УДК 621.391:519.72

Н.Ф. Казакова[1], Ю.В. Щербина[1]

[1]*Одесский государственный экономический университет , г. Одесса*

ПРОБЛЕМЫ ПОСТРОЕНИЯ КОМБИНИРОВАННЫХ ЛИНЕЙНЫХ ГЕНЕРАТОРОВ ПСЕВДОСЛУЧАЙНЫХ ЧИСЕЛ

Проведен анализ проблем, связанных с теоретическим и практическим обоснованием принципов построения комбинированных генераторов псевдослучайных последовательностей на основе регистров сдвига с обобщённой обратной связью.
Ключевые слова: потоковый шифр, комбинированный генератор, фильтр с памятью, квазигрупповой фильтр, период, распределение.

Постановка проблемы в общем виде и ее связь с важными научными и практическими задачами

В настоящее время генераторы псевдослучайных последовательностей (ПСП) находят самое широкое применение в задачах криптографии и моделирования. Они являются важной частью многих криптографических систем, к числу которых относятся формирование ключей, инициализация, генераторов, шифрование сообщений и одноразовых блокнотов и многое другое. В моделировании такие генераторы применяются достаточно давно и эффективно. Что же касается криптографических задач, то здесь требования к равномерности распределения вероятностей формируемых чисел значительно выше и этим определяется тот факт, что в настоящее время в этой области появляется все большее число новых идей и подходов.

Постановка проблемы в общем виде

Для решения проблем, связанных с формированием ПСП, использовались различные методы, среди которых наиболее значимыми можно выделить генераторы, в основу которых заложены сложные нелинейные преобразования и генераторы, построенные на основе линейных сдвиговых регистров с обратной связью (*linear feedback shift register* – LFSR). Первые являются более криптостойкими, однако их реализация требует использования трудоемких, с точки зрения вычислительных ресурсов, операций, что существенно снижает оперативность их работы. Вторые более экономичны, поскольку для их реализации применяются в основном сдвиговые, логические и линейные операции. Они экономно расходуют вычислительные ресурсы и, главное, обеспечивают высокую производительность формируемых случайных чисел. Однако, для обеспечения заданной стойкости, в состав таких генераторов приходится все же вводить дополнительные нелинейные функции и, в таком случае, они представляют собой некоторый инженерный компромисс между обоими подходами.

Анализ исследований и публикаций

Актуальность задачи, связанной с построением генераторов ПСП, отвечающих современным требованиям, привела к тому, что европейским криптологическим сообществом ECRYPT был объявлен открытый конкурс (2004-2008 г.г.) на разработку новых потоковых шифров – eSTREAM (ECRYPT Stream Cipher Project) [1]. Его целью было

выявление наиболее достойного соискателя на использование в качестве стандарта для стран европейского сообщества.

История создания генераторов ПСП началась давно. Начало теоретическому обоснованию методов их построения и оценки было положено в [2]. В ней рассмотрены свойства основных алгоритмов формирования псевдослучайных чисел, основанных на рекуррентных соотношениях, и обоснованы способы их статистического тестирования. При этом сделан вывод о том, что ни один из рассмотренных простых алгоритмов не обеспечивает приемлемой равномерности распределения вероятностей чисел, формируемых на выходе генератора. Именно поэтому дальнейшая история создания таких генераторов развивалась по пути создания комбинированных алгоритмов.

В работе [3] был дан подробный анализ проблем, связанных с объединением алгоритмов, построенных на основе LFSR, по сути являющихся линейными рекуррентными алгоритмами. Что же касается наиболее известных нелинейных алгоритмов, то была дана лишь краткая аннотация. К генераторам первого типа может быть применена некоторая общая математическая теория. В основу алгоритмов второго типа их авторы, как правило, закладывали обособленные математические задачи. Поскольку в настоящее время экономия вычислительных ресурсов, надежная криптостойкость и высокая производительность формирования псевдослучайных чисел являются наиболее актуальными задачами, разработчики потоковых систем шифрования все чаще склоняются к построению алгоритмов, сочетающих достоинства линейных и нелинейных преобразований.

Цель статьи – определение проблем, связанных с проектированием и оценкой качества рекуррентных генераторов ПСП комбинированного типа, объединяющих рекуррентные способы формирования шифрующих последовательностей с нелинейной фильтрацией выходного потока.

Изложение основного материала

Исторически так сложилось, что большинство потоковых шифров основаны на использовании линейных сдвиговых регистров с обратной связью. В те времена, когда вычислительная техника была дорогой и малодоступной, технически реализовать такой шифр было несложно. Сдвиговый регистр – это массив битов (ячеек памяти), а обратная связь – это набор сумматоров по модулю два (операция «исключающее ИЛИ»). Однако, как показано в [3], проблема заключается в том, что их программная реализация не особенно эффективна. При выборе образующих полиномов приходится избегать разреженных многочленов обратной связи, которые облегчают противнику их корреляционное вскрытие. Учитывая, что выход потокового шифра является побитовым, такой известный алгоритм как DES, за одну итерацию шифрует столько же текста, сколько потоковый шифр шифрует за 64 итерации. Здесь важна цена итерации, так что в каждом конкретном случае следует проводить соответствующую оценку.

До недавнего времени основной подход к проектированию потоковых шифров сводился к выбору нескольких LFSR с различными длинами многочленов обратной связи. Причем длины выбирались взаимно простыми, а многочлены – примитивными, что обеспечивало наибольшую длину периода формируемой последовательности. При этом функция, определяющая порядок формирования выходных символов, называется *комбинирующей функцией*, а генератор в целом – *комбинационным генератором*. В случае если генератор не является комбинационным и состоит из единственного LFSR, то его называют *фильтрующим генератором*.

Важным параметром, характеризующим составной генератор, является его линейная сложность (*linear complexity*). Она определяется как длина n самого короткого LFSR, который может имитировать выходную последовательность генератора. Любая последовательность, формируемая конечным автоматом над конечным полем и имеет конечную линейную сложность. Желательно, чтобы для достижения большей криптостойкости, комбинирующая функция составного генератора была сложной и нелинейной.

Однако возможна ситуация, при которой выходы отдельных LFSR – могут быть связаны общим ключевым потоком и вскрыты при помощи аппарата линейной алгебры. Атаки такого типа называют *корреляционным вскрытием*. В работе [4] показано, что можно точно определить корреляционную независимость, и что существует компромисс между *корреляционной независимостью* и *линейной сложностью*.

В области проектирования генераторов ПСП появляется все больше идей, основанных не на объединении разных LFSR, а на дополнительном нелинейном преобразовании выходного потока, получаемого на основе некоторого рекуррентного алгоритма. В соответствии с этим принципом, основной генератор, обладающий высокой производительностью, формирует по возможности длинную ПСП, а дополнительный фильтр преобразует ее с целью повышения криптографической стойкости. Объяснение такого подхода базируется на том факте, что современные вычислительные системы обладают ресурсами, способными к высокоскоростному целочисленному умножению. Линейные регистры с битовыми ячейками давно ушли в прошлое и в практических приложениях используется программная реализация, заложенного в них принципа. Однако в качестве ячеек регистра рассматриваются не битовые ячейки, а блоки памяти, равные по размеру величине машинного слова w. Как правило, $w = 32$. Такие генераторы называют регистрами сдвига с обобщённой обратной связью (*generalized feedback shift register* GFSR). Примерами генераторов ПСП, построенных по этому принципу, могут служить генератор Фибоначчи с запаздыванием [2] и «вихрь Мерсенна» (*Mersenne Twister* MT) [5].

Для удобства анализа всякий рекуррентный генератор может быть представлен как автомат с памятью $A = (S, F, O, o)$ с конечным числом состояний без входа, где S – конечное множество его состояний, отображение $f : S \to S$ – это функция переходов из текущего состояния в следующее, O – набор символов выходного алфавита и $o : S \to O$ – это выходная функция, отображающая его внутренние состояния в символы выходного алфавита. Вид формируемой последовательности определяется начальным состоянием автомата s_0, а переходы в следующие состояния происходят в соответствии с рекуррентным соотношением $s_i = f(s_{i-1})(n = 1, 2, 3, ...)$, при этом $o(s_0), o(s_1), o(s_2), ... \in O$.

Наличие отдельного выходного алфавита O объясняется тем, что, как показано в [2], не все символы на выходе генератора одинаково «случайны» и, поэтому, часто, для выравнивания распределения формируемой последовательности, отфильтровывают только старшие наиболее значимые разряды выходных слов (*Most Significant Bit* – MSB).

Задание начального состояния GFSR генератора – отдельная проблема. Так, например, MT-генератор требует заполнения 623 w-битных ячеек памяти и, в этом случае, приходится использовать отдельный генератор инициализации, формирующий на основе некоторой функции инициализации $\text{init}: K \to S$ значение $s_0 = \text{init}(k_i)$ из ключа $k_i \in K$, где K – пространство ключей. Обычно эта проблема решается путем выбора линейной функции перехода в двоичном поле Галуа $GF(2)$.

Чтобы получить безопасный генератор, с наибольшим периодом T, желательно, чтобы функции f и o были достаточно сложными. Однако для сложной функции f, анализ ее периода и распределения вероятностей в выходной последовательности представляет собой сложную задачу. Поэтом сначала определяют пространство состояний генератора как $S = GF\left(2^w\right)^n$, где n – степень образующего полинома GFSR, а, затем, выбирают функцию перехода f, период которой может быть определен методами линейной алгебры в полиномиальном исчислении.

В общем случае для GFSR, построенного на основе образующего полинома степени n, число его внутренних состояний равно $2^{wn} - 1$. Линейное отображение множес-

тва состояний генератора во множество его выходных слов $g: GF\left(2^w\right)^n \to GF\left(2^w\right)$ называют функцией обратной связи, а его переход из одного состояния в другое, эквивалентен рекурсии $x_{i+n} = = g(x_i, x_{i+1}, ..., x_{i+n-1})$, $i = 0,\ 1,\ 2,\ ...$. При этом выход GFSR задается отображением $o: S \to GF\left(2^w\right)$, $(x_1, ..., x_n) \to x_1$, которое не является секретным.

Программная реализация генераторов на основе GFSR с применением технологии циклических массивов [2], суть которой сводится не к перемещению данных в ячейках массива, а к вычислению новых значений индексов этих ячеек через значение функции g, позволяет сделать их достаточно высокоскоростным. При этом, величина n не оказывает влияния на скорость генератора. Однако, любая рекуррентная последовательность, порождаемая потоковым генератором достаточно уязвима, поэтому имеет смысл ввести в алгоритм некоторую нелинейность. Обычно эта задача решается выбором некоторого нелинейного представления $o: S \to O$. В этом контексте, выходная функция o называется фильтром.

Оценка нелинейности функции определяется ее алгебраической степенью, под которой понимается полиномиальная булева функция $h(c_1, c_2, ..., c_n)$ с переменными из поля $GF(2)$, определяемая как $h: GF\left(2^n\right) \to GF(2)$ или $h = \sum_{i \subset \{1,2,...,n\}} a_i c_i$, где $a_i \in GF(2)$, а c_i – это состояния ячеек памяти GFSR.

Пусть, например, $h_{i,n}(s_0)$ обозначают i-й бит на n-ном выходе генератора порождаемый из начального состояния $s_0 \in S = GF\left(2^w\right)$. Это значит, что злоумышленник может получить s_0 в результате решения системы уравнений $h_{i,n}(s_0) = o_{i,n}$ для неизвестного s_0 изменяя i и n. В данном случае $o_{i,n}$ – это наблюдаемый противником выход генератора. Алгебраические атаки такого типа описаны в [6].

Для достижения высокой производительности генератора, преобразованию o_i не достаточно количества доступных бит, составляющих состояния s_i и, кроме того, алгебраическая степень такого преобразования, также ограничивается числом доступных бит. Это уменьшает достоинства большого пространства состояний S. Проблема решается введением в состав генератора дополнительного конечного автомата с входом, представляющего собой фильтр с памятью.

В отличие от автомата без входа, конечный автомат с входом может быть представлен в виде кортежа $A = (S, U, f, O, o)$, где компоненты S, f, O и o имеют те же значения, что и в случае автомата без входа, а U представляет собой множество символов входного алфавита. При этом функция переходов имеет вид $f: U \times S \to S$. При начальном состоянии s_0 и входной последовательности $u_0, u_1, ... \in U$, изменения внутреннего состояния автомата с входом определяется рекурсией $s_i = f\left(u_{j-1}, s_{j-1}\right)$, $(i = 1, 2, 3, ...)$.

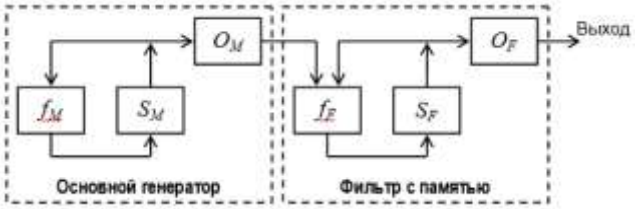

Рис. 1 – Комбинированный генератор с фильтром

Формально, комбинированный генератор с фильтром можно описать следующим образом. Пусть автомат без входа $A_M = (S_M, f_M, O_M, o_M)$ – это основной генератор на основе GFSR, порождающий некоторую ПСП. Автомат с входом $A_F = (S_F, U_F, f_F, O_F, o_F)$ будем называть фильтром с памятью. Поскольку символы ПСП поступают на вход фильтра, алфавиты O_M и U_F совпадают, то есть $O_M = U_F$. Для формирования выходной последовательности следует инициализировать и основной генератор и фильтр. Для этого требуется пара начальных состояний $s_{M,0} \in S_M$ и $s_{F,0} \in U_F$. Далее, фильтр A_F преобразует выходную последовательность основного генератора o_M в собственную выходную последовательность o_F. В итоге, вся эта конструкция в целом формально может быть представлена как автомат C без входа, называемый комбинированным генератором. При этом пространство внутренних состояний такого генератора составляет $S_M \times S_F$, переходная функция имеет вид:

$$f_C : (s_M, s_F) \to \left(f_M(s_M),\, f_F\left(o_M(s_M), s_F\right)\right),$$

а выходная функция:

$$o_C : (s_M, s_F) \to o_F(s_F) \in O_F.$$

В приведенном выше определении функция o_C зависит только от пространства состояний фильтра S_F, но, в принципе, она может быть представлена как функция от общего числа состояний S_M и S_F.

Характерным примером шифра, построенного в соответствии с описанным принципом, является шифр SNOW2.0 [7, 8]. Он представляет собой объединение LFSR-генератора с 512-битным пространством состояний и фильтра, содержащего четыре 8-битных S-блока, обеспечивающих выполнение арифметических операций в поле $GF\left(2^8\right)$.

Для удобства описания фильтров с памятью, введем понятие квазигруппового фильтра, под которым будем понимать автомат с входом и конечным числом состояний, функция переходов которого является двойной биекцией $f : U \times S \to S$.

Пусть, например, $U = S$ множество нечетных чисел в кольце $Z/2^{32}$ целых чисел, приведенных по модулю 2^{32} и пусть $f : U \times S \to S$ целочисленная операция умножения по модулю 2^{32}. Это мультипликативная группа в кольце $Z/2^{32}$, которая удовлетворяет введенному определению квазигруппы. Функция выхода для нее может быть представлена как отображение $o_F : S \to O_F$, выделяющее 8 старших бит из 32-разрядного целого числа. Реализация подобных высокоскоростных фильтров с памятью, осуществляющих целочисленное умножение 32-разрядных целых чисел, для современных процессоров особой сложности не представляет.

Период недвоичного генератора GFSR достаточно велик, и предварительная его оценка может быть выполнена следующим образом.

Пусть C – это скомбинированный генератор A_M и пусть $s_{M,0}$ его начальное состояние. Предположим, что смена состояний генератора происходит с периодом $T = Qq$ для простого Q и целого q. Пусть при этом S^O составляет часть множества состояний основного генератора S_M, то есть $S^O \subset S_M$, и пусть k целое число. Предположим далее, что k-элементный кортеж выходной функции основного генератора $o_M^{(k)} : S^O \to O_M^k$ сюръективен при ограниченном S^O.

Предположим так же, что A_F – квазигрупповой фильтр составного генератора C и r – максимальный прообраз некоторого элемента из: $o_F : S_F \to O_F$ в S_F, а именно:

$$r = \max_{b \in O_F} \frac{\#\left(o_F^{-1}(b)\right)}{\#\left(S_F\right)}.$$

Если $r^{-(k+1)} > q \cdot \left(\#S_F\right)^2$, то период выходной последовательности генератора C ненулевой и кратен Q. Доказательство этого утверждения приводится в [9].

Наконец, при создании комбинированных генераторов следует помнить, что возможны ситуации, когда шифруемое сообщение по своему объему меньше величины внутреннего состояния составного генератора. При этом комбинированный генератор становится неэффективен. Возможным решением этой проблемы может быть отдельный загрузчик на основе LFSR с малым числом состояний, выходы которого подключаются непосредственно к входам выходного фильтра комбинированного генератора C. В этом случае он просто заменяет основной генератор. В обычной ситуации, когда длина сообщения больше объема внутреннего состояния генератора, используется стандартный способ его инициализации. Причина такого разделения кроется в том, что при использовании стандартной схемы не приходиться беспокоится о предотвращении атак на загрузчик.

Выводы

На текущий момент одним из наиболее перспективных способов построения составных генераторов, на основе GFSR, является их компромиссное сочетание с нелинейными мультипликативными фильтрами, обладающими собственной конечной памятью. Применение такого подхода обосновывается тем фактом, что при достаточно большом периоде формируемых GFSR-генераторами последовательностей, число порождающих полиномов, определяющих тип обратной связи относительно невелико, что облегчает задачу криптоаналитикам. Введение же некоторой нелинейности – есть экономичный способ решения проблемы. Приводимое в статье аналитическое обоснование схемы такого типа генераторов, может быть использовано как рабочий инструмент в процессе их проектирования.

Литература

1. eSTREAM, the ECRYPT Stream Cipher Project [Електронний ресурс] // Портал : без назви. – Режим доступу \www/ URL : http://www.ecrypt.eu.org/stream/index.html. – Заголовок з екрану, доступ вільний, 18.05.2013.
2. Дональд Кнут. Искусство программирования для ЭВМ : підручник : пер. з англ. [Ю. В. Козаченко] / Дональд Кнут. – М. : Мир, 1977. – Т. 2. – 727 с. – ISBN 978-5-8459-0081-4.

3. Брюс Шнайер. Прикладная криптография. Протоколы, алгоритмы, исходные тексты на языке Си : монографія : пер. з англ. / Брюс Шнайер. – М. : Триумф, 2002. – 816 с. – ISBN 5-89392-055-4, 0-471-11709-9.
4. Siegenthaler, T. Gorrelation-Immunity of Nonlinear Combining Functions for Cryptographic Applications [Текст] / T. Siegenthaler // IEEE Transactions on Information Theory. – 1984. – №5. – V. IT-30. – P. 776-780. – ISSN відсутній.
5. Matsumoto, M. Mersenne Twister: a 623-dimensionally equidistributed uniform pseudo-random number generator [Текст] / M. Matsumoto, T. Nishimura // ACM Trans. on Modeling and Computer Simulation. – 1998. – №8. – P. 3-30. – ISSN відсутній.
6. Courtois, N. Fast algebraic attacks on stream ciphers with linear feedback Advances in Cryptology [Текст] / N. Courtois // CRYPTO-2003. – Springer-Verlag. – 2003. – №2729. – P. 176-194. – ISSN відсутній.
7. Ekdahl, P. SNOW-a new stream cipher [Текст] / P. Ekdahl, T. Johansson // Proc. of First Open NESSIE Workshop. – KU-Leuven, 2000. – 230 p. – ISSN відсутній.
8. Ekdahl, P. A New Version of the Stream Cipher SNOW [Текст] / P. Ekdahl, T. Johansson // Selected Areas in Cryptography. – Springer Verlag : LNCS. – 2002. – №2595. – p. 47-61. – ISSN відсутній.
9. Matsumoto, M. A Fast Stream Cipher with Huge State Space and Quasigroup Filter for Software [Текст] / Matsumoto M., Saito M., Nishimura T., Hagita M // Selected Areas in Cryptography. – Springer Verlag : LNCS. – 2007. – №4876. – p. 246-263. – ISSN відсутній.

Надійшла до редколегії 17.02.2013 р.

Рецензент: д.т.н., проф. Петров А.С.

Н. Ф. Казакова, Ю. В. Щербина
ПРОБЛЕМИ ПОБУДОВИ КОМБІНОВАНИХ ЛІНІЙНИХ ГЕНЕРАТОРІВ ПСЕВДОВИПАДКОВИХ ЧИСЕЛ

Проведено аналіз проблем, пов'язаних з теоретичним і практичним обґрунтуванням принципів побудови комбінованих генераторів псевдовипадкових послідовностей на основі регістрів зсуву з узагальненим зворотнім зв'язком.

Ключові слова: потоковий шифр, комбінований генератор, фільтр з пам'яттю, квазігруповий фільтр, період, розподіл.

N.F. Kazakova, Yu.V. Shcherbina
THE PROBLEM OF CONSTRUCTING A HYBRID LINEAR PSEUDO RANDOM NUMBER GENERATOR

The analysis of the problems associated with the theoretical and practical principles of justification of combined pseudo-random sequence generator based on the generalized shift registers with feedback.

Keywords: stream cipher, combined generator, filter with memory, quasigroup filter, multiplicative filter, CryptMT, period, distribution.