y)$ , given  $y$
- RSA (Rivest-Shamir-Adleman)
  - given  $N$ , need to find two *large* primes  $p, q$  s.t.  $N = pq$

# Knapsack Cryptosystems

---

- Merkle-Hellman scheme (1978)
- Bob creates a superincreasing knapsack  $S = \{s_1, \dots, s_n\}$ 
  - $s_i > \sum_{j=1}^{i-1} s_j$  for  $i = 2, \dots, n$
- chooses  $p, q$  with  $\gcd(p, q) = 1$
- private key –  $(S, p, q)$
- can find  $p^{-1} \pmod q$

# Knapsack Cryptosystems: Encryption

---

- computes  $a_i \equiv ps_i \pmod q$
- public key –  $A = \{a_1, \dots, a_n\}$
- to send message  $x = (x_1, \dots, x_n)$ ,  $x_i = 0$  or  $1$ ,
- Alice computes  $M = \sum_{i=1}^n a_i x_i$  (encryption)
- transmits  $M$  to Bob

# Knapsack Cryptosystems: Decryption

---

- Bob solves

$$\begin{aligned}\sum_{i=1}^n a_i x_i &\equiv M \pmod{q} \\ \Rightarrow p^{-1} \sum_{i=1}^n a_i x_i &\equiv p^{-1} M \pmod{q} \\ \Rightarrow \sum_{i=1}^n s_i x_i &\equiv M' \pmod{q}\end{aligned}$$

- easy to solve (as  $S$  is superincreasing)  
while  $M' > 0$  do  
    find largest  $s_i \in S$  s.t.  $s_i \leq M'$ ;  
    set  $x_i = 1$ ;  
    set  $M' \leftarrow M' - s_i$ ;  
end\_while

# An Example

---

- $n = 5$ ;  $S = \{1, 3, 7, 13, 26\}$ ;  
 $p = 467$ ,  $q = 523$ ;  $p^{-1} \bmod q \equiv 28$ ;  
 $A = \{467, 355, 131, 318, 113\} = a$
- for message  $x = (0, 1, 1, 0, 1)$ , Alice transmits  
 $M = ax = 599$
- Bob calculates  $M' \equiv p^{-1}M \bmod q = 36$

$$36 - 26 = 10 \quad \Rightarrow \quad x_5 = 1$$

$$10 - 7 = 3 \quad \Rightarrow \quad x_3 = 1$$

$$3 - 3 = 0 \quad \Rightarrow \quad x_2 = 1$$

# Knapsack Cryptosystems: Security

---

- to intercept, Eve has to solve

$$\sum_{i=1}^n a_i x_i = M$$

$$x_i \in \{0, 1\}, \quad i = 1, \dots, n$$

- 0–1 knapsack problem is NP-complete
- difficulty to solve knapsack implies security
- Merkle offered a \$100 prize for breaking the code!!

# Attacks using Diophantine Approximation

---

- Shamir (1982) used the superincreasing property of  $S$
- find  $p', q'$  such that  $\{p'A \bmod q'\}$  is superincreasing
- with  $A = \{467, 355, 131, 318, 113\}$ ,
  - $\frac{p'}{q'}$  must lie in one of 466 intervals  $\left[\frac{k}{467}, \frac{k}{467} + \frac{1}{2^4 467}\right]$  for  $k = 1, \dots, 466$
  - $\frac{p'}{q'}$  must lie in one of 354 intervals  $\left[\frac{k}{355}, \frac{k}{355} + \frac{1}{2^3 355}\right]$  for  $k = 1, \dots, 354$
  - intersect intervals, try many choices..
- for  $p' = 53$ ,  $q' = 990$ ,  $p'A \bmod q' = \{1, 5, 13, 24, 49\}$  and  $p'M \bmod q' = 67$  solves the problem!

# Attacks using Diophantine Approximation

---

- Adleman (1983) and Brickell, Lagarias, Odlyzko (1983)
- find integers  $k_1, \dots, k_\ell$  such that  $\frac{k_i}{a_i}$  approximates  $\frac{p}{q}$
- need only a few  $k_i$ 's
- solve an IP to find  $k_i$ 's (Lenstra's algorithm)
- IP created using superincreasing property of  $S$

# Attacks using Basis Reduction

---

- Lagarias and Odlyzko (LO) (1985)
- to solve  $ax = M$ , consider the lattice  $\mathbb{L}(B)$  generated by

$$B = \begin{bmatrix} I & 0 \\ Na & -NM \end{bmatrix}, \quad N \text{ large}$$

- $\hat{x} = \begin{bmatrix} x \\ 0 \end{bmatrix}$  is a shortest vector in  $\mathbb{L}(B)$
- apply LLL basis reduction to  $B$
- check for solution in the reduced basis

# Attacks using Basis Reduction

---

- density =  $\frac{n}{\max_{i=1,\dots,n} \log_2 a_i}$
- $\text{Prob}(\exists \tilde{x} = \begin{bmatrix} x' \\ y \end{bmatrix} \in \mathbb{Z}^{n+1} \mid ax' = My, \|x'\| \leq \|x\|) \rightarrow 0$  as  $n \rightarrow \infty$  **if** density  $< 0.647$
- LO method solves almost all knapsack problems of density  $< 0.647$  in poly time if there is a poly time oracle for solving the shortest vector problem
- claim: LLL acts like such an oracle in practice??
- Coster et al. (1991): density  $< 0.941$

# Attacks using Basis Reduction

---

- LaMacchia (1993) – empirical testing
- used Seysen-Lovász reduction
- instances had density  $< 1$  (needed for unique decoding?)
- $n/2$  of the  $x_i$ 's are set to 1
- for  $n = 106$ , with density = 0.393, ( $a_i$ 's had  $\approx 80$  digits),  
on an average
  - solved 50% instances
  - in 34147 seconds

# Attacks using Integer Programming

---

- try to solve the IP directly:

$$\sum_{i=1}^n a_i x_i = M$$

$$x_i \in \{0, 1\}, \quad i = 1, \dots, n$$

- solvers cannot handle the huge numbers involved
- try exact solvers? (Espinoza et al.)
- apply **Column Basis Reduction**
  - obtain reformulation with smaller numbers using BR
  - run CPLEX on reformulation

# Reasons for Insecurity

---

- structure – superincreasing sequence
- low density
- existence of lattices in which the correct solution is a shortest vector
- success of BR in finding the shortest vectors in such lattices (in practice)

# Why Knapsack??

---

- simple in construction
- fast encryption and decryption
  - addition/subtraction instead of multiplication/exponentiation
- claim that LLL is *highly likely* to find the shortest vector in *almost all* instances:
  - still too early to throw in the towel!!

# A New Knapsack Cryptosystem

---

- construction and structure
- an example
- security against BR-based attacks
- security against IP attacks
- further work

# New Cryptosystem: Construction

---

- pick  $r$  primes  $p_1, \dots, p_r$  (private)
- $\forall i$  pick  $m_i \leq p_i$  numbers distinct mod  $p_i$ ,  
put them in set  $S_i$  (Note: these numbers can be bigger than  $p_i$ )
- $\forall i$  find  $A_i = S_i \times \frac{p_1 \dots p_r}{p_i}$
- $n = \sum_{i=1}^r m_i$ ; knapsack coefficients are  
 $A = \{A_1, \dots, A_r\} = a$  (public)
- can receive messages (0–1  $n$ -vectors) which  
pick *one* element from  $A_i$  for each  $i$
- can further disguise  $a_{ij}$ 's using modular multiplication

# New Cryptosystem: Construction

---

- Eve needs to solve

$$\sum_{i=1}^r \sum_{j=1}^{m_i} a_{ij} x_{ij} = M$$

$$\sum_{j=1}^{m_i} x_{ij} = 1 \quad i = 1, \dots, r$$

$$x_{ij} \in \{0, 1\}, \quad i = 1, \dots, r, \quad j = 1, \dots, m_i$$

- can set  $\sum_{j=1}^{m_i} x_{ij} = v_i \quad i = 1, \dots, r$ , where  $v_i \geq 1$ , or

- can set  $\sum_{j=1}^{m_i} w_{ij} x_{ij} = v_i \quad i = 1, \dots, r$ , for some

$$w_{ij} \in \mathbb{Z}, \quad v_i \geq 1$$

# New Cryptosystem: Example

---

- $r = 2$ ,  $p_1 = 5$ ,  $p_2 = 7$ ;  
 $S_1 = \{21, 17, 13, 34, 25\}$ ,  $S_2 = \{22, 25, 31, 33\}$ ;
- $A_1 = S_1 \times 7 = \{147, 119, 91, 238, 175\}$ ,  
 $A_2 = S_2 \times 5 = \{110, 125, 155, 165\}$ ;
- all sums  $a_{1i} + a_{2j}$  are distinct mod 35
- Alice sends  $119 + 165 = 284 = M$
- $M = 284 \equiv 4 \pmod{5}$ , hence Bob needs  $7s_1 \equiv 4 \pmod{5}$   
i.e.,  $s_1 \equiv 12 \pmod{5} \equiv 2 \pmod{5}$   
looks up  $S_1$  to find  $17 \equiv 2 \pmod{5}$   
hence, first part of message is  $(0, 1, 0, 0, 0)$

# New Cryptosystem: Density

---

- Message space has  $\prod_{i=1}^r m_i$  messages
- restrictive, but helps to increase density!
- with  $m_i = m \forall i$ ,

$$\text{density} = \frac{rm}{\max_{i=1,\dots,n} \log_2 a_i} \geq \frac{rm}{((r-1) \log_2 p_{\max} + \log_2 R)}$$

where  $p_{\max} = \max_i p_i$  and  $R = \max_{i,j} S_{ij}$

- can choose  $m, p_i$ 's and  $R$  such that density is much bigger than 1!
- only  $m$  of the  $x_{ij}$ 's is 1

# Security against Basis Reduction

---

- a (simpler) test problem for Lagarias-Odlyzko method
  - $m_i = m \forall i$ ;  $I = [9R, 10R]$ ,  $J = [10R, 11R]$  for large  $R$
  - choose  $s_{i1} \in I$  randomly  $\forall i$ , let  $\sum_{i=1}^r s_{i1} = M$
  - for  $j = 2, \dots, m$ , choose  $s_{ij} \in J$  randomly for  $i = 1, \dots, r - 1$
  - set  $s_{rj} = \sum_{i=1}^{r-1} s_{ij} + 2s_{r1} - M$try LO method on

$$\sum_{i=1}^r \sum_{j=1}^m s_{ij} x_{ij} = M$$

$$\sum_{j=1}^m x_{ij} = 1 \quad i = 1, \dots, r$$

$$x_{ij} \in \{0, 1\}, \quad i = 1, \dots, r, \quad j = 1, \dots, m$$

# Test for Lagarias-Odlyzko Method

---

- apply LLL reduction to  $B = \begin{bmatrix} I & 0 \\ ND & -Nb \end{bmatrix}$ , where

$$D = \begin{bmatrix} s_1 & \dots & s_r \\ 1 \dots 1 & \dots & \\ & \dots & \\ & & 1 \dots 1 \end{bmatrix}, \text{ and } b = \begin{bmatrix} M \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

- solution is shortest vector in  $\mathbb{L}(B)$ :  $x_{i1} = 1 \forall i, \|\hat{x}\| = \sqrt{r}$
- there are  $m - 1$  “short” vectors in  $\mathbb{L}(B)$  with length  $\sqrt{r + 4}$
- Results:
  - LO method always finds  $\hat{x}$  for  $r \leq 6, m \leq 6, R \leq 10^8$
  - 100 trials with  $r = m = 20, R = 10^{20}$ , LO found  $\hat{x}$  in **none** of the instances!

# Security against Basis Reduction

---

- if cardinality constraints are ignored for the LO method
  - can prove that there exists exponentially many vectors in  $\mathbb{L}(B)$  equal or shorter in length to the solution vector
- performance of LLL similar on actual instances (with  $p_1, \dots, p_r$ )
- even block Korkine-Zolotarev (BKZ) reduction produced similar results
  - arguably, the strongest BR algorithm implemented

# Security against IP Reformulations

---

- original IP is  $\{x \in \{0, 1\}^n \mid Dx = b\}$  where

$$D = \begin{bmatrix} A_1 & \dots & A_r \\ 1 \dots 1 & & \\ & \dots & \\ & & 1 \dots 1 \end{bmatrix}, \text{ and } b = \begin{bmatrix} M \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

- rewrite IP as  $\{x \in \mathbb{Z}^n \mid Bx \begin{pmatrix} = \\ \leq \end{pmatrix} f\}$  where

$$B = \begin{bmatrix} D \\ -I \\ I \end{bmatrix}, \text{ and } f = \begin{bmatrix} b \\ 0 \\ e \end{bmatrix}$$

- using BKZ reduction, find reduced  $\tilde{B}, \tilde{f}$ ;  
try to solve  $\{y \in \mathbb{Z}^n \mid \tilde{B}y \begin{pmatrix} = \\ \leq \end{pmatrix} \tilde{f}\}$

# IP Reformulation Tests

---

\*\* – unsolved instances

† – CPLEX encountered numerical instability

$r$	$m$	$n = rm$	$R$	BRED	CPLEX	#BB
10	13	130	100	1	1	10
40	13	520	100	74	4	59
80	13	1040	100	1097	13	1110
150	13	1950	100	> 4 hrs	**	**
3	73	219	100	2	1	20
3	179	537	100	35	1	10
3	541	1623	100	2764	9	705
3	1229	3687	100	> 4 hrs	**	**
10	23	230	$10^{10}$	1	1	12
10	23	230	$10^{20}$	7	4	190
10	23	230	$10^{30}$	9	18	312
10	23	230	$10^{32}$	9	†	†

# Further Work

---

- a promising knapsack cryptosystem
- have clever ideas to defeat diophantine approximation
  - even under modular multiplication
- formalize hardness results for BR and IP reformulations
- need to consider other possible attacks
  - *better* BR algorithms
  - more numerically stable CPLEX
  - exact MIP solver (Espinoza et al.) and improvements