УДК 681.3.06: 519.248.681

Р. В. ОЛЕЙНИКОВ

ДИФФЕРЕНЦИАЛЬНЫЙ КРИПТОАНАЛИЗ АЛГОРИТМА ШИФРОВАНИЯ ГОСТ 28147-89

Симметричные блочные алгоритмы шифрования широко применяются в современных системах криптографической защиты информации. Первоначально они использовались исключительно для шифрования в блочном режиме, однако в настоящее время сфера их применения значительно расширилась. Блочные алгоритмы в качестве базовых примитивов используются при построении поточных шифров, криптографических генераторов псевдослучайных последовательностей и хеш-функций, а также при выработке кодов аутентификации сообщений. Учитывая высокую производительность симметричных шифров, их часто применяют вместе с несимметричными алгоритмами, получая эффективную реализацию модели взаимного недоверия и взаимной защиты. Стойкость, надежность и производительность современных систем криптографической защиты информации в значительной степени определяется характеристиками применяемых симметричных алгоритмов шифрования.

С принятием шифра DES [1] в качестве национального стандарта США в 1977г. наметился подход использовать в системах защиты исключительно стандартизированные алгоритмы. В пользу этого подхода свидетельствуют множество фактов взлома различных алгоритмов, которые не являлись стандартными и использовались без надлежащего анализа. Одним из последних примеров стал алгоритм шифрования A5/1, применяемый для обеспечения конфиденциальности в системе мобильной связи GSM. После детального анализа была обнаружена атака, позволяющая восстанавливать содержание зашифрованных сообщений на достаточно мощном персональном компьютере, хотя сам алгоритм A5/1 устойчив к силовым атакам.

Национальный институт стандартов США при проведении международного конкурса на алгоритм шифрования AES одним из главных требований считал обеспечение стойкости нового алгоритма к различным атакам. Это же условие является одним из основных в европейском проекте создания криптографических стандартов Nessie.

В нашей стране в качестве симметричного блочного шифра используется ГОСТ 28147-89. Кроме обеспечения конфиденциальности, он применяется для выработки кода аутентификации в ГОСТ 34.311-95, в свою очередь, используемом в стандарте цифровой подписи ГОСТ 34.310-95. Принципы проектирования алгоритма его разработчики оставили закрытыми, как и правила генерации долговременных ключей–одного из основных элементов, влияющих на стойкость шифра. Несмотря на более чем десятилетнюю историю, в открытой печати, как отечественной, так и зарубежной, практически отсутствуют сведения об анализе стойкости алгоритма, и нет ни одной публикации об успешной атаке всего алгоритма. В настоящей работе производится исследование стойкости ГОСТ 28147-89 к дифференциальному криптоанализу, и впервые описываются условия, при которых существует атака на алгоритм, более эффективная, чем полный перебор ключей.

Дифференциальный криптоанализ относится к классу атак с выбранными открытыми текстами. Предполагается, что криптоаналитик имеет возможность выбирать значения открытых блоков, подаваемые на вход шифратора и получать соответствующие им зашифрованные значения. При выполнении криптоанализа изучается прохождение разностей между обрабатываемыми текстами (блоками) через циклы шифрования алгоритма. Обычно для вычисления разности выбирается операция, обратная к операции введения ключа, что позволяет при расчете вероятности прохождения разности исключить влияние ключа. В классическом варианте атаки на алгоритм DES [2] используют операцию побитового сложения по модулю 2. Основным объектом атаки является нелинейное преобразование, выполняемое цикловой функцией. Шифр уязвим для дифференциального криптоанализа, если вероятность прохождения разностей через нелинейное преобразование будет достаточно высокой. Как правило, при изучении прохождения разностей, делается предположение о независимости используемых на каждом цикле шифрования подключей, что значительно упрощает анализ. Результатом атаки является вычисление подключей шифрования на последних циклах, по которым восстанавливается полный ключ шифрования.

Для дальнейшего изложения потребуется несколько определений [2]:

Разность (дифференциал) – результат некоторой операции над входными, выходными или промежуточными значениями двух параллельно шифруемых блоков. В ГОСТ 28147-89 ключ вводится с помощью операции сложения по модулю 2^{32}. Для вычисления разности возможно использование двух операций: вычитания по модулю 2^{32} или сложения по модулю 2. Атака с

использованием последней будет более простой, поэтому для вычисления разностей выбрана операцию сложения по модулю 2.

Совокупность разностей открытых и зашифрованных на одном ключе блоков (в дальнейшем обозначаемых как Ω_P и Ω_T), а также соответствующие им промежуточные значения на выходе каждого из n циклов шифрования называется ***n- цикловой характеристикой***. Каждая характеристика имеет определённую *вероятность*, в соответствии с которой при шифровании случайно выбранных открытых блоков с разностью, соответствующей характеристике, все разности при дальнейших преобразованиях, вплоть до зашифрованных блоков, будут также соответствовать характеристике. Если входная разность одной характеристики соответствует выходной разности другой, то их можно объединить. Итоговая вероятность полученной характеристики будет соответствовать произведению вероятностей исходных [2]. С точки зрения криптоанализа лучшая характеристика имеет наибольшую вероятность (но в то же время меньшую 1).

Верной парой называется совокупность открытых блоков X и X' и результат их шифрования Y и Y', разность которых, а также все промежуточные значения при шифровании соответствуют заданным характеристикой разностям.

Подстановкой (или S-блоком) в ГОСТ 28147-89 является перестановка чисел от 0 до 15 в произвольном порядке. Восемь подстановок формируют заполнение узлов замены (таблицу подстановок), используемую как долговременный ключ. Подстановка называется *активной* на i-м цикле, если в используемой характеристике на этом цикле перестановке соответствует ненулевая разность.

Таблицей распределения разностей называется математическая конструкция, описывающая дифференциальные свойства подстановки и позволяющая определить вероятность преобразования входной разности в выходную для заданного ***S***-блока. Для подстановок ГОСТа входные и выходные разности могут изменяться в диапазоне от 0h до 0Fh, а вероятность – принимать значения от 0/16 до 16/16.

Таблица восстановления входных значений – элемент, позволяющий по заданной входной и выходной разности определить все возможные значения на входе подстановки, которые могут дать заданное преобразование разностей.

Цикловая функция – основной элемент криптографического преобразования, итеративно повторяемый при шифровании. В ГОСТ 28147-89 цикловую функцию составляет сложение с ключом, подстановка и циклический сдвиг влево на 11 разрядов.

Сложностью атаки на алгоритм шифрования является количество необходимых операций шифрования для восстановления ключа при условии, что криптоаналитик обладает всей остальной необходимой для него информацией (описание алгоритма шифрования, определённое количество пар открытых и зашифрованных на искомом ключе блоков и т.д.). Сложность дифференциальной атаки зависит от вероятности лучшей характеристики. В свою очередь, вероятность характеристики зависит от числа циклов, количества активных подстановок на каждом цикле и вероятностей нужного преобразования разностей в активных ***S***-блоках. Поскольку количество циклов в ГОСТ 28147-89 равно 32, то критерием выбора лучшей характеристики будет минимальное количество активных подстановок и набольшая вероятность заданного преобразования в каждом ***S***-блоке.

Кратко остановимся на свойствах подстановок с точки зрения дифференциального криптоанализа. Наиболее вероятную характеристику будут формировать подстановки с максимальными вероятностями преобразования разностей. Поскольку долговременный ключ стандартом не определяется, то имеется возможность выбора таблицы подстановок с любыми свойствами, в том числе уязвимыми для дифференциальной атаки – с вероятностью преобразования разностей, равной 1. Пример перестановки, для которой входная разность, равная 01h, всегда преобразуется в выходную 02h, приведен в табл. 1. Приблизительную оценку количества перестановок ГОСТа, обладающих таким свойством, можно найти в [3].

Таблица 1

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
6	4	1	3	0	2	7	5	14	12	13	15	8	10	9	11

Более детальную зависимость преобразований входных разностей в выходные для приведенной подстановки можно увидеть из распределения разностей в таблице 2. В левой колонке выбирается входная разность $\Delta_{вх}$, в верхней строке – выходная $\Delta_{вых}$, а число на пересечении выбранной колонки и столбца даёт количество входных пар, соответствующих заданному преобразованию (пустая ячейка означает, что выбранное преобразование невозможно). Разделив это число на 16 (общее количество пар, формирующих разность), можно получить вероятность преобразования $p_{пр}$ выбранных разностей в ***S***-блоке.

Таблица 2

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	16															
1			16													
2		4		4					8							
3		4		4		8										
4			4		12											
5			12		4											
6		8				4		4								
7				8		4		4								
8									8				4		4	
9											8		4		4	
A										4		4				8
B										4		4		8		
C									4		4				8	
D									4		4		8			
E										8				4		4
F												8		4		4

Имея таблицы распределения разностей всех 8 перестановок, можно определить вероятность преобразования произвольных входных разностей для цикловой функции ГОСТа в произвольные выходные. Рассмотрим подробнее прохождение разностей через цикл шифрования (рис. 1).

Сначала разность попадает на сумматор, где к обрабатываемому полублоку добавляется подключ. Операция введения ключа (сложение по модулю 2^{32}) является нелинейной по отношению к операции вычисления разности (сложению по модулю 2), поэтому при преобразовании возможно искажение разностей. В [2] для решения этой проблемы (при рассмотрении модификаций алгоритма DES) предлагается строить таблицу распределения разностей, при этом полагая, что значение ключа является случайной равномерно распределённой величиной. Однако это не решает проблемы, поскольку при выполнении дифференциальной атаки все блоки шифруются на одном ключе, который является постоянным.

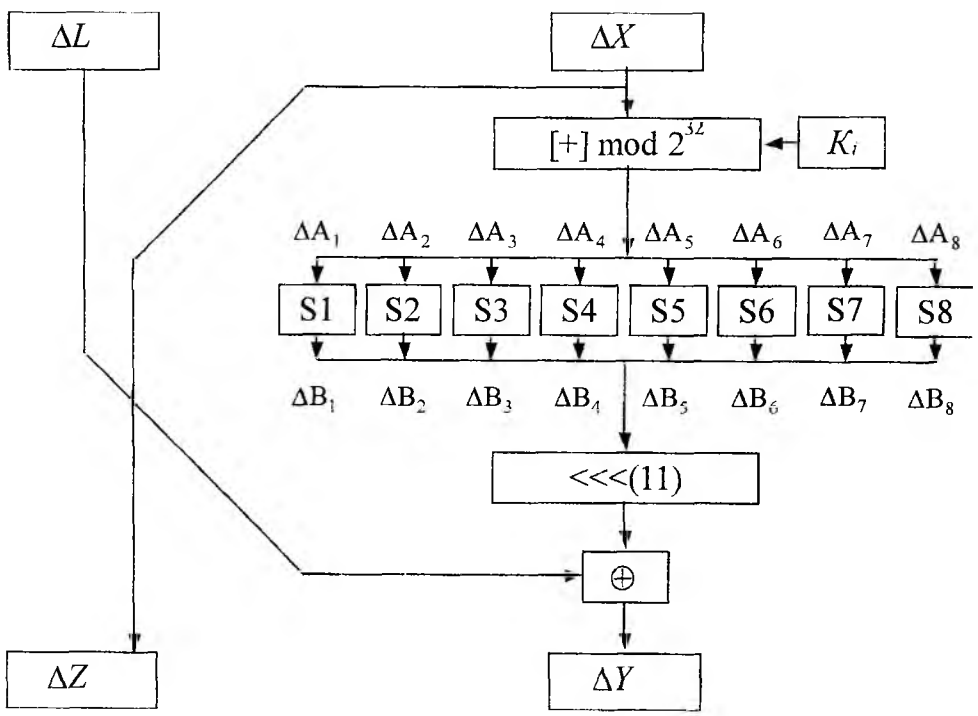

Рис.1

Более предпочтительным вариантом является построение отдельных таблиц распределения разностей для каждого из значений ключа, поскольку разные ключи дают разные вероятности преобразования. Вычислить, проанализировать и сохранить таблицу по размеру сумматора (2^{32}x2^{32}) представляется достаточно ресурсоемкой задачей, поэтому имеет смысл разбить входную 32-разрядную разность на 8 блоков по 4 бита (по размеру ***S***-блока). При таком размере вектора разности нужно построить всего 16 таблиц (для всех 4-битовых значений ключа от 0 до 0Fh размера 16x16). Пример преобразования разностей для ключа, равного 0Eh, приведен в табл. 3 (для полного анализа преобразования на сумматоре требуется, соответственно, 16 таких таблиц).

Можно отметить, что для выбранного ключа входная разность 01h всегда преобразовывается в выходную 01h (вероятность преобразования p_s=16/16=1). Эта особенность сохраняется для всех четных 4-битовых блоков подключа. Для нечетных блоков подключа это преобразование является невозможным (вероятность преобразования p_s=0/16=0). Следующей важной особенностью таблиц распределения разностей при ключевом преобразовании так же, как и при преобразовании на ***S***-блоках, является прохождение нулевой разности через сумматор без искажения. Одним из наиболее оптимальных вариантов построения характеристики будет комбинирование максимального количества нулевых 4-битовых разностей с минимальным количеством единичных с прохождением разностей через сумматор без искажения.

Таблица 3

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	16															
1		16														
2							8								8	
3								8								8
4					8								8			
5						8								8		
6			8								8					
7				8								8				
8									16							
9										16						
A							8								8	
B								8								8
C					8								8			
D						8								8		
E			8								8					
F				8								8				

В случае активной младшей подстановки на ключ шифрования накладывается дополнительно ограничение: для прохождения разностей через сумматор без искажения младший бит подключа должен быть равным нулю. Вероятность прохождения разности 01h через сумматор без искажения во всех остальных тетрадах (4-битовых блоках) равна 1/2 из-за возникновения переносов из младших разрядов в старшие при сложении с ключом по модулю 2^{32}. При этом предполагается, что нулевые и единичные биты ключа имеют равную вероятность появления, а в используемых характеристиках разности 01h всегда предшествует нулевая, что определяет возникновение синхронного переноса (одинакового в двух шифруемых блоках). Перенос в активную тетраду из предыдущей возникает с вероятностью 1/2, что эквивалентно прибавлению единицы к активным четырём битам ключа. При использовании четного ключа получается искажение разностей. Для нечетного ключа это можно представить как преобразование в четный и соответствующее прохождение разностей через сумматор без искажения. Отсюда получается вероятность преобразования, равная 1/2, для всех активных тетрад, в которые возможен перенос из младших разрядов вне зависимости от конкретных значений битов ключа.

В среднем при сложении перенос распространяется не более, чем на 2 разряда, что меньше длины анализируемых блоков. Поэтому преобразования 4-битовых разностей при добавлении 32-битового подключа можно считать независимыми событиями, и при вычислении вероятности

преобразования разности для всего сумматора следует перемножить вероятности для каждого 4-битового элемента.

Исходя из вышеизложенного, можно определить вероятность прохождения разности заданного вида (и вычисленной по модулю 2) через ключевой сумматор без искажения:

$$p_{np} = \left(\frac{1}{2}\right)^{k_a}, \tag{1}$$

где k_a – количество активных (ненулевых) тетрад в разности при условии, что младшая тетрада неактивна или в текущем цикле шифрования добавляется чётный подключ (с младшим битом, равным нулю). Если это условие не выполняется, разность обязательно будет искажена (вероятность прохождения равна нулю).

После сумматора разность попадает на ***S***-блоки. Поскольку они не определены стандартом, их можно выбрать со следующим свойством: как и для подстановки, приведенной в таблице 1, входная разность, равная 01h, всегда преобразуется в выходную 02h. В этом случае разность проходит через подстановку без изменений.

Затем в цикловой функции следует циклический сдвиг влево на 11 разрядов. Фактически это является переименованием двоичных переменных, представляющих каждый разряд разности. После сдвига активные тетрады займут другие позиции, а активная разность 02h будет преобразована в 01h. Пример прохождения разности $\Omega_{[}$ = 00101000h через один цикл шифрования приведен в таблице 4, при этом вероятность преобразования при сложении с ключом вычислена по формуле (1).

Таблица 4

Операция	Входное значение	Выходное значение	Вероятность преобразования
Сложение с подключом	00101000h	00101000h	2^{-2}
Подстановка	00101000h	00202000h	1
Сдвиг влево	00202000h	01000001h	1

Итоговая вероятность прохождения разностей через один цикл шифрования (вероятность одноцикловой характеристики) будет равна произведению вероятностей преобразования на каждом из элементов:

$$p_u = p_{sm} \cdot p_s \cdot p_r = 2^{-2}. \tag{2}$$

Отсюда следует, что, в среднем, в одном из четырёх шифрований на одном цикле будет выполняться нужное преобразование разностей.

Характеристику можно распространить на все 32 цикла шифрования путём комбинирования одноцикловых. Опять же, итоговая вероятность построенной характеристики будет равна произведению вероятностей характеристик, её составляющих:

$$p_\Omega = \prod_{i=1}^{32} p_i \,. \tag{3}$$

Размер блока в ГОСТ 28147-89 составляет 64 бита, что даёт 2^{64} возможных входных значений. Поэтому вероятность характеристики, пригодной для дифференциальной атаки, ограничена снизу значением $p_\Omega = 2^{-64}$. Если для заданного набора подстановок не существует характеристик, удовлетворяющих выбранному условию, тогда алгоритм неуязвим для дифференциального криптоанализа. Для подстановок с дифференциальными свойствами, аналогичными свойствам подстановки, приведенной в таблице 1, вероятность 32-цикловой характеристики (с младшими активными битами в тетрадах) будет находиться в пределах от $p_{\Omega_{\min}} = 2^{-154}$ (Ω_P = 0000 0000 1111 1111h и Ω_T = 1111 1111 0000 0000h) до $p_{\Omega_{\max}} = 2^{-33}$ (Ω_P=0000 0010 0100

0001h и Ω_T= 0000 0010 0101 0001h). Пример прохождения характеристики (Ω_P=00000000 00000001h, Ω_T=00100000 00010100h, $p_\Omega = 2^{-38}$) приведен в [3].

При выполнении атаки выбирается случайный открытый блок X, вычисляется соответствующий характеристике блок $X' = X \oplus \Omega_P$, после чего оба блока зашифровываются на секретном (искомом) ключе. В среднем требуется выполнить $(p_\Omega)^{-1}$ шифрований для того, чтобы полученные значения Y и Y' удовлетворяли соотношению $Y \oplus Y' = \Omega_T$. Найдя верную пару, можно получить значения входных и выходных разностей в активном ***S***-блоке на последнем цикле.

Левая половина зашифрованного блока соответствует входному значению цикловой функции на последнем цикле, поэтому известны значения X и X' для 32-го цикла (см. рис.1). $\Delta X = X \oplus X'$ можно вычислить непосредственно из зашифрованных блоков, но это значение уже известно, поскольку для верной пары характеристики известны все промежуточные разности. Предположив, что при сложении с ключом разность не исказилась, можно разделить ΔX на 8 значений по четыре бита, получив при этом входные разности ΔA_i для каждого ***S***-блока. Значение ΔL и ΔY также известно из характеристики, и из них вычисляется $\Delta B = \Delta L \oplus \Delta Y$. Разбив ΔB по ***S***-блокам, получим значения выходных разностей для каждого из них.

Как следует из табл. 2, только часть входных значений может вызвать заданное преобразование разностей. Например, входная разность $\Delta_{вх}$= 0Ah преобразуется в выходную $\Delta_{вых}$= 09h только четырьмя входными значениями. Соответственно, по известным входным и выходным разностям можно определить подмножество возможных входных значений. Удобнее всего для этого использовать таблицы восстановления входных значений. Пример для выходной разности $\Delta_{вых}$= 09h приведен в табл. 5. В левой колонке выбирается входная разность $\Delta_{вх}$, в верхней строке входное значение A_i, а число на пересечении выбранной колонки и столбца показывает вероятность заданного преобразования разностей. Пропущенные строки и пустые ячейки означают, что для выбранных значений преобразование невозможно. Для восстановления входных значений подстановки при любых выходных разностях требуется, соответственно, 16 таких таблиц.

Таблица 5

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
A					1	1									1	1
B	1	1									1	1				
E			1	1			1	1	1	1			1	1		

Получив верную пару, в соответствии с характеристикой можно определить ΔA_i и ΔB_i для активного на последнем цикле ***S***-блока (см. рис. 1). По разностям можно определить все входные значения, причем, чем меньше вероятность преобразования $\Delta A_i \rightarrow \Delta B_i$, тем меньше остаётся вариантов входных значений A_i. Поскольку $\Delta X = \Delta Z$, а также $X = Z$ и $X' = Z'$, можно вычислить варианты нескольких битов подключа шифрования на последнем цикле, соответствующие активному ***S***-блоку: $K_8^{S_i} = A_i - Z_i (\mathrm{mod}16)$. В верной паре присутствуют два шифртекста Y и Y' (соответственно два значения Z и Z'), для которых выполняется определение вероятных битов ключа шифрования. Найдя пересечение возможных значений, можно дополнительно уменьшить их количество. Для вероятности преобразования разности на ***S***-блоке, равной 2/16, остаётся одно возможное (верное) значение четырёх битов ключа для одной найденной верной пары.

Заметим, что при использовании преобразования разностей с вероятностью 1 (см. таблицу 2) все входные значения вызовут заданное преобразование, что не позволит ограничить количество возможных входных значений. Поэтому, хотя для построения характеристики используются переход с вероятностью 1, для восстановления ключей необходимо использовать искажение разностей при преобразовании на сумматоре в ходе последнего цикла шифрования. В этом случае удаётся скомбинировать характеристику с высокой вероятностью и переход с низкой вероятностью на последнем цикле, что в результате даст возможность определить несколько битов ключа со сравнительно низкими вычислительными затратами. Для переходов с вероятностью, меньшей 1,

вероятность 32-цикловых характеристик в ГОСТе будет меньше 2^{-64}, из чего следует, что с высокой степенью вероятности верной пары для таких характеристик вообще не существует.

При поиске верной пары для характеристики задаётся разность для открытых блоков, а после выполнения шифрования проверяется разность между зашифрованными блоками. В случае, когда нужно найти пару, соответствующую характеристике до предпоследнего цикла и последующим искажении на сумматоре, выполняются несколько дополнительных проверок. Значение левой половины зашифрованного блока (соответственно и разницы между ними) останется без изменения ($X = Z$, см. рис. 1). Отсюда же следует, что значение ΔL с высокой вероятностью соответствует характеристике. После сумматора изменится активная разность ΔA_i. Также возможно изменение разностей в блоках, старших по отношению к активному. Значение разности на выходе S блоков можно получить по формуле $\Delta B = (\Delta Y \oplus \Delta L) >>> 11$ (результат сложения по модулю 2 сдвигается вправо на 11 разрядов). Здесь ΔL известна из характеристики, а ΔY вычисляется по правой половине зашифрованных блоков. Из ΔB находится искомая разность ΔB_i. Разность на входе S-блока определяется искажением на ключевом сумматоре. Рассматривая различные варианты искажения разности на сумматоре, можно найти преобразование на S-блоке, дающее однозначное определение ключевых битов, причём вероятность нахождения верной пары для такой характеристики будет лишь немного меньше (порядок разности вероятностей примерно равен 2^{-3}).

Используя приведенную технику, можно восстановить 28 из 32 битов на последнем цикле шифрования. Для получения оставшихся четырёх битов необходимо использовать искажение разностей на предпоследнем цикле шифрования. Контроль соответствия характеристике обрабатываемых пар выполняется путём сравнения уже известных значений на выходе предпоследнего цикла. Они получаются после расшифрования с использованием вычисленных битов ключа шифрования.

После получения всех 32 битов подключа на последнем цикле шифрования можно выполнить расшифрование на 1 цикле и затем применить описанную методику к ГОСТу с 31 циклом для получения следующих 32 битов ключа. Продолжая атаку, возможно вычислить все 256 битов сеансового ключа. Сложность вычисления составит менее 2^{50} эквивалентных операций шифрования с использованием алгоритма ГОСТ 28147-89. Дополнительно повысить эффективность атаки позволит квартет-метод из [2].

Описанное нападение применимо лишь к "слабым" сеансовым ключам, описанным в [3]. Дополнительным условием является применение "слабых" долговременных ключей, оптимизированных для выполнения дифференциальной атаки.

Рассмотренная атака была реализована для ГОСТа с восемью циклами шифрования. Поиск подключа на последнем цикле шифрования занимал несколько минут на ЭВМ класса 486SX-33. Криптоанализ полного 32-циклового алгоритма, по нашим оценкам, займёт время порядка одного месяца при использовании вычислительной сети из 500 компьютеров класса Celeron.

Для защиты ГОСТ 28147-89 от атак дифференциального криптоанализа достаточно использование долговременных ключей, не позволяющих строить характеристики с вероятностью, превышающей 2^{-64}. Методы формирования таких подстановок предполагается рассмотреть в одной из будущих публикаций.

Список литературы: 1. *FIPS PUB 46-3. Federal Information Processing Standards Publication.* U.S. Department Of Commerce/National Institute of Standards and Technology. Data Encryption Standard (DES). 1999. 2. *Biham E., Shamir A.* Differential Cryptanalysis of the Data Encryption Standard. Springer-Verlag. New York. 1993. 3. *Долгов В.И., Лисицкая И.В., Олейников Р.В.*. "Слабые" ключи в алгоритме шифрования ГОСТ 28147-89 // Радиотехника. 2000. Вып 114. С. 63–68.

Харьковский государственный технический университет радиоэлектроники

Поступила в редколлегию 19.03.2001