A PRELIMINARY REPORT ON THE CRYPTANALYSIS OF MERKLE-HELLMAN KNAPSACK CRYPTOSYSTEMS[+]

E. F. Brickell, J. A. Davis and G. J. Simmons

Sandia National Laboratories
Albuquerque, New Mexico 87185

INTRODUCTION

In April, 1982, Adi Shamir caused a furor with the announcement [1] of "A Polynomial Time Algorithm for Breaking Merkle-Hellman Cryptosystems." Like many others who received his "extended abstract," members of the mathematics department at the Sandia National Laboratories undertook a careful study of both the algorithm and the underlying mathematical concepts. This paper summarizes some of our findings. In order to meet the deadline for Crypto'82 the style will be deliberately telegraphic -- and informal. A complete paper will be presented this fall at the Twelfth Conference on Numerical Mathematics and Computing at Winnipeg, the proceedings of which will be published in Congressus Numerantium. It should also be remarked that in discussions with Adi at Crypto'82, we learned that some of the results presented here as new were also arrived at independently by him in the period since his original announcement.

Finally, we shall adopt Adi's notation, etc., wherever possible, so as to assist the reader in making the transition from one paper

+ This work performed at Sandia National Laboratories supported by the U. S. Department of Energy under contract number DE-AC04-76DP00789.

Copyright (c) 1998, Springer-Verlag

to the other. Since Adi's paper was given in the same session at Crypto'82 as this one, and will accompany it in the Proceedings, we assume the reader to be familiar with Adi's development and notation, and will only discuss those points that are essential to illustrate arguments made here.

A WAY OUT FOR THE KNAPSACK DESIGNER

In this section we show how the designer of a Merkle-Hellman knapsack can, with a feasible amount of precomputation, generate a hard knapsack that is immune to the cryptanalytic algorithm described in Adi's "extended abstract."

Given an n weight hard knapsack $\{a_i\}$ derived from a Merkle-Hellman easy knapsack [2] (super increasing sequence of weights) by the modular (trapdoor) transformation:

$$a_i \equiv W_0 b_i \pmod{M_0} \quad .$$

Adi normalizes the inverse modular functions

$$f_i(z) = a_i z - \left[\frac{a_i z}{M_0}\right] M_0$$

by dividing out the unknown (to the cryptanalyst) modulus M_0 to obtain the functions

$$f_i(v) = a_i v - [a_i v] = a_i v - c_i$$

where $0 \le v = \frac{z}{M_0} \le 1$, $0 \le f_i(v) \le 1$ and c_i is an integer, $c_i < a_i$. The $f_i(v)$ are piecewise linear (sawtooth) functions with slope a_i having a_i zeroes uniformly spaced $\frac{1}{a_i}$ apart on the v axis. At $v_0 = \frac{W_0^{-1}}{M_0}$, i.e., at the multiplicative inverse of W_0, one of the functions must have a value $< \frac{2^n}{M_0} \approx 2^{-n}$, since $b_i < 2^n$ in a Merkle-Hellman easy knapsack. Hence the minima of $f_i(v)$ closest to v_0 on the left cannot be further away from v_0 than $\frac{2^n}{M_0 a_i} \approx 2^{-3n}$. Note that we have assumed a_i is the modular image of b_i in accordance with Adi's usage. Similarly the closest minima to the left of v_0 of $f_i(v)$, $i = 2, 3$ and 4 (corresponding to b_2, b_3 and b_4) are no

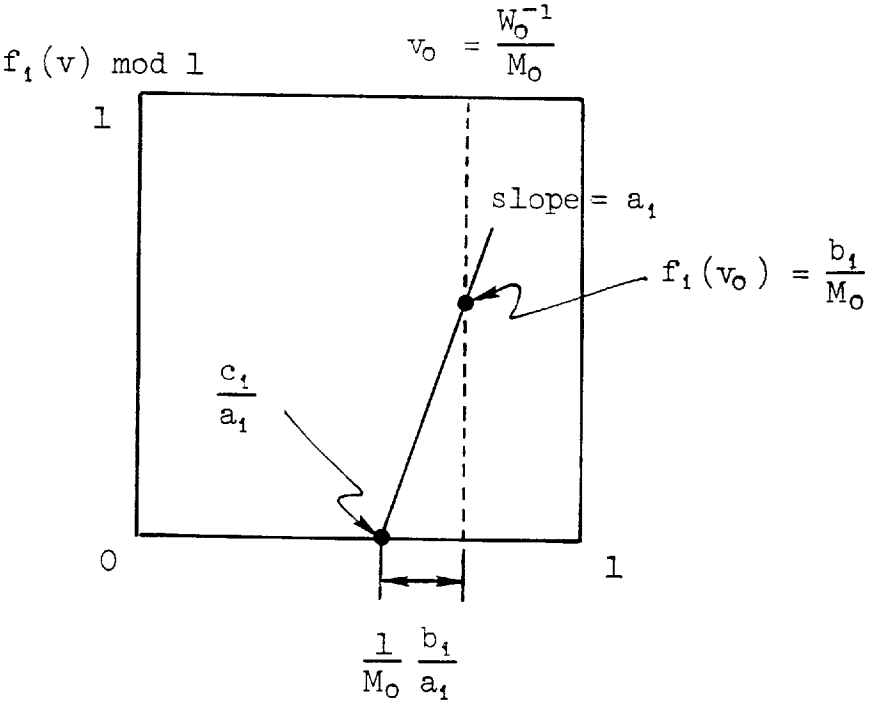
Copyright (c) 1998, Springer-Verlag

further away than $\frac{2^{-n+i-1}}{M_0 a_i}$. The value of v_0 is of course unknown, and in fact is what the cryptanalyst is trying to localize, however the distance between pairs of the minima of the $f_i(v)$, $i = 1, 2, 3$, and 4, closest to v_0 can be expressed. Adi uses approximations of these distances to form three inequalities.

$$\left.\begin{aligned} 0 \le \frac{c_1}{a_1} - \frac{c_2}{a_2} &\le 2^{-3n+1} \\ 0 \le \frac{c_1}{a_1} - \frac{c_3}{a_3} &\le 2^{-3n+2} \\ 0 \le \frac{c_1}{a_1} - \frac{c_4}{a_4} &\le 2^{-3n+3} \end{aligned}\right\} \qquad (1)$$

where $0 \le c_i$ is the number of minima of $f_i(v)$ between 0 and a value of v where the four functions simultaneously satisfy $f_i(v) < 2^{-n+i-1}$. The algorithm is based on the expectation that if for some choice of four weights from $\{a_i\}$ the minima of the $f_i(v)$ do accumulate so close together, i.e., if four integers c_i exist satisfying (1) then this "(almost certainly unique)" point is close to $\frac{W_0^{-1}}{M_0}$.

In keeping with the stated objective of this short paper, we ignore questions of uniqueness, etc., and concentrate instead on how the knapsack designer can cause the system (1) to fail to have a solution. Consider one of the functions $f_i(v)$ in the vicinity of v_0.



Copyright (c) 1998, Springer-Verlag

Since v_0 is the normalized multiplicative inverse of W_0, $f_i(v_0) = \frac{b_i}{M_0}$ for all i. The slope of $f_i(v)$ is a_i everywhere except at the zeroes of $f_i(v)$, hence at v_0 in particular. The distance between v_0 and the minima of $f_i(v)$ closest to v_0 on the left is therefore precisely

$$\frac{c_i}{a_i} \frac{1}{M_0} = v_0 - \frac{c_i}{a_i}$$

so that the closest minima of $f_i(v)$ are located at

$$\frac{c_i}{a_i} = v_0 - \frac{b_i}{a_i} \frac{1}{M_0} \quad .$$

If these values are substituted into system (1) and all terms multiplied through by M_0, one obtains the equivalent system of inequalities

$$\begin{aligned} 0 \le \frac{b_2}{a_2} - \frac{b_1}{a_1} &\le M_0 2^{-3n+1} \\ 0 \le \frac{b_3}{a_3} - \frac{b_1}{a_1} &\le M_0 2^{-3n+2} \\ 0 \le \frac{b_4}{a_4} - \frac{b_1}{a_1} &\le M_0 2^{-3n+3} \end{aligned} \quad . \qquad (1')$$

The essential point on which Adi's algorithm hinges is that the expected values of the ratios $\frac{b_i}{a_i}$, for the smallest weights in a Merkle-Hellman knapsack, are extremely small; i.e., $\frac{b_1}{a_1} \approx 2^{-n}$, $\frac{b_2}{a_2} \approx 2^{-n+1}$, etc. Consequently, with a very high probability, for randomly constructed trapdoor knapsacks, the difference of the smallest ratios are correspondingly small. The form of (1'), and the preceding remarks, however, suggest a way out for the knapsack designer. Roughly speaking, if he can cause one of the weights b_2, b_3, or b_4 to map into an unexpectedly small a_i, the slope of the corresponding $f_i(v)$ will be smaller than expected and the closest minima of $f_i(v)$ will be further to the left of v_0 than expected -- which can cause one or more of the inequalities in (1') to fail. We make this intuitive notion precise in two ways.

First, the designer can force one of the inequalities in (1') to fail by solving the linear inequalities

Copyright (c) 1998, Springer-Verlag

$$\left.\begin{aligned} \frac{M_0}{2} < W_0 b_1 - c_1 M_0 = a_1 < M_0 \\ 0 < W_0 b_1 - c_1 M_0 = a_1 < A \\ 2^{2n-1} < M_0 < 2^{2n} \\ W_0 < M_0 \end{aligned}\right\} \qquad (2)$$

for one of the $a_1 = 2$, 3 or 4 -- where A is chosen to cause $\frac{b_1}{a_1}$ to be larger than expected. The result -- as seen from (1') -- will be to cause the distance between the minima of $f_1(v)$ and $f_1(v)$ closest to v_0 to be too far apart for the corresponding inequality in (1) to be satisfied. Computationally this is an easy problem for the designer -- so long as he need only force a small number of values of the a_1 (only one in the present discussion) to fall in a restricted range, i.e., < A. We next examine the likelihood that (2) has a solution.

Under the assumption that the values $\frac{a_1}{M_0}$ resulting from the modular mappings

$$a_1 \equiv W_0 b_1 \qquad (\text{mod } M_0)$$

for randomly chosen W_0 and M_0, subject to the already stated constraints, are well approximated by independent uniformly distributed random variables, A_1, on the unit interval, we have proven the following theorem.

Theorem 1

For $b_j > b_1$

$$P_{1,j}(z) = P\left[\left(\frac{b_1}{A_j} - \frac{b_1}{A_1}\right) > z\right] = \frac{b_1}{zM_0}\left[\alpha - \frac{b_1}{zM_0}\,\ell n\left(1 + \alpha\,\frac{zM_0}{b_1}\right)\right] + (1-\alpha) \qquad (3)$$

where

$$\alpha = \begin{cases} 1 & \text{if} \quad zM_0 \geq b_j - b_1 \\ \dfrac{b_1}{b_j - zM_0} & \text{otherwise} \end{cases} .$$

Copyright (c) 1998, Springer-Verlag

A Monte Carlo Evaluation of (3) using $z = M_0 2^{-3n+i}$ from (1') and randomly drawn Merkle-Hellman easy weights b_i, modulii M_0 and multipliers W_0 yielded the probabilities

$$P_{12}(z) = 0.30$$

$$P_{13}(z) = 0.36$$

$$P_{14}(z) = 0.41$$

for $n \geq 50$. In other words the probability that the system of inequalities (1) would naturally fail to have a solution is $P_f \approx 0.74$. Even more significant though is the ease with which the designer can find a solution to system (2) which will result in a hard knapsack that is immune to cryptanalysis by algorithm (1).

AN ALTERNATE APPROACH TO INVERTING MERKLE-HELLMAN KNAPSACKS

In the preceding section we showed that it is not necessarily true that in the vicinity of a point at which the four functions $f_i(v)$ are simultaneously very small that all of the minima of the $f_i(v)$ are also very close together. It is true, however, that if one considers some fraction αn of the $f_i(v)$ whose values all satisfy

$$f_i(v_0) < 2^{-n+\alpha n}$$

at v_0, that many, but not necessarily all, of the closest minima will accumulate near v_0. We shall not pursue this approach further -- except to remark that we have proven theorems that strongly suggest that finding k minima of the $f_i(v)$ crowded sufficiently close together is a good indication of a neighborhood to a solution -- and that the knapsack designer appears to be faced with a computationally infeasible task in eluding this algorithm with $k = 4$ or 5, $\alpha \approx 1/6$ and $n \approx 100$.

Instead of looking for accumulations of minima of the $f_i(v)$ as indicators of regions in which solutions have a high probability of existing, we focus instead on the distribution of the values of

Copyright (c) 1998, Springer-Verlag

$f_i(v)$. At v_0, since the a_i are modular images of a super increasing sequence, the values of the $f_i(v_0)$ are exponentially dense near 0. On the other hand at values of v away from v_0 the values of the $f_i(v)$ are nearly uniformly distributed in the unit interval. What this says is that if we consider a bound $B \ll 1$, then in the vicinity of v_0, there will be $\approx n + \log_2 B$ of the $f_i(v)$ whose values are simultaneously less than B while away from v_0 the probability of $f_i(v) < B$ for any i will be roughly proportional to B. We shall make these remarks more precise in a moment -- but the basic idea should be clear. Choose B sufficiently large that the $\ell \approx n + \log_2 B$ of the $f_i(v)$ that have values less than B in the neighborhood of v_0 due to the exponential accumulation of function values at v_0 are a fixed fraction, αn, of the total, but small enough that the probability of k ($k = 4$ or 5) of those αn $f_i(v)$ being less than B at points away from v_0 is essentially zero.

It might at first seem an impossible task to search the continuous unit interval for values of v at which the $f_i(v)$ accumulate, however we can greatly simplify the task by the following argument that reduces the search to a finite -- albeit large -- number of values of v. Given k weights from the hard knapsack and a bound $B \ll 1$, assume that at some point v' all k of the related functions satisfy $f(v') < B$. The inequalities will continue to hold for all $v'' \le v \le v'$, where v'' is the location of the closest minima below v' of any of the $f_i(v)$ since at v'' at least one of the functions has value zero and the others have decreased in value by $a_i(v'-v'')$.

$$v'' = \max_{i=1}^{k} \left\{ \frac{\lceil a_i v' \rceil}{a_i} \right\} .$$

The search for points at which the functions simultaneously have values less than B can therefore be restricted to those points at which one or more of the $f_i(v)$ have minima, i.e., rationals of the form

$$\frac{c_i}{a_i} \qquad \text{where} \qquad \begin{matrix} 0 \le c_i < a_i \\ 1 \le i \le k \end{matrix} .$$

Copyright (c) 1998, Springer-Verlag

As already remarked there are a_1 such minima of $f_1(v)$, however the total number of minima for a collection of functions may be significantly less than the sum of the associated a_1 since some points may be multiple minima.

Given a subset of k weights, A, define the functions S_j;

$$\left.\begin{aligned} S_1 &= \sum_{a_1 \in A} (a_1) - \binom{k}{1} \\ S_2 &= \sum_{\substack{a_1, a_j \in A \\ a_1 \neq a_j}} (a_1, a_j) - \binom{k}{2} \\ &\vdots \\ S_j &= \sum (\text{GCD of j element subsets of A}) - \binom{k}{j} \\ &\vdots \\ S_k &= (A) - \binom{k}{k} \end{aligned}\right\} \qquad (4)$$

If $\hat{a} = \max\{a_1\}$ then trivially $f_1(v) < B$ in the interval

$$\left(0 < v < \frac{B}{\hat{a}}\right) .$$

Rationals in this interval are solutions for the hard knapsack only in the (improbable) event that the $\{a_1\}$ are already a super increasing sequence, i.e., the "hard" knapsack is in fact trivial. We therefore exclude the open interval $(0 < v < \frac{1}{\hat{a}})$ at the origin in the following discussion.

Theorem 2

Given a k element subset A of the weights $\{a_1\}$, the number of points at which precisely j of the associated functions $f_1(v)$ have minima is given by

Copyright (c) 1998, Springer-Verlag

$$\left.\begin{aligned} N_j &= \sum_{i=0}^{k-j} (-1)^i \binom{j+i}{j} S_{j+i} \qquad 1 \le j < k \\ &\text{and the total number of minima by} \\ N_{min} &= \sum_{i=1}^{k} N_i = \sum_{i=1}^{k} (-1)^{i-1} S_i \end{aligned}\right\} . \tag{5}$$

For a randomly drawn set of k weights, the expected number of open intervals in which all k of the associated functions have values less than B is therefore

$$N_s = \sum_{i=1}^{k} N_i \cdot 2^{(k-i)\log_2 B} \tag{6}$$

on the assumption that those functions not equal to zero are uniformly -- and independently -- distributed on the unit interval. The detailed behavior of the N_i is an exceedingly difficult number theoretic question -- hinging as it does on the distribution of common divisors among the elements of $\{a_i\}$, however, the following argument shows that k is greater than 3 and is almost certainly no greater than five under our assumptions about the a_i.

For the purposes of illustration, let $B = 2^{-n+n/6}$, so that $\approx n/6$ of the $f_i(v_0)$ have values less than $2^{-5/6\,n}$ at a point away from v_0 and 0 is

$$P_f = \frac{N_s}{N_{min}} = \frac{1}{N_{min}} \sum_{i=1}^{k} N_i 2^{-5/6(k-i)n} . \tag{7}$$

If $N_1 \gg N_2 2^{-1\log_2 B}$ so that the points at which only one function has a minima exponentially dominate the others, then N_s can be approximated by

$$N_s \approx 2^{(k-1)\log_2 B} \sum_{i=1}^{k} a_i$$

or by

Copyright (c) 1998, Springer-Verlag

$$N_s \approx 2^{n/6(17-5k)+\log_2 k}$$

in the present example. $k = 3$ and $n \geq 100$ results in an intolerable number of false alarms, while $k = 4$ -- if the assumption about N_1 exponentially dominating the N_i -- results in

$$N_s \approx 2^{-n/2+2} = 3.6 \times 10^{-15}$$

for the case of $n = 100$, i.e., an exceedingly small number of false solution intervals. Our experience with Monte Carlo computations bears out this expectation for randomly generated Merkle-Hellman knapsacks.

To summarize, what we have shown is that if the bound is taken to be $B = 2^{-5/6\,n}$, so that $n/6$ of the $f_i(v_0)$ are less than B, then the probability that even four randomly chosen $f_i(v)$ will be less than B at a point away from v_0 is very small.

It is worth remarking that choosing B so as to cause αn of the $f_i(v_0)$ to be less than B avoids the $O(n^4)$ search problem of Adi's original proposal. Since the method described here will be satisfied by any subset of k out of the αn a_i corresponding to the $f_i(v_0) < B$, the probability of hitting such a subset at random is

$$P_h = \frac{\binom{\alpha n}{k}}{\binom{n}{k}} = \alpha^k \prod_{i=1}^{k-1} \frac{(n-i\alpha^{-1})}{(n-i)} \quad . \tag{8}$$

If $k = 4$ and $\alpha = 1/6$, then for all $n \geq 52$ the probability of hitting a subset of four of the desired a_i is

$$P_h \geq \frac{1}{2 \cdot 6^4}$$

or conversely the expected number of sets of four weights that need to be tried before a success is less than 2592. If $k = 4$ and $\alpha = 1/10$, then for all $n \geq 91$ the upper bound on the expected number of tries increases to 20,000.

Copyright (c) 1998, Springer-Verlag

With these preliminaries out of the way, we are now ready to describe the algorithm by which we propose to localize v_0. Given a set of four a_i, it has already been shown that it is only necessary to consider those points $\frac{c_i}{a_i}$, $0 \leq c_i < a_i$, at which a minima of at least one of the $f_i(v)$ occurs. The values of the other functions at these minima are

$$f_j\left(\frac{c_i}{a_i}\right) = \left(\frac{c_i}{a_i}\right) a_j - \left[\frac{c_i}{a_i} a_j\right] = \frac{a_j}{a_i} c_i - c_j \quad . \tag{9}$$

Hence, the inequalities that need to be satisfied in order for all four functions to have values less than the bound B are:

$$\left.\begin{aligned} 0 \leq \frac{c_i}{a_i} - \frac{c_j}{a_j} \leq \frac{B}{a_j} \\ 0 \leq \frac{c_i}{a_i} - \frac{c_k}{a_k} \leq \frac{B}{a_k} \\ 0 \leq \frac{c_i}{a_i} - \frac{c_l}{a_l} \leq \frac{B}{a_l} \end{aligned}\right\} \quad . \tag{10}$$

As we have already remarked, inequalities (10) are very similar in appearance to Adi's system (1). In fact, if we replace B by the tightest possible upper bound for the functions $f_i(v_0)$, i = 2, 3 and 4 and assume that the corresponding a_i are as large as possible we get precisely system (1). The important point is not that under these extreme conditions (10) is equivalent to (1), but rather that (10) describes a search (on the vertical axis) for accumulations of values of the $f_i(v)$ rather than a search (on the horizontal axis) for accumulations of minima of the $f_i(v)$.

We use the same substitution of values for the $\frac{c_i}{a_i}$ used earlier in obtaining the system (1') from (1) to obtain a system of inequalities equivalent to (10):

$$\left.\begin{aligned} 0 \leq \frac{b_j}{a_j} - \frac{b_i}{a_i} \leq \frac{M_0}{a_j} B \\ 0 \leq \frac{b_k}{a_k} - \frac{b_i}{a_i} \leq \frac{M_0}{a_k} B \\ 0 \leq \frac{b_l}{a_l} - \frac{b_i}{a_i} \leq \frac{M_0}{a_l} B \end{aligned}\right\} \quad . \tag{10'}$$

Copyright (c) 1998, Springer-Verlag

When put in this form (10'), it is easy to see why systems (1') and (10') behave differently. In (10'), if an a_j, a_k or a_ℓ is caused to be smaller than expected, which would have caused the corresponding inequality in (1') to fail to have a solution, the right-hand bound grows proportionally since the corresponding weight a appears in the denominator.

The weight a_i -- whose associated function $f_i(v)$ was assumed to have a minima at the test points -- is distinguished in (10), so that for each choice of four weights, there are four systems of the form of (10) whose solution must be attempted: letting each weight take the role of a_i in turn. Setting $B = 2^{-n+an}$ and multiplying the inequalities in (10) through by $2^n a_i a_j$, etc., we obtain the equivalent system --

$$\left.\begin{aligned} 0 \le 2^n a_j c_i - 2^n a_i c_j \le a_i 2^{an} \\ 0 \le 2^n a_k c_i - 2^n a_i c_k \le a_i 2^{an} \\ 0 \le 2^n a\ c_i - 2^n a_i c \le a_i 2^{an} \end{aligned}\right\} \qquad (10')$$

in which the coefficients are integers with no more than 5n bits -- which can be solved in precisely the same way and with the same computational difficulty as in Adi's proposal.

All other aspects of this algorithm are the same as in Adi's method -- only the means of localizing the regions in which to test for solutions has been changed. The reader is referred to Adi's paper for the essential details involved in subdividing intervals and testing for super increasing sequences that satisfy the unique invertibility condition.

CONCLUSION

The reason that we have titled this paper "A Preliminary Report on the Cryptanalysis of Merkle-Hellman Knapsack Cryptosystems," is primarily our concern that just as the knapsack designer could elude cryptanalysis by Adi's original algorithm, that he may also be able

Copyright (c) 1998, Springer-Verlag

to elude the alternative algorithm described here. Our investigation of this question has only just begun. However, there are several worrisome areas. For example, in the analysis of N_s and P_f, we used the assumption that the a_i were randomly distributed. The designer may be able to force the values of the a_i in such a way as to invalidate this assumption and hence to invalidate statements we have made about k, P_f, N_s, N_{min}, N_j, etc., that were true for randomly constructed $\{a_i\}$. Similarly, in (10) and (10'), there is an implicit assumption -- in the discussion of the solution to these systems which we bypassed by referring the reader to Adi's paper -- that not many of the a_i's corresponding to the αn, $f_i(v_0) < B$ are much smaller than expected -- in the as yet unproven expectation that it will be difficult for the designer to cause this assumption to fail and to at the same time make the hard knapsack still be "hard" against direct solution.

Finally, we conclude with the observation (caution) that it is extremely dangerous to impute properties to a deliberately designed cryptosystem -- even though these properties can be shown to hold with very high probability for almost all randomly chosen cryptosystems based on the same algorithm.

REFERENCES

1. Shamir, Adi, "A Polynomial Time Algorithm for Breaking Merkle-Hellman Cryptosystems," (extended abstract) 7 p. research announcement, April, 1982.

2. Merkle, R. and M. Hellman, "Hiding Information and Receipts in Trapdoor Knapsacks," IEEE Transactions on Information Theory IT-24 (1978), p. 525-530.

Copyright (c) 1998, Springer-Verlag

Copyright (c) 1998, Springer-Verlag