Smart Hill Climbing Finds Better Boolean Functions

William Millan, Andrew Clark and Ed Dawson
Information Security Research Centre
Queensland University of Technology
GPO Box 2434, Brisbane, Queensland, Australia 4001
Facsimile: 61-7-3221 2384
Email: {millan,aclark,dawson}@fit.qut.edu.au

Abstract

Block and stream ciphers are made from Boolean functions that usually require a compromise between several conflicting cryptographic criteria. Although some constructions exist to generate Boolean functions satisfying one or more criteria, such as balance and high nonlinearity, there are often drawbacks to them such as low nonlinear order. In this paper we present a new algorithm for simple modification of a Boolean function truth table to improve both nonlinearity and balance. We also show how to modify a balanced function in two truth table positions so that the nonlinearity is increased and the balance is maintained. When the algorithm fails to find an improvement, one does not exist, and we have then identified a locally maximum function. We present results comparing the probability distributions of random functions with that of locally maximum functions found by our algorithms, and also comment on how the number of steps required to find a local maximum is affected by increasing the number of variables.

1 About Boolean Functions

Let $f(x)$ denote the binary truth table ($f(x) \in \{0,1\}$) and $\hat{f}(x)$ the corresponding *polarity truth table*, $\hat{f}(x) \in \{1,-1\}$. We have $\hat{f}(x) = (-1)^{f(x)} = 1 - 2f(x)$. The Hamming weight of a Boolean function is the number of ones in the binary truth table, or equivalently the number of -1s in the polarity truth table. A balanced function has the same number of zeroes and ones in the truth table. Balance is a primary cryptographic criterion: an imbalanced function has sub-optimum unconditional entropy (ie. it is correlated to a constant function). We define the imbalance of a Boolean function as $I_f = \sum_x \hat{f}(x)$. The correlation between a function and the constant zero function is simply $\frac{I_f}{2^n}$, which is a value between -1 and 1. A function with zero imbalance is balanced and has no correlation to the constant functions.

Every function has a unique representation in the Algebraic Normal Form (ANF) as the binary coefficient vector of a fixed (positive) polarity Reed-Muller expansion (for example, see [5]). The ANF describes a two level circuit: an XOR sum of AND products. The *nonlinear order* or just *order* of a Boolean function is the size of the largest product term in the ANF. Order zero functions are constant, affine functions have order 1, and linear functions are those affine functions with a zero constant term in their ANF. The exclusive-or operation is linear; a linear function is an XOR sum of variables. We may specify a linear function by an n-bit vector ω that selects the variables in this sum: $L_\omega(x) = \omega_1 x_1 \oplus \cdots \oplus \omega_n x_n$.

The Hamming distance to linear functions is an important cryptographic property, since ciphers that employ nearly linear functions can be broken easily by a variety of methods (for example see [7, 4]). In particular, both differential and linear cryptanalysis techniques [2, 8] are resisted by highly nonlinear functions. Thus the minimum distance to any affine function is an important indicator of the cryptographic strength of a Boolean function. The *nonlinearity* of a Boolean function is this minimum distance, or the distance to the set of affine functions. We note that complementing the output will not change the nonlinearity of any Boolean function, so we need to consider the magnitude of the correlation to all linear functions, of which there are 2^n.

The Hamming distance between a pair of functions can be determined by evaluating both functions for all inputs and counting the disagreements. This process has complexity $O(2^n)$. It follows that determining the nonlinearity in this naive fashion will require $O(2^{2n})$ function evaluations, which is infeasible even for small n. However, a tool exists that enables the calculation of all linear correlation coefficients in $O(n2^n)$ operations. This is the fast Walsh-Hadamard Transform, and its uses in cryptography and elsewhere are well known [1, 13].

Let $\hat{F}(\omega)$ denote the Walsh-Hadamard Transform (WHT) of a Boolean function. Its calculation is defined as $\hat{F}(\omega) = \sum_x \hat{f}(x)\hat{L}_\omega(x)$. It is clear from this definition that the value of $\hat{F}(\omega)$ is closely related to the Hamming distance between $f(x)$ and the linear function $L_\omega(x)$. In fact the correlation to the linear function is given by $c(f, L_\omega) = \frac{\hat{F}(\omega)}{2^n}$. The nonlinearity N_f of $f(x)$ is related to the maximum magnitude of WHT values WH_{max}, by $N_f = \frac{1}{2} * (2^n - WH_{max})$. Clearly in order to increase the nonlinearity, we must decrease WH_{max}. A function is uncorrelated with linear function $L_\omega(x)$ when $\hat{F}(\omega) = 0$. We would like to find a Boolean function that has all WHT values equal to zero, since such a function has no correlation to any affine function. However, it is known [9] that such functions do not exist. A well known theorem, widely attributed to Parseval [6], states that the sum of the squares of the WHT values is the same constant for every Boolean function: $\sum_\omega \hat{F}^2(\omega) = 2^{2n}$. Thus a tradeoff exists in minimising affine correlation. When we alter a function so that its correlation to some affine function is reduced, the correlation to some other affine function is increased.

It is known that the Bent functions [12] satisfy the property that $|\hat{F}(\omega)| = 2^{\frac{n}{2}}$ for all ω. Bent functions exist only for even n, and they attain the maximum possible nonlinearity of $N_{bent} = 2^{n-1} - 2^{\frac{n}{2}-1}$. It is an open problem to determine an expression for the maximum nonlinearity of functions with an odd number of inputs. It is known that, for n odd, it is possible to construct a function with nonlinearity $2^{n-1} - 2^{\frac{n-1}{2}}$ by concatenating Bent functions. It is known that for $n = 3, 5, 7$ that this is in fact the upper bound of nonlinearity. The only value of n for which it is known that this value is not the upper bound is $n = 15$ [10, 11]. We note that determining the covering radius of a Reed-Muller code is the same problem as finding an upper bound on low order approximation. It is a well known open problem to find the covering radius of Reed-Muller codes, so it is not known to what extent functions may resist low order approximations.

There seem to be no formal tools for low order nonlinear approximation, so we leave this difficult area, and instead concentrate on improving the nonlinearity of Boolean functions in a systematic way.

In this paper, we present algorithms that provide a list of truth table positions that, if complemented, will result in a Boolean function with higher nonlinearity. The approach is based on the observation that small changes to a truth table result in small magnitude changes to the WHT values. In particular, a single truth table complementation will cause every $\hat{F}(\omega)$ to alter by ± 2. Two truth table changes will cause $\Delta\hat{F}(\omega) \in \{-4, 0, 4\}$. We use these facts in the next section to prove conditions required for small changes to increase nonlinearity. When two changes are made, the Hamming weight can be maintained while nonlinearity is increased.

These techniques provide a fast way of hill-climbing the Boolean function terrain to locate highly nonlinear Boolean functions that would be difficult to obtain by a purely random search or exhaustive hill climbing.

2 Improving Nonlinearity

Consider altering a function $f(x)$ by complementing the output for a single input x_1, with the nonlinearity increasing. We define the 1-Improvement Set of $f(x)$, 1-IS_f , as the set of all inputs such that complementing the corresponding output of any one of them will increase the nonlinearity of the function.

Definition 1 *Let $g(x) = f(x) \oplus 1$ for $x = x_1$ and $g(x) = f(x)$ for all other x. If $N_g > N_f$ then $x_1 \in$ 1-IS_f.* □

If 1-IS_f is empty, the function is a 1-local maximum for nonlinearity. Of course all Bent functions are global maxima, so their 1-Improvement Sets are empty. There also exist sub-optimum local maxima that will be found by hill climbing algorithms. It is computationally intensive to exhaustively alter truth table positions, find new WHTs and so determine the set 1-IS_f, so we seek a fast, systematic way to determine the 1-Improvement Set of a given Boolean function from its truth table and Walsh-Hadamard transform. In this section we present easily checked conditions for an input x to be in the 1-Improvement Set.

Definition 2 *Let $f(x)$ be a Boolean function with Walsh-Hadamard Transform $\hat{F}(\omega)$. Let WH_{max} denote the maximum absolute value of $\hat{F}(\omega)$. There will exist one or more linear functions $L_\omega(x)$ that have minimum distance to $f(x)$, and $|\hat{F}(\omega)| = WH_{max}$ for these ω. Let us define the following sets:*

$$\begin{aligned} W_1^+ &= \{\omega : \hat{F}(\omega) = WH_{max}\} \text{ and} \\ W_1^- &= \{\omega : \hat{F}(\omega) = -WH_{max}\}. \end{aligned}$$

We also need to define sets of ω for which the WHT magnitude is close to the maximum.

$$\begin{aligned} W_2^+ &= \{\omega : \hat{F}(\omega) = WH_{max} - 2\}, \\ W_2^- &= \{\omega : \hat{F}(\omega) = -(WH_{max} - 2)\}, \\ W_3^+ &= \{\omega : \hat{F}(\omega) = WH_{max} - 4\}, \text{ and} \\ W_3^- &= \{\omega : \hat{F}(\omega) = -(WH_{max} - 4)\}. \end{aligned}$$

□

When a truth table is changed in exactly one place, all WHT values are changed by +2 or -2. It follows that in order to increase the nonlinearity we need to make the WHT values in set W_1^+ change by -2, the WHT values in set W_1^- change by +2, and also make the WHT values in set W_2^+ change by -2 and the WHT values in set W_2^- change by +2. The first two conditions are obvious, and the second two conditions are required so that all other $|\hat{F}(\omega)|$ remain less than WH_{max}. These conditions can be translated into simple tests.

Theorem 1 *Given a Boolean function $f(x)$ with WHT $\hat{F}(\omega)$, we define sets $W^+ = W_1^+ \cup W_2^+$ and $W^- = W_1^- \cup W_2^-$. For an input x to be an element of the Improvement Set, all of the following conditions must be satisfied.*

(i) $f(x) = L_\omega(x)$ for all $\omega \in W^+$

and

(ii) $f(x) \neq L_\omega(x)$ *for all* $\omega \in W^-$.

If the function $f(x)$ *is not balanced, and we wish to reduce the imbalance, we impose the additional restriction that*

(iii) when $\hat{F}(0) > 0$, $f(x) = 0$, *else* $f(x) = 1$. □

Proof: We start by considering the conditions to make WHT values change by a desired amount. When $\hat{F}(\omega)$ is positive, there are more 1 than -1 in the polarity truth table, and more 0 than 1 in the binary truth table of $f(x) \oplus L_\omega(x)$. It follows that to make $\Delta\hat{F}(\omega) = -2$, we must change any single 0 to 1 in the truth table of $f(x) \oplus L_\omega(x)$. This means that we select an x to change such that $f(x) = L_\omega(x)$. We desire a -2 change for all WHT values with $\omega \in W^+$, so this proves condition (i). A similar argument proves condition (ii). A function is balanced when $\hat{F}(0) = 0$, so to reduce the imbalance we must select x according to condition (iii). □

We often seek to improve the nonlinearity of balanced Boolean functions, while retaining balance. Clearly this requires an even number of truth table changes. We now present the conditions on a pair of inputs x_1, x_2 so that complementing both their function values causes an increase in nonlinearity, without changing the Hamming weight. We define the 2-Improvement Set, 2-IS_f, as the set of all such input pairs. A function for which no pair satisfies these conditions is said to be a 2-local maximum.

Theorem 2 *Given a Boolean function* $f(x)$ *with WHT* $\hat{F}(\omega)$, *we define sets* $W_1 = W_1^+ \cup W_1^-$, $W_{2,3}^+ = W_2^+ \cup W_3^+$ *and* $W_{2,3}^- = W_2^- \cup W_3^-$. *A pair of inputs* (x_1, x_2) *is in the 2-Improvement Set of* $f(x)$ *if and only if all of the following conditions are satisfied:*

(i) $f(x_1) \neq f(x_2)$

(ii) $L_\omega(x_1) \neq L_\omega(x_2)$ *for all* $\omega \in W_1$

(iii) $f(x_i) = L_\omega(x_i)$, $i \in \{1,2\}$, *for all* $\omega \in W_1^+$

(iv) $f(x_i) \neq L_\omega(x_i)$, $i \in \{1,2\}$, *for all* $\omega \in W_1^-$

(v) for all $\omega \in W_{2,3}^+$, *if* $L_\omega(x_1) \neq L_\omega(x_2)$ *then* $f(x_i) = L_\omega(x_i)$, $i \in \{1,2\}$

(vi) for all $\omega \in W_{2,3}^-$, if $L_\omega(x_1) \neq L_\omega(x_2)$ then $f(x_i) \neq L_\omega(x_i)$, $i \in \{1,2\}$ □

Proof: Condition (i) is required to maintain the Hamming weight. Conditions (ii),(iii) and (iv) are proven similarly to theorem 1. In order to stop the correlation to other linear functions increasing too much, we require that $\Delta\hat{F}(\omega) \neq +4$, for all $\omega \in W_{2,3}^+$, and it follows that not both of $f(x_i) \oplus L_\omega(x_i) = 1$, or equivalently that at least one of $f(x_i) \oplus L_\omega(x_i) = 0$. Consequently,

$$[f(x_1) \oplus L_\omega(x_1)][f(x_2) \oplus L_\omega(x_2)] = 0,$$

and expanding this, noting from (i) that $f(x_1)f(x_2) = 0$, we have

$$f(x_1)L_\omega(x_2) \oplus f(x_2)L_\omega(x_1) \oplus L_\omega(x_1)L_\omega(x_2) = 0.$$

We need to consider four cases to find the exact conditions for this expression to be satisfied:

(a) When $L_\omega(x_1) = L_\omega(x_2) = 0$ the expression is satisfied and no further conditions on (x_1, x_2) are required.

(b) When $L_\omega(x_1) = L_\omega(x_2) = 1$ the expression becomes $f(x_1) \oplus f(x_2) = 1$ which is equivalent to condition (i).

(c) When $L_\omega(x_1) = 0$ and $L_\omega(x_2) = 1$ the expression becomes $f(x_1) = 0$.

(d) When $L_\omega(x_1) = 1$ and $L_\omega(x_2) = 0$ the expression becomes $f(x_2) = 0$.

Combining (a)-(d) we see that when $L_\omega(x_1) \neq L_\omega(x_2)$ for $\omega \in W_{2,3}^+$ we require that $f(x_i) = L_\omega(x_i)$, $i = 1, 2$, thus proving condition (v). The Proof of (vi) is similar. □

The following theorem shows how to modify the WHT of a Boolean function that has been altered in a single truth table position, with complexity $O(2^n)$. We note that the algorithm for incremental improvement of Boolean functions suggested in [3] recomputes the WHT after every single bit change regardless of whether that change improves the nonlinearity. Our algorithms are superior on two counts - every change is an improvement and the new WHT is found n times faster.

Theorem 3 *Let $g(x)$ be obtained from $f(x)$ by complementing the output for a single input, x_1. Then each component of the WHT of $g(x)$, $\hat{G}(\omega) = \hat{F}(\omega) +$*

$\Delta(\omega)$, *can be obtained as follows: If* $f(x_1) = L_\omega(x_1)$, *then* $\Delta(\omega) = -2$, *else* $\Delta(\omega) = +2$

Proof: When $f(x_1) = L_\omega(x)$, we have $(-1)^{f(x_1) \oplus L_\omega(x_1)} = 1$, which contributes to the sum in $\hat{F}(x_1)$. Changing the value of $f(x_1)$ changes this contribution to -1, so $\Delta\hat{F}(\omega) = -2$. Similarly when $f(x_1) \neq L_\omega(x)$, $\Delta\hat{F}(\omega) = +2$. $\square$

3 Implementation and Results

In this section the implementation details for the one step improvement and two step improvement algorithms are given - **HillClimb** and **HillClimb2**. We note that condition (ii) of Theorem 2 is redundant, and is not referred to in the implementation of that algorithm.

- **HillClimb(BF, WHT)**
  1. Determine maximum value of the Walsh-Hadamard transform $\mathrm{WH}_{\mathrm{max}}$.
  2. By parsing the WHT find the values of ω which belong to the sets W_1^+, W_1^-, W_2^+ and W_2^-. At the completion of this step there should be two lists: $W^+ = W_1^+ \cup W_2^+$ and $W^- = W_1^- \cup W_2^-$. NB. Either (but not both) of W^+ and W^- may be empty.
  3. For i in $0 \ldots 2^n - 1$, do
     (a) Let b_i denote the i^{th} bit in the truth table of BF.
     (b) Parse the sets W^+ and W^- ensuring that conditions (iii) and (iv) in Theorem 2 are satisfied - if not skip to Step 3e.
     (c) We have a candidate for improvement. Complement b_i in the truth table of BF (denote the resulting boolean function BF'), update the WHT (becoming WHT') by using Theorem 3, and call **HillClimb(BF', WHT')**.
     (d) Skip to Step 4.
     (e) Increment i $(i = i + 1)$.
  4. BF represents a 1-local maximum - terminate processing.

- **HillClimb2(BF, WHT)**
  1. Determine maximum value of the WHT WH_{max}.
  2. By parsing the WHT obtain the sets W_1^+, W_1^-, $W_{2,3}^+$ and $W_{2,3}^-$.
  3. For i in $0 \ldots 2^n - 1$, do
     (a) Let b_i denote the i^{th} bit in the truth table of BF.
     (b) Parse the sets W^+ and W^- ensuring that conditions (iii) and (iv) in Theorem 2 are satisfied - if they are add i to c_{b_i}.
  4. For each element of c_0, do
     (a) For each element of c_1, do
        i. Check conditions (v) and (vi) of Theorem 2 and if they are satisfied complement the corresponding bits in the truth table (call the resulting truth table BF'), find the adjusted WHT (becoming WHT') by applying Theorem 3 twice and call **Hill-Climb2(BF', WHT')**. Skip to Step 5.
  5. BF represents a 2-local maximum - terminate processing.

We now present examples of the distribution of nonlinearity for random functions and random balanced functions, compared with the nonlinearity of locally maximum functions obtained by our two algorithms. In Figures 1 and 2 we compare random functions with the maxima found by **HillClimb**, for 8 and 12 input variables respectively. From these graphs it is clear that random functions have a smooth, bell-shaped distribution, whereas 1-local maxima are much more likely to have an even value for nonlinearity. It is also clear that hill climbing will find highly nonlinear functions much more easily than random search.

Figures 3 and 4 illustrate the performance of random generation versus hill climbing, when confined to balanced functions only, for functions with 8 and 12 input variables respectively. Note that **HillClimb2** conserves Hamming weight. For our tests we started with balanced functions so that the 2-local maxima were also balanced. This allows a direct comparison with randomly generated balanced functions. It is easy to show that the nonlinearity of a

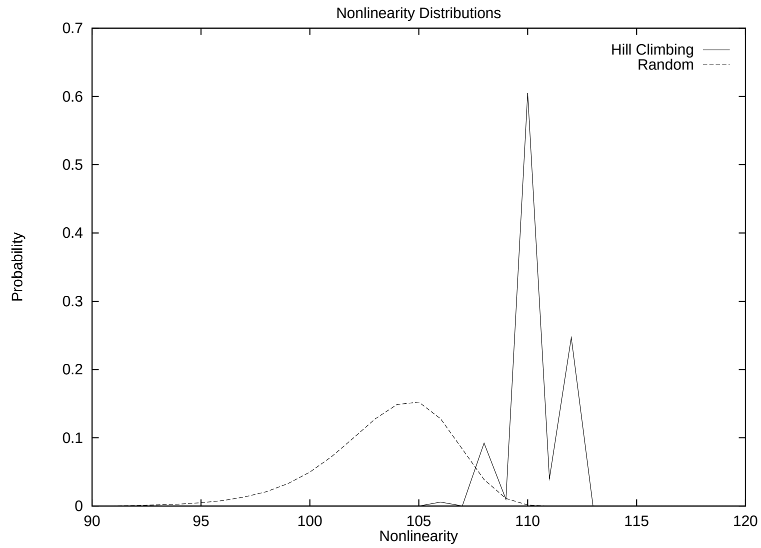

Figure 1: A Comparison of Hill Climbing with Random Generation, n=8.

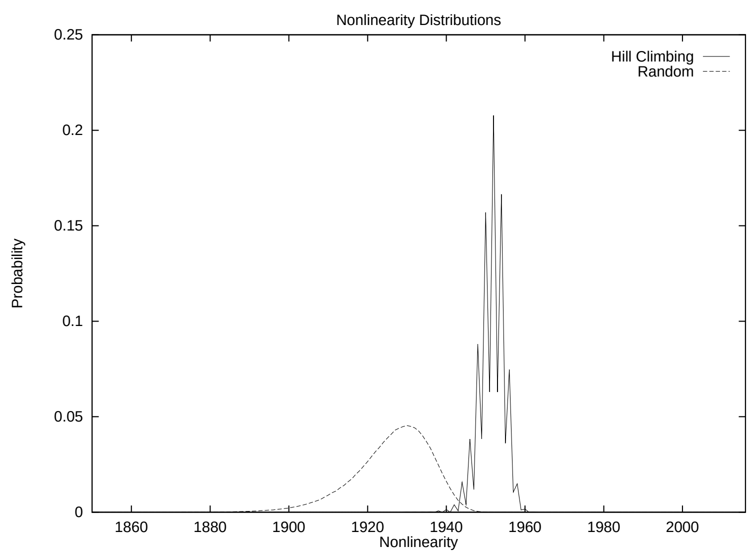

Figure 2: A Comparison of Hill Climbing with Random Generation, n=12.

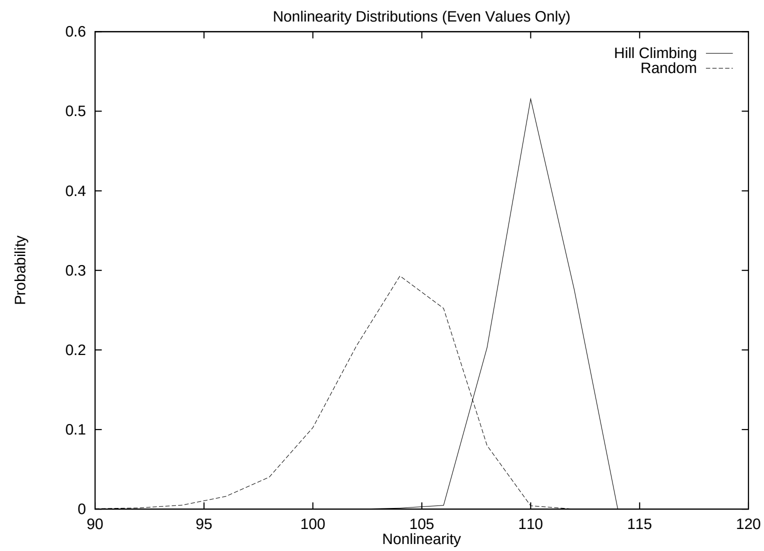

Figure 3: A Comparison of Hill Climbing with Random Balanced Generation, n=8.

balanced function is always even. For simplicity we only show results for even values of the nonlinearity in these graphs.

Figure 5 shows how the average number of steps to find a local maximum is changing with the number of variables. These results suggest that as n increases the distance from a random function to a local maximum is increasing in an exponential-like manner. The implication of this is that these hill climbing algorithms will be more effective for large n. It follows from Theorem 3 that making n hill climbing steps can be done in approximately the same time as a single random generation and complete fast WHT. The relative efficiency of the hill climbing algorithm improves as n increases.

4 Conclusion

We have presented two useful algorithms for the improvement of Boolean functions. With these tools it is now feasible to perform hill climbing to obtain locally maximum functions. In conjunction with heuristic search methods, these tools provide a means to find strong Boolean functions for cryptographic applications.

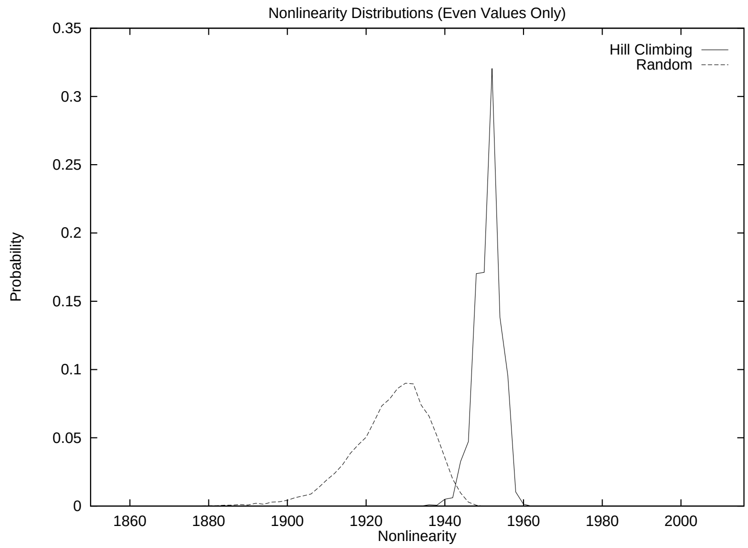

Figure 4: A Comparison of Hill Climbing with Random Balanced Generation, n=12.

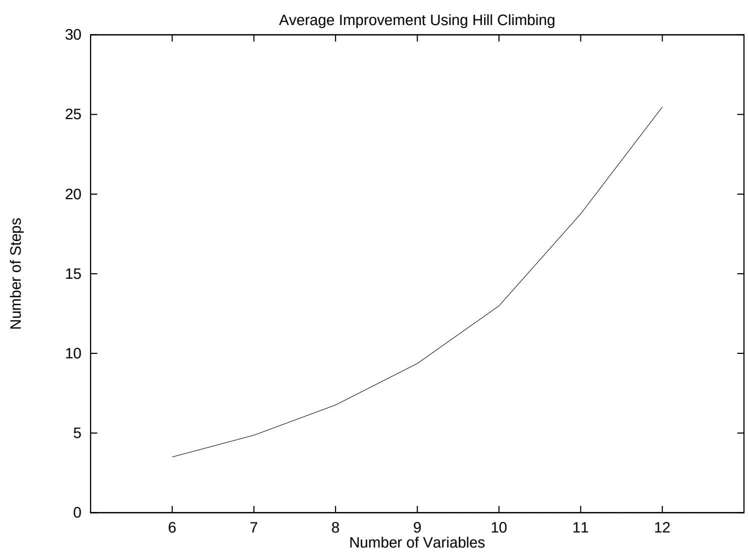

Figure 5: Average Distance To Local Maxima For Various n.

Several open problems remain. One is to determine the general relationship between the nonlinearity of a Boolean function and the size of its Improvement Set. Clearly local maxima have empty sets and local minima have full sets, but for arbitrary functions the relationship is not clear. Initial experiments have shown that two functions with the same nonlinearity can have different sized Improvement Sets, and the results suggest that the function closer to a local maximum has a smaller Improvement Set, so that in a hill climbing algorithm it may be of benefit to maintain as large a set as possible, thus avoiding grossly sub-optimum local maxima.

Smart hill climbing may be adapted to generate Boolean functions satisfying other cryptographic criteria, and to improve the nonlinearity of bijective S-boxes. These topics are the subject on ongoing research.

References

[1] K.G. Beauchamp. *Applications of Walsh and Related Functions*. Academic Press, 1984.

[2] E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems. In *Advances in Cryptology - Crypto '90, Proceedings, LNCS*, volume 537, pages 2–21. Springer-Verlag, 1991.

[3] R. Forre. Methods and Instruments for Designing S-Boxes. *Journal of Cryptology*, 2(3):115–130, 1990.

[4] J.Dj. Golic. Linear Cryptanalysis of Stream Ciphers. In *Fast Software Encryption, 1994 Leuven Workshop, LNCS*, volume 1008, pages 154–169, December 1994.

[5] D.H. Green and G.A. Khuwaja. Simplification of switching functions expressed in Reed-Muller algebraic form. *IEE Proceedings, Pt E.*, 139(6):511–518, November 1992.

[6] F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error Correcting Codes*. North-Holland Publishing Company, Amsterdam, 1978.

[7] J.L. Massey. Shift-Register Synthesis and BCH Decoding. *IEEE Transactions on Information Theory*, 15(1):122–127, January 1969.

[8] M. Matsui. Linear Cryptanalysis Method for DES Cipher. In *Advances in Cryptology - Eurocrypt '93, Proceedings, LNCS*, volume 765, pages 386–397. Springer-Verlag, 1994.

[9] W. Meier and O. Staffelbach. Nonlinearity Criteria for Cryptographic Functions. In *Advances in Cryptology - Eurocrypt '89, Proceedings, LNCS*, volume 434, pages 549–562. Springer-Verlag, 1990.

[10] N.J. Patterson and D.H. Wiedemann. The Covering Radius of the $(2^{15}, 16)$ Reed-Muller Code is at least 16276. *IEEE Transactions on Information Theory*, 29(3):354–356, May 1983.

[11] N.J. Patterson and D.H. Wiedemann. Correction to 'the covering radius of the $(2^{15}, 16)$ Reed-Muller code is at least 16276'. *IEEE Transactions on Information Theory*, 36(2):443, March 1990.

[12] O.S. Rothaus. On Bent Functions. *Journal of Combinatorial Theory (A)*, 20:300–305, 1976.

[13] G-Z. Xiao and J.L. Massey. A Spectral Characterization of Correlation-Immune Combining Functions. *IEEE Transactions on Information Theory*, 34(3):569–571, May 1988.