EXTENSION OF BRICKELL'S ALGORITHM FOR BREAKING HIGH DENSITY KNAPSACKS

F. Jorissen, J. Vandewalle, R. Govaerts

Katholieke Universiteit Leuven,
Department of Electrical Engineering, ESAT Laboratory
K. Mercierlaan 94, B-3030 Heverlee, Belgium

INTRODUCTION

A knapsack (or subset-sum) problem that is useful for cryptographic purposes, consists of a set of n positive integers $a = \{a_1, a_2, \ldots a_n\}$, called the knapsack a, and a sum s. The density d of a knapsack is defined to be $n/\log_2(a_i)_{max}$. The knapsack problem then consists of finding the set, if any, of binary numbers $x = \{x_1, x_2, \ldots, x_n\}$, such that $\Sigma x_i.a_i = s$.

In {2}, E.F.Brickell presented an algorithm for breaking knapsacks of low density. It was expected {2} that an improved version of it would solve most knapsacks of density less than .54. For knapsacks of increasing density, the probability of having so-called "small coefficient identities" (SCI's) in the knapsack also increases. The presence of these SCI's appears to be the reason for the failure of Brickell's algorithm for knapsack problems of high density.

In this paper, a simple technique is proposed for circumventing this problem effectively, so that the algorithm becomes capable of solving knapsacks of higher densities.

Tests for low dimensions have shown that the effectiveness of the algorithm can thus be increased to solve knapsacks of densities even >.9 with high probability and reasonable supplementary computer power. Since the number of available knapsacks decreases very quickly with its density (cfr. fig. p.7), only a relatively small set of high density knapsacks of low dimension remains difficult to solve.
Moreover, all the advantages of Brickell's algorithm are preserved. The most important of these is that the algorithm is applicable to any knapsack public-key cipher based on the knapsack problem mentioned above, whether the cipher already exists or is to be invented.

D. Chaum and W.L. Price (Eds.): Advances in Cryptology - EUROCRYPT '87, LNCS 304, pp. 109-115, 1988.

THE EXTENSION OF BRICKELL'S ALGORITHM

We shall now give a very brief overview of Brickell's algorithm and the improvements that have been accomplished. More details can be found in {2},{1}. Brickell made the basic observation that a knapsack problem $\Sigma x_i.a_i=s$ is in fact a linear equation in n binary unknown x_i and with non-zero sum s. He concluded that if n linear equations ($\Sigma x_i.y_{ji} = s_j(k)$) could be found such that:
- they are not linearly dependent of each other
- they all conform to the solution x of the knapsack problem
- that finding such set of equations is computationally feasible

that then each knapsack (problem) for which this is possible may be regarded as solved.

In the first part of his algorithm, a technique is described to construct a nxn matrix Y from which the rows y_j correspond to the coefficients of the n equations in the n unknown x_i. The first row y_1 of Y consists of the elements $b_i = a_i.W \bmod M$. The other rows of Y are derived by applying modular mappings with the small sum property on previous rows of Y.{2},{1} These modular mappings are calculated with the L.L.L.-algorithm {3}.

In the second part of his algorithm it is proven that $\Sigma x_i.b_i \in \{s', s'+M, \ldots, s'+(n-1)M\}$, with $s' = s.W \bmod M$. Thus n integers $s'(k) = s'+k.M$ $(0 \leq k \leq n-1)$ are derived, exactly one of which corresponds to the correct binary solution x. Because all rows of Y, except for the first one, have been calculated by applying modular mappings with the small sum property on previous rows, we can now calculate {2},{1} from the set of possible sums of the first row s'(k), a new set of possible sums $s_j(k)$ for every other row j, with: $s_j(k) = \Sigma x_i.y_{ji}$, assuming that $\Sigma x_i.b_i = s'(k)$.
(for exactly one constant value of k, $s_j(k)$ will be the sum of the new knapsack problem $\Sigma x_i.y_{ji} = s_j(k)$ with the same x as the original knapsack problem, and this for every row y_j of Y.) For each of the values k we can thus compose a nx1 matrix S(k) and we can calculate the possible solutions x as nx1 matrices X from the matrix problems Y.X = S(k). If indeed s is the sum of a subset of the a_i's then one of the vectors S(k) will give a correct <u>binary</u> solution X for $Y^{-1}.S(k)$.

If the first part of the algorithm can be completed then it is always straightforward to execute the second part and to solve the knapsack problem.

In experiments run by Brickell, his algorithm appeared to be always successful except when there were identities satisfied by the a_i's of the form:

$\Sigma\alpha_i.a_i = 0$, with $\Sigma|\alpha_i| \leq n$.

Such an identity is called a "small coefficient identity" (SCI). It appears that some "dangerous" types of these SCI's are easily "inherited" from one vector to another through the use of modular mappings, and in particular by modular mappings with the small sum property.{1} Brickell also demonstrated that the expected number of SCI's of a given knapsack increases with its density.

Through experiments it has come clear to us that the basic algorithm of Brickell fails for knapsacks of high densities because the given knapsack contains one or more SCI's. Since the first row y_1=b of Y, consisting of the numbers b_i, is derived from the numbers a_i through a modular mapping, some of the SCI's present in the knapsack a are inherited by the first row of Y. Since all the other rows of Y are analogously derived through modular mappings with the small sum property of the previous rows, it may happen that one or more of the dangerous SCI's of the initial knapsack are present in all the possible rows of Y. As a consequence of this, it becomes impossible to find a matrix Y with linearly independent rows and the algorithm fails.

In order to be able to still use Brickell's algorithm in these cases, we propose to construct the first row b of Y such that its expected number of SCI's is smaller than unity. Hereto we proved that the density of b is so small for practical dimensions n that it would normally not contain any SCI's if it was constructed randomly. It is also proven that, by construction, b can exclusively inherit SCI's present in a, for practical dimensions {1}. It may therefore be concluded that the SCI's of b are mainly introduced by the non-random character of b.

We therefore propose to construct b as:

$$b_i = a_i.W \mod M + k_i.M.$$

The numbers k_i can initially be regarded as relatively small random numbers.

It can be proven that as a consequence of this:

$$\Sigma x_i.b_i \in \{s', s'+M, \ldots, s'+(n-1+\Sigma k_i)\}.$$

So it remains possible for this choice of b to determine $\Sigma x_i.b_i$, but $n+\Sigma k_i$ vectors S(k) have to be tested instead of n.

It is therefore advantageous to keep Σk_i small, e.g. by restricting all the k_i's to $k_i \in [1,K]$, in which K is a small positive integer value.

It has been mathematically and experimentally verified that this new technique has several advantages:

- b can still exclusively contain SCI's of the knapsack a.
- Therefore, b can only contain SCI's if at least one of the SCI's of a is still present in b, the probability of which decreases with increasing K.
- The average supplementary computer power needed, however, can be roughly estimated as the processing of (K-1).n/2 extra vectors S(k), and the implications of larger values of the b_i's, mainly on the execution time of the L.L.L.-algorithm.

These results have been verified experimentally for knapsacks of low dimensions n, in two ways.

1. Knapsack problems with knapsacks of low dimensions and high density have been tested.

For example the knapsack problems with knapsack:

a = { 2 , 5 , 6 , 10 , 20 }, of - dimension n = 5
- high density d = 1.16

could not be solved with the original algorithm of Brickell due to the inheritance of the SCI α = { 0 , 0 , 0 , 2 , -1 } from a to b and from there to all possible rows of the matrix Y.
This knapsack was however solved with the better choice of b.

2. Computer tests have been run on the effectiveness of both the original and the new version of Brickell's algorithm.

The purpose of these Monte-Carlo tests was to gain experimental results on the probability that b contains SCI's, for both algorithms.
As parameters were chosen the dimension n and the density d, plus, for the new algorithm, the value of K and the number of trials T (i.e. the number T of vectors k that may be chosen for a given knapsack to try to find a SCI-free vector b for it). While considering the figures, we must bear the following in mind:

- All figures represent dimension 5. Calculations were also made for dimension 8. For this higher dimension, the figures appear to remain flat till higher densities but then fall steeper.
  This may indicate worse efficiency for high dimensions.
  This observation is in line with results obtained by Lagarias and Odlyzko.
- The figures, however, give rather pessimistic results for low dimensions, since:
  - Pessimistic approximations had to be made to calculate them.
  - It has been assumed that finding a SCI-free vector b is a necessary condition for solving a knapsack. To our experience however, a lot of knapsacks with b contaminated by SCI's could be solved,since only for dangerous SCI's the probability is high that they are inherited by all possible rows of Y.
  - In case only one SCI is transported through all possible rows of Y, a special countermeasure can be taken. {2}
  - The set of knapsacks that will remain unsolved will contain a relatively large fraction of "useless" knapsacks, since for increasing density, the probability also increases that knapsacks are not one-to-one.

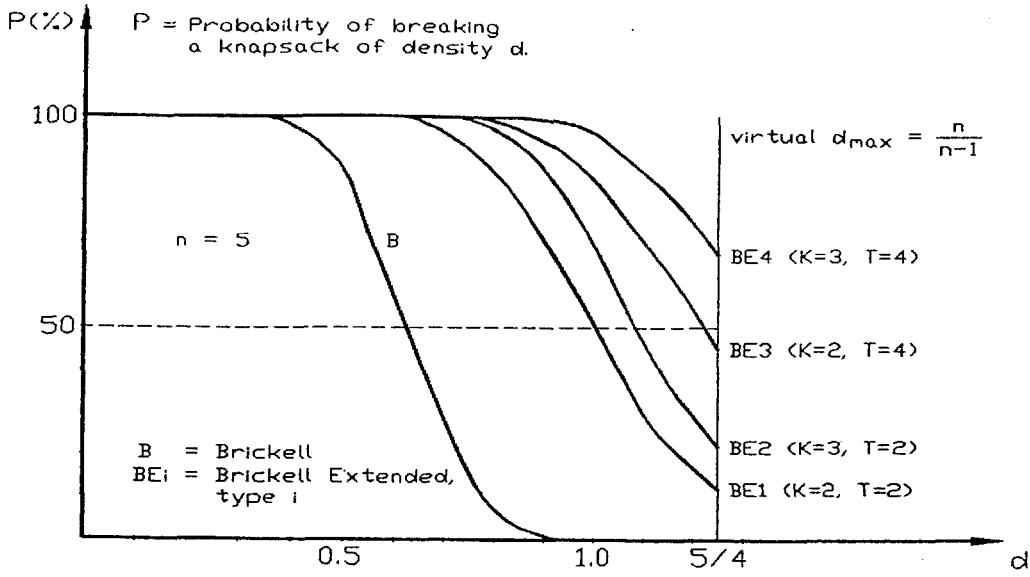
CONCLUSION

In our work, Brickell's algorithm {2} has been extended to the use of better vectors b. {1} This makes the algorithm capable of solving low dimensional knapsacks of high densities larger than .9, with reasonable supplementary computer power.
Since the efficiency of this improvement has only been tested for low dimensions, and considering contradictory results from Lagarias and Odlyzko for high dimensions, it remains to be tested to what extent our improvement remains efficient for increasing dimensions.

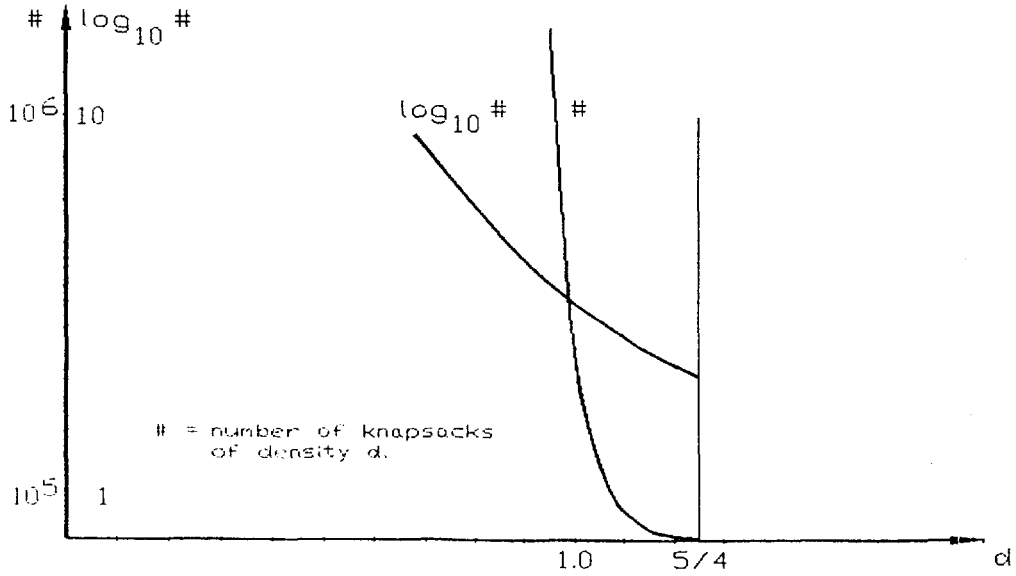
ACKNOWLEDGEMENTS

We are grateful to Y. Desmedt for many suggestions during the thesis work and to S. Claeys for his help with the test programs. We also wish to thank the company CRYPTECH for its support and stimulating interaction. Last but not least, we thank E. Brickell and A. Odlyzko for the discussion on this work.

Figures



Test results of the extended algorithm of Brickell: Probability of breaking a knapsack, with both algorithms, and number of available knapsacks as functions of its density.



REFERENCES

{1} F. Jorissen, "De cryptanalyse van knapzak publieke sleutel geheimschriftvormende algoritmes", M.Sc. Thesis, Katholieke Universiteit Leuven, Belgium, May 1985.

{2} E.F. Brickell, "Solving Low Density Knapsacks", Sandia National Laboratories, Albuquerque, USA.

{3} A.K. Lenstra, H.W. Lenstra Jr., and L. Lovasz, "Factoring Polynomials with Rational Coefficients", Mathematische Annalen, Vol.261, no.4, pp.515-534, 1982.