УДК 519.14: 004.272.26

Д.А. Никитин, К.В. Дьяконов

О распределении значений нелинейности булевых функций[1]

Описываются результаты работы по определению возможных значений нелинейности булевых функций многих переменных с использованием вычислений на кластере. Результаты могут быть полезны для алгебры и криптографии, в частности, могут применяться для выбора узлов замен блочных шифров с заданной величиной нелинейности.

Для достижения устойчивости шифра к линейному и дифференциальному криптоанализу необходимо, чтобы все преобразования бит в ходе шифрования были, во-первых, как можно более нелинейными (то есть далеко отстоящими в смысле Хэмминга от множества аффинных булевых функций), во-вторых, обладали лавинным эффектом. Количественными характеристиками, выражающими устойчивость шифрования к этим методам криптоанализа, являются нелинейность и динамическое расстояние, определённые для процедуры преобразования бит [1].

При выборе узлов замен полезно знать максимальное значение нелинейности, достижимое при замене группы бит заданной длины. Фактически узел замены блочного шифра – это булева функция от определённого количества переменных. Пока не существует аналитического способа определения возможных значений нелинейности булевых функций от определённого числа переменных N, либо нелинейности конкретной булевой функции. Поэтому приходится вычислять нелинейность, основываясь на его определении, то есть практически перебором: вычисляется расстояние Хэмминга каждой функции от всех линейных функций того же числа переменных, и затем выбирается наименьшее из полученных расстояний. При выборе узла замены блочного шифра, таким образом, приходится осуществлять подобный перебор для всех нелинейных булевых функций N переменных.

Данная работа преследует следующие цели:

- разработать инструменты для эффективного вычисления нелинейности функции от большого числа переменных;
- разработать инструменты для эффективного нахождения значений нелинейностей большого массива функций;
- набрать статистику о распределении функций с высоким значением нелинейности по всему множеству функций;
- выработать рекомендации по осуществлению более эффективного поиска максимально нелинейных функций.

Вычислительная сложность данной задачи не позволяет решать её за приемлемое время с использованием персонального компьютера. В частности общее количество булевых функций от N переменных равно 2^{2^N}, а размер одной функции (хранимой в виде её таблицы истинности) – 2^N бит, то есть 2^{N-3} байт. Например, для N=32 размер одной функции – 512 Мбайт. В настоящее время известны возможные значения нелинейности для функций от 15 переменных и менее [2]. Для функций от большего числа переменных известны оценки сверху и снизу для значения максимальной нелинейности, а также имеется известный

[1] Работа выполнена в рамках ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009-2013 г. (контракт П1032 от 27.05.2010).

максимум нелинейности для некоторых N. Поскольку для вычислений необходимо всё множество аффинных функций, то важен размер таблицы истинности всего такого множества (табл. 1). При этом с точки зрения практической криптографии интересны функции от 8 булевых переменных, 16-ти, 32-х и более.

Таблица 1. Зависимость размера таблицы истинности множества всех аффинных булевых функций и от количества булевых переменных.

N	4	5	6	7	8	9	10	11	16	17	18
размер	256 бит	1024 бит	512 байт	2 Кб	8 Кб	32 Кб	128 Кб	512 Кб	512 Мб	2048 Мб	8 Гб

Для нахождения значений нелинейности наиболее разумным предполагается использовать вычисления на кластере, так как характер описанного алгоритма переборный – и он может быть подвергнут распараллеливанию. В нашей работе были использованы вычислительные ресурсы СФУ. Исходя из характеристик предоставленных ресурсов (14 узлов по 8 Гб ОЗУ и 8 ядер на каждом) и приведённых выше таблиц, были определены допустимые значения N, для которых все необходимые для решения данные смогут поместиться в оперативной памяти узла.

Например, для определения всех возможных значений нелинейности для всех функций от заданного количества переменных необходимо постоянное нахождение в памяти всего массива аффинных функций. Поэтому полный перебор возможен для функций от 18 переменных.

В случае же определения значения нелинейности одной функции, необходимо чтобы она была постоянно в памяти, а линейные функции можно генерировать по ходу проверки. В этом случае размер необходимой памяти определяется суммарным размером двух функций. Отсюда следует, что вычисление нелинейности одной функции с учётом ограничения размера оперативной памяти в 8 Гб возможна для функций от 35 переменных (табл. 2).

Таблица 2. Необходимый объём памяти при вычислении нелинейности одной функции.

N	28	29	30	31	32	33	34	35	36
размер	64 Мб	128 Мб	256 Мб	512 Мб	1Гб	2Гб	4Гб	8Гб	16Гб

Программа была разработана с использованием системы MC#, разработанной в Институте программных систем РАН [3, 4]. Данная система является надстройкой к широко распространённому языку C# и предназначена для написания параллельных и распределённых программ, в том числе и исполняемых на графических процессорах.

Язык программирования MC# основан на модели асинхронного параллельного программирования, впервые введенной в языке Polyphonic C# [5].

Разработанная программа может работать в 4-х режимах:

1) полный перебор всех булевых функций от указанного количества переменных и нахождение значений нелинейности для них;
2) случайный перебор булевых функций и нахождение значений нелинейности для них;
3) нахождение значения нелинейности заданной булевой функции;
4) генерация случайной булевой функции от указанного количества переменных.

Были получены следующие результаты.

Во-первых, было посчитано количество функций с каждым значением нелинейности для N=3-5. Результаты сведены в табл. 3:

Таблица 3. Количество функций с различной нелинейностью для N=3, 4, 5.

Нелинейность	N=3	N=4	N=5
0	16	32	64
1	128	512	2048
2	112	3840	31744
3		17920	317440
4		28000	2301440
5		14336	12888064
6		896	57996288
7			215414784
8			647666880
9			1362452480
10			1412100096
11			556408832
12			27387136

Из неё видно, что хотя с ростом N доля функций с максимальной нелинейностью падает, но всё же их количество довольно велико.

Далее было исследовано, как распределены такие функции по всему множеству функций. Для этого каждая функция представлялась в виде целого числа из диапазона $0\ldots2^N-1$. Такое представление вытекает из хранения функции в памяти: каждая функция хранится в виде её таблицы истинности, где каждая ячейка таблицы – 0 или 1, т.е. один бит. Например, булева функция, принимающая значение 0 при любых значениях аргументов, будет в этом случае представлена в виде числа 0. Распределение анализировалось двумя способами: графически и в табличном виде. Для сокращения размера таблиц использовалось агрегирование на интервалах: то есть всё множество функций, представленное в виде диапазона целых чисел $0\ldots2^{2^N}-1$, разбивалось на равные поддиапазоны, и затем анализировалось количество нелинейных функций, попавших в каждый диапазон.

Для N=3 распределение в графическом виде выглядит следующим образом (рис. 1):

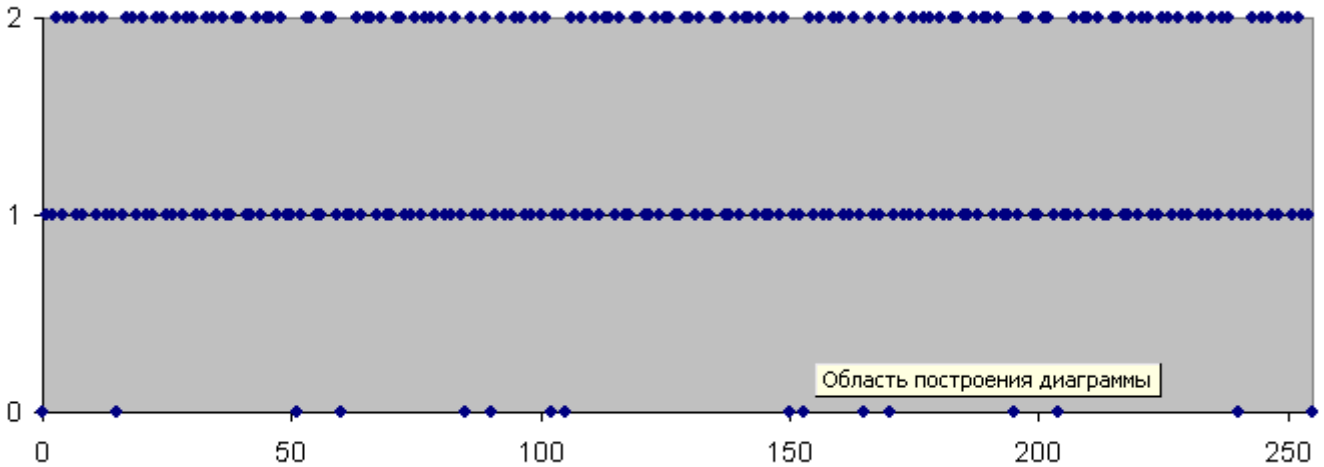

Рис. 1. График значений нелинейности всех булевых функций от 3-х переменных.

Если разделить весь диапазон на 4 поддиапазона, то оказывается, что количество функций каждой нелинейности в каждом диапазоне равно (табл. 4):

Таблица 4. Распределение нелинейностей функций по равным диапазонам (*N*=3, 4 диапазона)

Диапазон \значение нелинейности	0-63	64-127	128-191	192-255
0	4	4	4	4
1	32	32	32	32
2	28	28	28	28

Для *N*=4 распределение в графическом виде выглядит следующим образом (рис. 2).

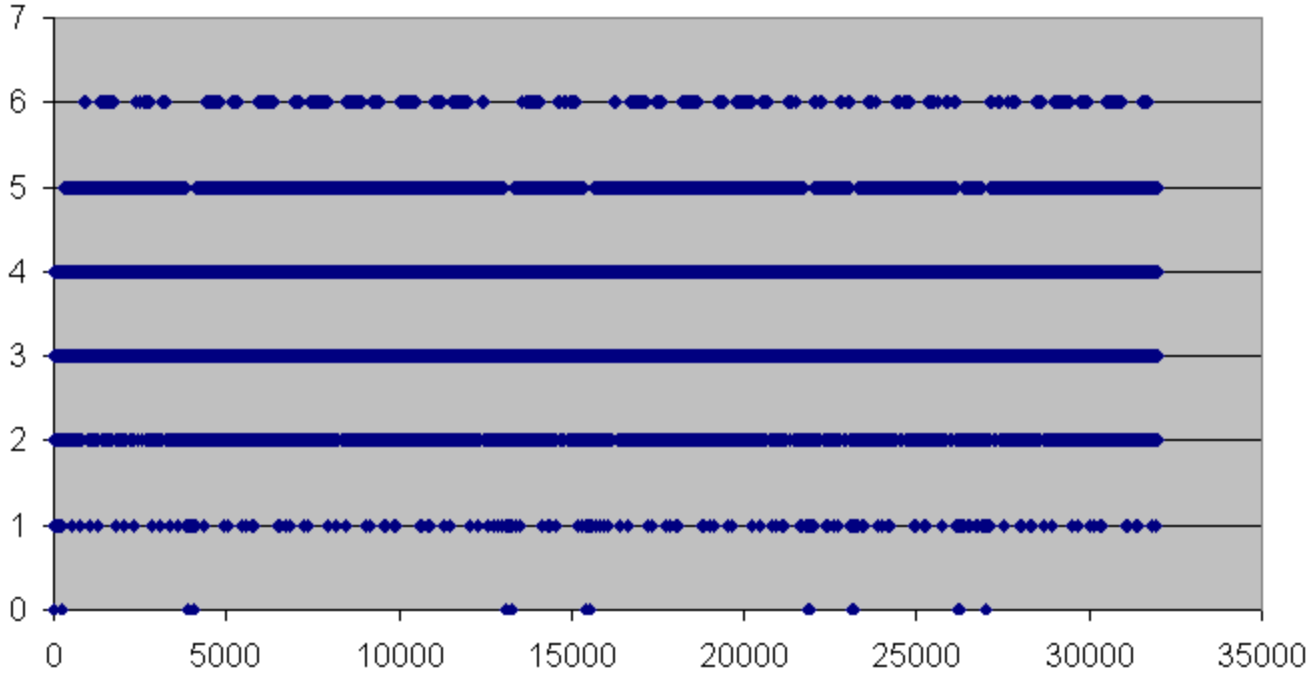

Рис. 2. График значений нелинейности булевых функций от 4-х переменных (первая половина диапазона).

Поскольку графически уже сложно разглядеть какую-то закономерность, более интересным представляется запись распределения в табличном виде (табл. 5, 6).

Таблица 5. Распределение нелинейностей функций по равным диапазонам (*N*=4, 4 диапазона)

Диапазон \ нелинейность	0-16383	16384-32767	32768-49151	49152-65535
0	8	8	8	8
1	128	128	128	128
2	960	960	960	960
3	4480	4480	4480	4480
4	7000	7000	7000	7000
5	3584	3584	3584	3584
6	224	224	224	224

Таблица 6. Распределение нелинейностей функций по равным диапазонам (*N*=4, 16 диапазонов)

Диапазон \ нелинейность	0-4095	4096-8191	8192-12287	12288-16383	16384-20479	20480-24575	24576-28671	28672-32767
0	4	0	0	4	0	4	4	0
1	48	16	16	48	16	48	48	16
2	288	192	192	288	192	288	288	192
3	1168	1072	1072	1168	1072	1168	1168	1072
4	1708	1792	1792	1708	1792	1708	1708	1792
5	832	960	960	832	960	832	832	960
6	48	64	64	48	64	48	48	64
Диапазон \ нелинейность	32768-36863	36864-40959	40960-45055	45056-49151	49152-53247	53248-57343	57344-61439	61440-65535
0	0	4	4	0	4	0	0	4
1	16	48	48	16	48	16	16	48
2	192	288	288	192	288	192	192	288
3	1072	1168	1168	1072	1168	1072	1072	1168
4	1792	1708	1708	1792	1708	1792	1792	1708
5	960	832	832	960	832	960	960	832
6	64	48	48	64	48	64	64	48

Для *N*=5 графически интервал $0\ldots2^{32}-1$ уже изображать не имеет смысла, поэтому использовался только табличный метод (табл. 7, 8).

Таблица 7. Распределение нелинейностей функций по равным диапазонам (*N*=5, 4 диапазона)

Диапазон \ нелинейность	0-1073741823	1073741824-2147483647	2147483648-3221225471	3221225472-4294967295
0	16	16	16	16
1	512	512	512	512
2	7936	7936	7936	7936
3	79360	79360	79360	79360
4	575360	575360	575360	575360
5	3222016	3222016	3222016	3222016
6	14499072	14499072	14499072	14499072
7	53853696	53853696	53853696	53853696
8	161916720	161916720	161916720	161916720
9	340613120	340613120	340613120	340613120
10	353025024	353025024	353025024	353025024
11	139102208	139102208	139102208	139102208
12	6846784	6846784	6846784	6846784

Таблица 8. Распределение нелинейностей функций по равным диапазонам (*N*=5, 16 диапазонов)

Диапазон \ нелинейность	0-268435455	268435456-536870911	536870912-805306367	805306368-1073741823
0	8	0	0	8
1	224	32	32	224
2	3072	896	896	3072
3	27552	12128	12128	27552
4	181952	105728	105728	181952
5	943712	667296	667296	943712
6	3999744	3249792	3249792	3999744
7	14215968	12710880	12710880	14215968
8	41417880	39540480	39540480	41417880
9	85182080	85124480	85124480	85182080
10	86969344	89543168	89543168	86969344
11	33848192	35702912	35702912	33848192
12	1645728	1777664	1777664	1645728
Диапазон \ нелинейность	1073741824-1342177279	1342177280-1610612735	1610612736-1879048191	1879048192-2147483647
0	0	8	8	0
1	32	224	224	32
2	896	3072	3072	896
3	12128	27552	27552	12128
4	105728	181952	181952	105728
5	667296	943712	943712	667296
6	3249792	3999744	3999744	3249792
7	12710880	14215968	14215968	12710880
8	39540480	41417880	41417880	39540480
9	85124480	85182080	85182080	85124480
10	89543168	86969344	86969344	89543168
11	35702912	33848192	33848192	35702912
12	1777664	1645728	1645728	1777664
Диапазон \ нелинейность	2147483648-2415919103	2415919104-2684354559	2684354560-2952790015	2952790016-3221225471
0	0	8	8	0
1	32	224	224	32
2	896	3072	3072	896
3	12128	27552	27552	12128
4	105728	181952	181952	105728
5	667296	943712	943712	667296
6	3249792	3999744	3999744	3249792
7	12710880	14215968	14215968	12710880
8	39540480	41417880	41417880	39540480
9	85124480	85182080	85182080	85124480
10	89543168	86969344	86969344	89543168
11	35702912	33848192	33848192	35702912
12	1777664	1645728	1645728	1777664

Диапазон \ нелинейность	3221225472-3489660927	3489660928-3758096383	3758096384-4026531839	4026531840-4294967295
0	8	0	0	8
1	224	32	32	224
2	3072	896	896	3072
3	27552	12128	12128	27552
4	181952	105728	105728	181952
5	943712	667296	667296	943712
6	3999744	3249792	3249792	3999744
7	14215968	12710880	12710880	14215968
8	41417880	39540480	39540480	41417880
9	85182080	85124480	85124480	85182080
10	86969344	89543168	89543168	86969344
11	33848192	35702912	35702912	33848192
12	1645728	1777664	1777664	1645728

Как видно, при делении на 4 диапазона, булевы функции всех значений нелинейности распределены по ним абсолютно одинаково. При делении на 16 диапазонов появляется два типа распределения, которые не сильно отличаются в части функций с высокой нелинейностью. Для N=5 было также построено распределение на 1024 диапазона. Отобразить его здесь не представляется возможным, однако можно сказать, что оно продолжает тенденцию, обозначившуюся на представленных выше распределениях: среди всех 1024 диапазонов присутствует всего лишь 6 различных видов распределений, и функции с высокой степенью нелинейности распределены довольно равномерно по всем диапазонам, так что нельзя выделить какую-то часть множества функций, где таких функций значительно больше, чем вне его, или наоборот, где таких функций нет или почти нет.

Другой особенностью, замеченной в распределении, было то, оно симметрично относительно середины диапазона. Именно поэтому на втором рисунке было изображено только половина всего множества функций.

Также были произведены замеры скоростей определения нелинейности отдельной функции, в зависимости от её размерности. Напомним, что основным ограничивающим фактором в этом случае является размер массива аффинных функций, и распараллеливание выполняется именно разделением этого массива по потокам при вычислении расстояния от исследуемой функции до каждой функции из этого массива. Измерение выполнялось случайной генерацией булевой функции и последующим вычислением её нелинейности на кластере (112 ядер). Результаты были получены следующие.

N = 15: время проверки нелинейности одной функции – около 2 сек. (больше времени ушло на создание потоков и передачу данных). Нелинейность 2-х проверенных функций – 16022 и 16024.

N = 16: время проверки нелинейности одной функции – около 2,5 сек. (половина времени уходит на создание потоков и передачу данных). Нелинейность 2-х проверенных функций – 32193 и 32244.

N = 17: время проверки нелинейности одной функции – около 5,5 сек. Нелинейность 2-х проверенных функций – 64702 и 64728.

N = 18: время проверки нелинейности одной функции – около 13,5 сек. Нелинейность 2-х проверенных функций – 129841 и 129856.

N = 19: время проверки нелинейности одной функции – около 41 сек. Нелинейность 2-х проверенных функций – 260321 и 260437.

N = 20: время проверки нелинейности одной функции – около 163 сек. Нелинейность 2-х проверенных функций – 521639 и 521864.

N = 21: время проверки нелинейности одной функции – около 398 сек. Нелинейность 2-х проверенных функций – 1044713 и 1044858.

N = 22: время проверки нелинейности одной функции – около 1573 сек. (26 мин). Нелинейность проверенной функции – 2091477.

N = 25: время проверки нелинейности одной функции – около 100470 сек. (1 сутки 3 ч 54 мин).

Как видим, после того, как размер задачи станет таким, что время вычислений будет преобладать над временем на вспомогательные действия (N=19), время вычисления растёт линейно: увеличивается в 4 раза, при увеличении N на 1. Это объясняется тем, что задача отлично распараллеливается по данным, а также при данных значениях N задача пока что умещается целиком в оперативную память.

Библиографический список:

1) Чалкин Т.А. Разработка методики выбора параметров для алгоритма построения узлов замен блочного шифра ГОСТ 28147-89 // Актуальные проблемы безопасности информационных технологий: материалы III Международной научно-практической конференции / под общей ред. О.Н. Жданова, В. В. Золотарева; Сиб. гос. аэрокосмич. ун-т. – Красноярск, 2009. – С. 33-38.

2) А.Е.Жуков Нелинейность булевых функций. Пособие по курсу «Криптографические методы защиты информации» МГТУ им. Н.Э. Баумана: 2002.

3) Guzev V., Serdyuk Y. Asynchronous parallel programming language based on the Microsoft .NET platform. PaCT-2003, LNCS, 2763, Springer, pp. 236-243.

4) http://www.mcsharp.net.

5) http://research.microsoft.com/en-us/um/people/nick/polyphony.htm.

D.A. Nikitin, K.V. Dyakonov

On distribution of Boolean functions nonlinearity values

The results of work on estimation of multivariable Boolean functions nonlinearity potential values with using of cluster computing. Results may be useful for algebra and cryptography, specifically for selecting of block cipher substitution blocks with defined nonlinearity value.