

ПОДСТАНОВОЧНО-ПЕРЕСТАНОВОЧНЫЕ ШИФРЫ

В работе предложен новый метод шифрования данных, основанный на преобразовании и перестановке блоков текста, причем блоки преобразуются и переставляются внутри массива большего размера в зависимости от текущего состояния рабочего ключа и шифруемой информации. Предложенный метод легко реализуется программным способом и обеспечивает рассеивание и перемешивание информации в пределах всего шифруемого массива данных.

Введение

Цель данного исследования – получение способа шифрования данных, легко реализуемого программным способом и обеспечивающего рассеивание и перемешивание информации в пределах всего шифруемого массива данных, размер которого ограничивается, вообще говоря, лишь размерами системной памяти используемой ЭВМ.

Разработанный способ может быть реализован в виде алгоритма шифрования, который работает с блоком данных и ключом произвольной длины. В алгоритме используется перестановка m -битных блоков внутри обрабатываемого массива данных большого размера (сравнимого с объемом системной памяти компьютера), бинарная операция XOR m -битных блоков и их циклический сдвиг, причем каждое последующее преобразование зависит от предыдущего. Шифр сочетает в себе элементы блочных шифров, поточных шифров и шифров перестановки. Для уменьшения корреляции между открытым текстом и шифртекстом на текст накладывается гамма, для инициализации которой используется дополнительный входной массив произвольной длины. Инициализация массива гаммы проводится однократно на начальном этапе работы алгоритма, в процессе инициализации используются однонаправленные функции. Полученный в итоге шифртекст зависит от открытого текста, ключа и массива гаммы. Алгоритм может оперировать m -битными блоками любого размера. От размера блока зависят скорость и перемешивающие характеристики алгоритма. На данный момент написаны реализации алгоритма, работающие с 16-ти и 32-х битными блоками, что, по-видимому, является оптимальным выбором.

¹⁾ Московский Инженерно-Физический Институт, кафедра «Защита информации», rus_mail2@mtu-net.ru

²⁾ Московский Государственный Технический Университет им. Н.Э.Баумана, кафедра «Информационная безопасность», azhukov@iu8.bmstu.ru

Особенность данного алгоритма заключается в том, что в процессе работы алгоритма *ключевая информация постоянно изменяется в процессе ее использования в зависимости от шифруемых данных*: изменение секретного ключа происходит в зависимости от открытого текста, самого ключа и массива гаммы.. Вместе с шифртекстом нужно хранить (или передавать) информацию, необходимую для восстановления ключа расшифрования, роль которого выполняет значение ключа в конце работы алгоритма. В качестве такой информации можно использовать $K' \oplus K$ – сумму секретного ключа и ключа расшифрования. Величина $K' \oplus K$ может также рассматриваться как хэш-функция и использоваться для построения системы электронной цифровой подписи.

1. Подстановочно-перестановочное преобразование информации

В основу метода положена идея соединения двух типов преобразований: псевдослучайных преобразований отдельных блоков шифруемого массива данных (информационного массива) и псевдослучайных перестановок преобразованных блоков в пределах всего информационного массива. Поставленная задача достигается тем, что при преобразовании очередного блока данных вначале вычисляется новое значение рабочего ключа, которое зависит как от значения рабочего ключа на предыдущем этапе работы, так и от значения информационного блока, преобразуемого в данный момент. Затем по вычисленному значению рабочего ключа выбирается еще один блок массива данных, номер которого является псевдослучайной величиной. Оба блока подвергаются совместным псевдослучайным преобразованиям, после чего их образы меняются местами в информационного массива данных. Под *псевдослучайностью* понимается то, что без знания текущего значения рабочего ключа невозможно определить совместно с каким блоком будет преобразовываться данный информационный блок и каким преобразованиям они будут подвержены.

Формальное описание подстановочно-перестановочных шифров

Предлагаемый способ криптографического преобразования может быть описан следующим формальным образом.

Будем обозначать через $f(x|y_1, \dots, y_n)$ функцию (отображение) множества $X \times Y_1 \times \dots \times Y_n$ в множество Z , обладающую тем свойством, что при любых фиксированных значениях элементов y_1, \dots, y_n , где $y_i \in Y_i$, соответствующее отображение множества X в множество Z обратимо. Если зафиксировать значения элементов

y_1, \dots, y_n , ($y_i \in Y_i$), то соответствующее обратное отображение будем обозначать через $f^{-1}(z|y_1, \dots, y_n)$, т.е. если $f(x|y_1, \dots, y_n)=z$, то $f^{-1}(z|y_1, \dots, y_n)=x$.

Пусть M – двоичный массив данных, подлежащий криптографическому преобразованию; пусть объем этого массива – $n \cdot m$ бит, где m – некоторая фиксированная величина. Размер массива данных может быть произвольным и ограничивается лишь размером системной памяти ЭВМ.

Вначале весь массив данных M размера $n \cdot m$ бит разбивается на n блоков (M_0, M_1, \dots, M_{n-1}), называемых в дальнейшем **информационными блоками**, по m бит в каждом блоке.

Алгоритм шифрования состоит из $T \geq 1$ циклов шифрования: сначала выполняется 1-й цикл шифрования, затем 2-й и т.д. до T -го цикла. Каждый цикл шифрования состоит из n итераций: при выполнении t -го цикла шифрования в начале выполняется 0-я итерация, затем 1-я итерация и т.д. до $(n-1)$ -й итерации. После выполнения $(n-1)$ -й итерации t -го цикла шифрования алгоритм начинает выполнять 0-ю итерацию $(t+1)$ -го цикла. После выполнения $(n-1)$ -й итерации T -го цикла шифрования алгоритм заканчивает работу.

Пусть K – секретный ключ. В начале работы алгоритма ключевая информация помещается в рабочий массив, содержимое которого в процессе работы алгоритма постоянно изменяется. Будем обозначать заполнение этого массива при i -ой итерации t -го цикла работы алгоритма через $K_i^{(t)}$ и называть его значением **рабочего ключа** при i -ой итерации t -го цикла. Преобразование, которому алгоритм шифрования подвергает массив данных M в результате i -ой итерации t -го цикла задается (определяется) следующим образом:

1. Вычисляется значение рабочего ключа для i -ой итерации t -го цикла шифрования: $K_i^{(t)} = f(K_{i-1}^{(t)} | M_i)$ для всех $i \geq 1$,
При этом: для $i = 0$, $t \geq 2$, полагаем $K_{-1}^{(t)} = K_{n-1}^{(t-1)}$;

для $i = 0$, $t = 1$, полагаем $K_{-1}^{(1)} = K$.

2. Определяется номер блока, который будет преобразовываться совместно с блоком M_i : $j = \varphi(K_i^{(t)})$.
3. Вычисляются новые значения блоков M_i и M_j :

$$A = F_1(M_j | M_i, K_i^{(t)})$$

$$B = F_2(M_i | K_i^{(t)})$$

$$M_i = A$$

$$M_j = B$$

4. Если $i < n-1$, то осуществляется переход к итерации $i+1$ того же цикла t ;
 Если $i = n-1$ и $t < T$, то осуществляется переход к итерации 0 цикла $t+1$;
 Если $i = n-1$, $t = T$, то работа алгоритма заканчивается.

Значения блоков M_0, M_1, \dots, M_{n-1} после завершения последней итерации последнего цикла шифрования образуют массив шифртекста, который передается (хранится) вместе со значением хэш-функции

$$K' = k(K_{n-1}^{(T)} | K),$$

которая одновременно служит для получения ключа расшифрования $K_{n-1}^{(T)}$.

Алгоритм расшифрования преобразует массив данных M , который также разбит на блоки M_0, M_1, \dots, M_{n-1} по m бит в каждом и начинается с определения значения рабочего ключа $K_{n-1}^{(T)}$: $K_{n-1}^{(T)} = k^{-1}(K' | K)$.

Затем продельвается T циклов расшифрования, начиная с T -го цикла расшифрования и заканчивая 1-м циклом расшифрования. Каждый цикл расшифрования состоит из n итераций, начиная с выполнения $(n-1)$ -й итерации и заканчивая выполнением 0-й итерации. После 0-й итерации t -го цикла расшифрования выполняется $(n-1)$ -я итерация $(t-1)$ -го цикла расшифрования. После выполнения 0-й итерации 1-го цикла расшифрования алгоритм расшифрования заканчивает работу.

Преобразование, которому алгоритм расшифрования подвергает массив данных M в результате i -ой итерации t -го цикла расшифрования задается следующим образом:

1. Определяется номер блока, который будет преобразовываться совместно с блоком M_i : $j = \varphi(K_i^{(t)})$.
2. Вычисляются новые значения блоков M_i и M_j :

$$A = F_2^{-1}(M_j | K_i^{(t)})$$

$$B = F_1^{-1}(M_i | A, K_i^{(t)})$$

$$M_i = A$$

$$M_j = B$$
3. Вычисляется значение рабочего ключа для $(i-1)$ -ой итерации t -го цикла шифрования:

$$K_{i-1}^{(t)} = f^{-1}(K_i^{(t)} | M_i) \text{ для всех } i \geq 1;$$

 При этом: для $i = 0$, $t < T$, полагаем $K_{n-1}^{(t)} = K_{n-1}^{(t+1)}$;

$$\text{для } i = n-1, t = T, \text{ полагаем } K_{n-1}^{(T)} = k^{-1}(K' | K).$$

4. Если $i < 0$, то осуществляется переход к итерации $i-1$ того же цикла t ;
 Если $i = 0$ и $t > 1$, то осуществляется переход к $(n-1)$ -й итерации цикла $t-1$;

Если $i = 0$, $t = 1$, то работа алгоритма расшифрования заканчивается.

Значения блоков M_0, M_1, \dots, M_{n-1} после завершения 0-й итерации 1-го цикла расшифрования образуют массив исходного открытого текста.

Отметим еще раз следующие особенности предлагаемого метода:

- рабочий ключ для каждой итерации алгоритма вычисляется по значению преобразуемого блока и значению рабочего ключа на предыдущей итерации;
- информационный блок, преобразуемый во время очередной итерации, преобразуется совместно с другим блоком, выбор которого из общего массива данных зависит от значения рабочего ключа на данной итерации;
- два преобразованных на данной итерации блока переставляются в общем массиве данных.

Благодаря такому решению обеспечивается рассеивание и перемешивание информации в пределах всего преобразуемого массива данных, что делает невозможным применение статистических методов криптоанализа, в том числе методов разностного криптоанализа и линейного криптоанализа.

2. Выбор функций и реализация

При конкретной реализации предлагаемого способа криптографического преобразования информации, необходимо задать легко реализуемые обратимые функции

$$f(K_{i-1}^{(t)} | M_i), F_1(M_j | M_i, K_i^{(t)}), F_2(M_i | K_i^{(t)}), k(K_{N-1}^{(T)} | K),$$

а также легко реализуемую функцию выбора номера блока $\varphi(K_i^{(t)})$.

Ниже для одного из вариантов выбора указанных функций приводится описание соответствующей реализации подстановочно-перестановочного способа шифрования.

2.1. Входные данные

Пусть M – двоичный массив данных, подлежащий криптографическому преобразованию, в дальнейшем этот массив будет называться **информационным массивом**; пусть объем этого массива – $n \cdot t$ бит, где t – некоторая фиксированная величина. Размер информационного массива может быть произвольным и ограничивается лишь размером системной памяти ЭВМ. Весь массив данных M размера $n \cdot t$ бит разбивает-

ся на n блоков (M_0, M_1, \dots, M_{n-1}) по m бит в каждом блоке, в дальнейшем эти блоки будут называться **информационными блоками**.

Пусть K – двоичный массив данных, образующих секретный ключ, в дальнейшем этот массив будет называться **ключевым массивом**; пусть объем этого массива – $s*m$ бит, где m – та же величина, что и выше. Весь массив K разбивается на s блоков (K_0, K_1, \dots, K_{s-1}) по m бит в каждом блоке, в дальнейшем эти блоки будут называться **подключами**. В процессе работы алгоритма заполнение массива K меняется при переходе к каждой следующей итерации; его заполнение при выполнении i -ой итерации t -го цикла работы алгоритма является значением рабочего ключа $K_i^{(t)}$. **Рабочим подключем** называется блок массива K , используемый в преобразованиях на i -ой итерации t -го цикла работы алгоритма.

Для изменения информационных блоков и подключей в процессе работы алгоритма используются дополнительные блоки, которые выбираются псевдослучайным образом из дополнительного массива G , проинициализированного в начале работы алгоритма и называемого **массивом гаммы**. Пусть объем этого массива – $p*m$ бит, где m – та же величина, что и выше. Весь массив G разбивается на p блоков (G_0, G_1, \dots, G_{p-1}) по m бит в каждом блоке, в дальнейшем эти блоки будут называться **блоками гаммы**.

2.2. Используемые обозначения

Будем пользоваться следующими обозначениями:

- X' – преобразованный массив X ;
- $(X)RRot(Y)$ – циклический сдвиг X вправо на Y позиций;
- $(X)LRot(Y)$ – циклический сдвиг X влево на Y позиций;
- Операция $+$ это либо \oplus (XOR), либо $\boxed{+}$ (сложение по модулю 2^m , где m – размер блоков в битах);
- $\nu_1, \nu_2, \pi_1, \pi_2 : (Z_2)^m \rightarrow Z_m$;
- $\lambda_1, \mu_1 : (Z_2)^m \rightarrow Z$;
- $\lambda_2, \mu_2, \psi_1, \psi_2 : (Z_2)^m \rightarrow Z_p$;
- $\varphi : (Z_2)^m \rightarrow Z_n$.

1.2.2. Инициализация гаммы

На начальном этапе происходит инициализация массива гаммы. Данная процедура необходима для того, чтобы обеспечить выработку качественной гаммы в процессе шифрования. Изначально массив гаммы заполняется любыми данными, например, текстом или псевдослучайной последовательностью. Инициализация представляет собой многократное однонаправленное изменение начального заполнения массива гаммы (например, операциями циклического сдвига, векторного сложения, поразрядного XOR и т.д.) с использованием секретного ключа. Инициализация снимает жесткие ограничения на длину и начальное заполнение массива гаммы, что позволяет использовать в качестве начального заполнения как текстовые, так и бинарные фрагменты.

Инициализация гаммы: $G' = I_K(G)$ – однократная инициализация дополнительного массива в зависимости от ключа шифрования.

1.2.3. Зашифрование

Зашифрование информационного массива производится последовательным считыванием информационных блоков и подключей (из соответствующих массивов) с их последующим преобразованием. i -я итерация алгоритма совместно обрабатывает очередной информационный блок M_i и псевдослучайно выбранный информационный блок M_j . При преобразовании очередного блока данных вначале вычисляется новое значение рабочего ключа, которое зависит как от значения рабочего ключа на предыдущем этапе работы, так и от значения информационного блока, преобразуемого в данный момент. Затем по вычисленному значению рабочего ключа выбирается еще один блок массива данных, номер которого является псевдослучайной величиной. Оба блока подвергаются совместным псевдослучайным преобразованиям, после чего их образы меняются местами в информационном массиве данных. Таким образом, при i -й итерации алгоритма происходят следующие преобразования ключевого массива K и информационного массива M :

Преобразование массива K : Вычисляется значение рабочего ключа для очередной итерации очередного цикла шифрования:

- $K'_i \pmod s = K_i \pmod s + M_i$ – изменение текущего подключа в зависимости от текущего блока текста;
- $K'_{i+1} \pmod s = K_{i+1} \pmod s + K^{\lambda_1 \left(K'_i \pmod s \right) + G^{\lambda_2 \left(K'_i \pmod s \right)}}$ – изменение следующего по порядку подключа в зависимости от текущего подключа;

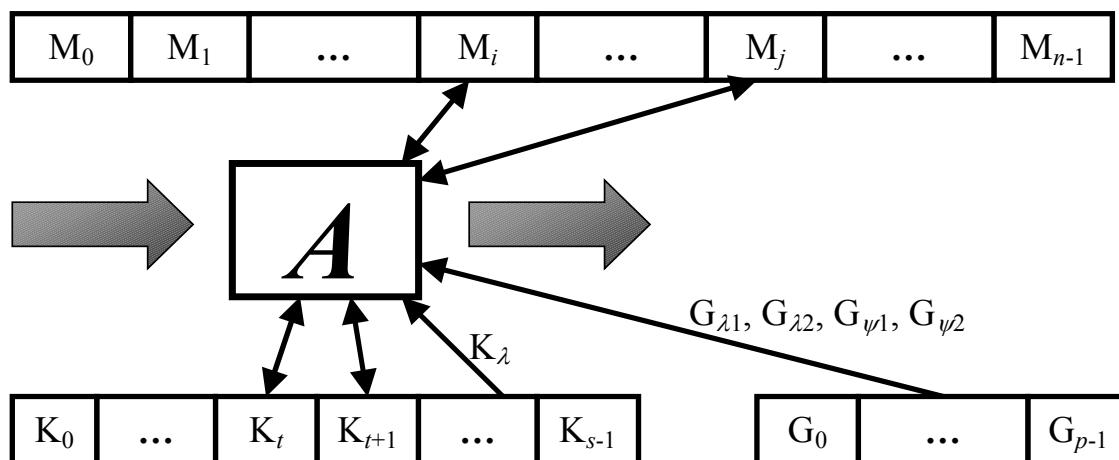
- $K''_i(\text{mod } s) = K'_i(\text{mod } s) + K_{\mu_1(K'_{i+1}(\text{mod } s))} + G_{\mu_2(K'_{i+1}(\text{mod } s))}$ – изменение текущего подключа в зависимости от следующего по порядку использования подключа;
- $K''_{i+1}(\text{mod } s) = K'_{i+1}(\text{mod } s) \text{RRot}(\nu_1(K''_i(\text{mod } s)))$ – циклический сдвиг следующего подключа вправо на количество разрядов, определяемое текущим подключом;
- $K'''_i(\text{mod } s) = K''_i(\text{mod } s) \text{RRot}(\nu_2(K''_{i+1}(\text{mod } s)))$ – циклический сдвиг текущего подключа вправо на количество разрядов, определяемое следующим подключом.

Преобразование массива M :

Сложение текущего и псевдослучайно выбранного (в зависимости от текущего подключа) блока текста с блоком, псевдослучайно выбранным (в зависимости от текущего подключа) из дополнительного массива, циклический сдвиг вправо на количество разрядов, определяемое текущим подключом, и их взаимная перестановка:

- $M_i = \left(M_{\varphi(K'''_i(\text{mod } s))} + G_{\psi_1(K'''_i(\text{mod } s))} \right) \text{RRot}(\pi_1(K'''_i(\text{mod } s)))$
- $M_{\varphi(K'''_i(\text{mod } s))} = \left(M_i + G_{\psi_2(K'''_i(\text{mod } s))} \right) \text{RRot}(\pi_2(K'''_i(\text{mod } s)))$

n последовательных итераций образуют один цикл шифрования. В результате работы одного цикла шифрования каждый информационный блок будет гарантированно подвергнут преобразованию хотя бы один раз. Заполнение ключевого массива в конце работы алгоритма является **ключом расшифрования** K' . Значение $K' \oplus K$ является значением хэш-функции для нашего текста.



3. Программная реализация и тестирование алгоритма

На основе предлагаемой схемы может быть построено множество алгоритмов, параметры которых будут зависеть от конкретных требований. Для достижения высокой скорости работы можно, например, упростить используемые функции, зафиксировать размеры ключа и массива гаммы, а для повышения криптостойкости наоборот – усложнить функции и увеличить число циклов шифрования.

3.1. Реализация и скорость

Один из вариантов предложенной схемы был реализован на языке С в среде разработки приложений Visual Studio 6.0 для работы под управлением операционной системы Windows 9x или Windows NT.

Реализованный вариант алгоритма не был оптимизирован по скорости, но даже в этом случае скорость шифрования/расшифрования составила 10 Мбайт/сек (80 Мбит/сек) для процессора Intel Celeron с тактовой частотой 845 MHz. Скорость определялась путем шифрования/расшифрования 1000 раз одного и того же файла объемом в 1 Мбайт. Благодаря тому, что в алгоритме используются простые (быстрые) операции, возможно значительное увеличение скорости шифрования, путем упрощения используемых функций, при незначительной потере криптостойкости.

Для определения криптографических характеристик алгоритма был проведен ряд тестов.

3.2. Равномерность распределения байтов шифртекста

Проверка равномерности распределения байтов шифртекста производилась с помощью критерия согласия хи-квадрат К.Пирсона. Проведенные тесты показали, что при шифровании текстового и бинарного файлов размера 1 Мб каждый равномерность распределения байтов шифртекста достигается если:

- длина ключа – не менее 5 байтов,
- длина массива гаммы – не менее 45 байтов,
- число главных циклов – не менее 2-х,

причем увеличение параметров сверх указанных не ведет к заметному увеличению равномерности распределения.

3.3. Эффект размазывания

Другой важной характеристикой является эффект размазывания [1]. Проверилось изменение шифртекста в зависимости от изменения 1 бита открытого текста или 1 бита ключа в любой позиции. В идеале должна измениться примерно половина битов шифр-текста. Тест показал $\approx 50\%$ изменение. Тестировались данные различных типов. Позиция измененного бита значения не имеет.

3.4. Имитостойкость

Имитостойкость [2] проверялась путем определения влияния изменения одного бита шифртекста на расшифрованный текст. Разница побитового сравнения файлов после расшифрования, в одном из которых был предварительно изменен 1 бит, составила 50%. Позиция измененного бита значения не имеет.

3.5. Анализ свойств накладываемой гаммы

Так как на блоки текста накладываются блоки, псевдослучайным образом выбранные из массива гаммы (т.е. фактически на текст накладывается гамма), необходимо проанализировать накладываемую последовательность. Тестировалась последовательность размером 3 Кбайта, т.е. 24000 бит, что удовлетворяет требованиям тестов. Использовались пять основных тестов для проверки качества псевдослучайных последовательностей [3], [4]:

- Monobit test
- Two-bit test
- Poker test
- Runs test
- Autocorrelation test

Исследуемая последовательность удовлетворительно прошла все тесты. Это говорит о том, что исследуемая последовательность близка по своим свойствам к случайной и вполне удовлетворяет требованиям данного алгоритма.

Для определения линейной сложности исследуемой последовательности использовался алгоритм Берлекемпа-Мессе [5], [6]. Известно, что для случайной последовательности длины n математическое ожидание линейной сложности равно $n/2$ [7]. Тест показал, что линейная сложность исследуемой последовательности длины n равна $n/2$ почти для всех значений n , причем отклонение от $n/2$ ни разу не превысило 3.

Выводы

1. Впервые предложен метод шифрования, основанный на одновременном преобразовании двух блоков текста, причем вместе с текущим по очереди блоком преобразуется блок, номер которого выбран псевдослучайным образом (в зависимости от текущего состояния рабочего ключа), при этом преобразованные блоки переставляются внутри массива большего размера.
2. Рабочий ключ изменяется после обработки каждого блока текста, причем его изменение зависит от самого текста и предыдущего значения рабочего ключа. Раскрытие подключа на одном из тактов работы не определяет даже части ключевого массива, так как в дальнейшем он пересчитывается.
3. Изменение одного бита открытого текста ведет к изменению рабочего ключа и, следовательно, к изменению всех последующих преобразований.
4. Конечное значение рабочего ключа является ключом расшифрования. Ключ расшифрования зависит от ключа шифрования и открытого текста. Для каждого открытого текста он свой. Ключ расшифрования сложенный с ключом шифрования (или преобразованный иным способом) играет роль хэш-функции.
5. Предложенный метод шифрования с изменением рабочего ключа в зависимости от шифруемых данных, перестановкой блоков данных и последующим наложением гаммы из-за своих особенностей, отличающих его от большинства стандартных алгоритмов, значительно усложняет задачу криптоанализа. Поскольку алгоритм осуществляет перемешивание информации в пределах большого массива данных, становятся практически неприменимыми статистические подходы к анализу шифрующих алгоритмов, в частности разностный (дифференциальный) и линейный криптоанализ.
6. Поскольку для опробования ключа потребуется расшифровать (или зашифровать) весь массив данных, возможность машинного перебора секретных ключей существенно осложнится, что служит защитой от атаки «грубой силой».
7. Частичное совпадение двух открытых текстов не дает совпадения частей соответствующих шифр-текстов. Изменение открытого текста даже в одном бите ведет к лавинному изменению шифр-текста ($\approx 50\%$), причем лавинный эффект происходит не на уровне одного блока, а на уровне всего текста.
8. Тестированием алгоритма шифрования установлено его соответствие криптографическим требованиям для применения в различных информационных системах, од-

нако новизна предлагаемого метода шифрования требует в будущем его всестороннего анализа (и, при необходимости, дальнейшего совершенствования).

Таким образом:

- Данный алгоритм может использоваться в качестве симметричного алгоритма шифрования. Благодаря своим качествам, а именно: универсальности, скорости, простоте реализации, он может использоваться для построения широкого спектра защищенных информационных систем с различными требованиями к криптографическим средствам защиты.
- Этот алгоритм можно также использовать для организации симметричной криптосистемы с сеансовыми ключами, что повышает защищенность информационной системы. В качестве ключа зашифрования сеанса может использовать, например, ключ расшифрования предыдущего сеанса.
- Описанный алгоритм может использоваться для выработки хэш-функции. На его основе может быть создана система электронной цифровой подписи.

Литература

1. Preneel B., van Leekwijk W., van Linden L., Govaerts R., Vandewalle J. Propagation characteristics of boolean functions. Proc. Eurocrypt-90. Lect. Notes in Comp. Sci., v.473, p.161-173 (1991)
2. Месси Дж.Л. Введение в современную криптографию. ТИИЭР, т.76, №5, с.24-42 (1988)
3. Menezes A.J., van Oorschot P.C., Vanstone S.A. Handbook of applied cryptography. CRC Press, 1996
4. Варфоломеев А.А., Жуков А.Е., Пудовкина М.А. Поточные криптосистемы. Основные свойства и методы анализа стойкости. М., МИФИ, 2000
5. Massey J. Shift-register synthesis and BCH decoding. IEEE Trans. Inf. Th., v.IT-15, №1, p.122-127 (1969)
6. Лидл Р., Нидеррайтер Г. Конечные поля. М., Мир, 1988
7. Rueppel R. Analysis and design of stream ciphers. Springer-Verlag, 1986