11. Попов А.А. Информационные соотношения между элементами пространства сигналов, построенного на обобщенной булевой алгебре с мерой// Вісник Державного університету інформаційно-комунікаційних технологій.-2007.-5, №2.-С. 175-184.
12. Биркгоф Г. Теория решеток. М.: Наука, 1984. — 568 с.
13. Тихонов В.И. Статистическая радиотехника. М.: Радио и связь, 1982. – 624 с.
14. Левин Б.Р. Теоретические основы статистической радиотехники. В 3-х т. Т.1. М.: Сов. радио, 1969. – 752 с.
15. Прудников А.П., Брычков Ю.А., Маричев О.И. Интегралы и ряды. В 3-х т. Т.1. Элементарные функции. М.: ФИЗМАТЛИТ, 2002. – 632 с.

Рецензент: Скрипник Л.В.
Поступила 12.12.2011

УДК 519.254

Борода А.В.
ГУИКТ

ПРИНЦИПЫ ПОСТРОЕНИЯ ПОТОЧНЫХ ШИФРСИСТЕМ И ПОДХОДЫ К ОЦЕНКЕ ИХ СТОЙКОСТИ

Современная шифрсистема (далее – ШС) представляет собой электронное устройство, которое осуществляет шифрование (расшифрование) информации и реализует криптопротокол согласования ключей криптографических алгоритмов приемной и передающей стороны. Для обеспечения достаточной гибкости в плане возможности изменения криптографических параметров, при технической реализации шифрсистем широко используется комбинация soft- и hardware технологий, предусматривающая выполнение части криптографического алгоритма аппаратным способом, а часть программными средствами. Обычно аппаратным способом реализуется *основа* криптографического алгоритма, то есть та часть, которая непосредственно производит шифрование и расшифрование данных, а предварительное преобразование ключевой информации при формировании *рабочего ключа* криптоалгоритма перед началом шифрования (расшифрования) выполняется программно.

Элементами *рабочего ключа* являются переменные параметры криптографического алгоритма, которые зависят от *ключевой установки* шифрсистемы. Ими могут быть переменные коммутаторы и подстановки, двоичные функции, таблицы замены, начальные заполнения регистров генератора гаммы и т.д. Под *ключевой установкой* обычно понимается совокупность *секретного ключа* и *разового ключа* шифрования (последний передается в открытом виде).

Шифрсистема является *поточной*, если она последовательно преобразует символы открытого текста в символы шифртекста с помощью шифра замены, который зависит от ключа и от внутреннего состояния шифрсистемы (номера такта шифрования). При работе поточной шифрсистемы для каждого шифруемого знака необходимо указать соответствующий ему шифр замены. Поэтому в конструкции поточных шифрсистем обычно выделяют два основных узла. Первый из них осуществляет выбор шифрующего преобразования, а второй выполняет собственно шифрование очередного знака открытого текста. Первый узел вырабатывает последовательность номеров шифрующих преобразований, то есть управляет порядком процедуры шифрования. Вырабатываемую им управляющую последовательность, представляющую собой номера используемых преобразований, обычно называют *управляющей гаммой* или *гаммой шифрования* (run keys), а сам этот узел – *генератором гаммы* (Run Key Generator). Второй узел, в

соответствии со знаком гаммы шифрования, реализует процедуру шифрования текущего знака. Этот узел называется *узлом шифрования* или *узлом криптографического преобразования.*

Наряду с перечисленными узлами – генератором гаммы и узлом шифрования, – важную роль в обеспечении криптографической стойкости шифрсистем выполняет *подсистема инициализации* рабочего ключа. Она отвечает за сохранение номинальной мощности ключевой установки при ее преобразовании в рабочий ключ и обеспечение "хороших" криптографических свойств рабочего ключа.

1. Необходимые характеристики подсистемы инициализации рабочего ключа.

Инициализация рабочего ключа обычно разделяется на два этапа. На первом осуществляется определенное преобразование исходной ключевой информации – ее «расширение» в последовательность большего объема. В ходе второго этапа выполняется формирование элементов рабочего ключа.

Первый этап строится таким образом, чтобы соблюдались 2 криптографических принципа Шеннона: «смешения» и «рассеивания». В применении к ключам шифрования принцип «смешения» состоит в том, чтобы изменение любого ключа в незначительном числе позиций приводило к существенному изменению рабочего ключа шифрования. Рассеивание подразумевает влияние каждого знака ключа на множество (все) знаки генерируемой гаммы. Обычно расширитель строится на базе линейного преобразования большой размерности. В ходе «расширения» символы ключа заносятся в элементы расширителя, после чего тот «прокручивается» Ψ тактов. «Расширенная» последовательность образуется путем съема значений с некоторого элемента расширителя в последующих C тактах работы. Из соображений секретности, число тактов прокрутки может выбираться переменным и зависеть от ключа шифрования.

Операция «расширения» ключевой информации используется в тех случаях, когда для построения рабочего ключа требуется больше данных, чем содержится в исходных ключах. По сути такой расширитель представляет собой датчик псевдослучайных чисел, в котором в качестве «зерна» используется ключевая информация. К расширителю предъявляются такие же требования, как и к «хорошему» датчику *псевдослучайных чисел.* Выходная последовательность расширителя должна иметь большой период, чтобы исключить возможность повторений рабочего ключа. Расширенная информация должна обладать максимальной линейной сложностью, чтобы в случае вскрытия отдельных элементов рабочего ключа было затруднено определение других элементов. Это требование тождественно тому, что расширение не порождает эквивалентных ключей, то есть число генерируемых расширителем последовательностей должно совпадать с мощностью ключевой установки.

«Расширенная» последовательность используется при формировании переменных таблиц замены и начальных заполнений регистров криптоалгоритма. Для построения зависящих от ключа шифрования двоичных функций, коммутаторов и подстановок применяются различные детерминированные алгоритмы, которые в качестве исходных данных используют определенные участки расширенной последовательности. Формирование коммутаторов и подстановок сводится к построению перестановок элементов некоторого алфавита.

Основное требование, которое предъявляется к алгоритмам построения элементов рабочего ключа – это соблюдение биективности отображения соответствующих входных данных алгоритма в построенные элементы рабочего ключа. Невыполнение этого требования приводит к неравновероятному распределению определенных частей рабочего ключа, что может быть использовано для создания алгоритмов их направленного

опробования. Если такой алгоритм позволяет снижать сложность опробования в определенное число раз, тогда можно говорить о коэффициенте снижения стойкости соответствующих элементов рабочего ключа.

2. Особенности построения генератора гаммы.

Генератор гаммы поточной шифрсистемы представляет собой автономный автомат с конечным числом состояний, который в каждом такте работы вырабатывает знак гаммы шифрования в зависимости от рабочего ключа и своего внутреннего состояния. Генерируемая гамма должна удовлетворять ряду основных требований: обладать большим периодом, большой линейной сложностью и равномерным распределением.

Работа любого автономного автомата с заданным множеством состояний определяется функцией переходов из состояния в состояние, функцией выхода и начальным состоянием. В генераторе гаммы вид этих двух функций, а также начальное состояние зависит от рабочего ключа. При конструировании генератора гаммы осуществляется выбор таких функций перехода и выхода, а также способа формирования рабочего ключа, которые гарантируют, что генерируемая генератором гамма будет удовлетворять основным требованиям.

Для этого генераторы гаммы строятся в соответствии с 2-мя криптографическими принципами Шеннона: смешение и рассеивание. Смешение реализуется путем использования нелинейных преобразований усложнения. Рассеивание подразумевает влияние каждого знака ключа на множество (все) знаки гаммы. Таким образом, конструкция генератора гаммы подразумевает использование нелинейных преобразований, которые значительно усложняют проведение раздельной криптоаналитической атаки на узлы генератора.

Нелинейность может вводиться явным образом путем задания нелинейных функций перехода и (или) выхода. Когда нелинейной является функция перехода, возможности по расчету характеристик генератора гаммы являются весьма ограниченными, поскольку теоретически этот подход разработан слабо. Достаточно хорошо исследованы те автоматы, у которых функция перехода является линейной, а нелинейной будет только функция выхода. В этом случае принято конструктивно разделять генератор гаммы на две подсистемы: *задающий генератор* (driving part) и *блок усложнения* (combining part), [2].

Задающий генератор продуцирует последовательность состояний с большим периодом и хорошими статистическими свойствами. К примеру, задающий генератор может состоять из совокупности линейных регистров сдвига максимального периода.

Блок усложнения обеспечивает большую линейную сложность гаммы без потери хороших вероятностных свойств с тем, чтобы сделать невозможными любые линейные криптоаналитические атаки типа процедуры Берлекемпа-Месси. Работа блока усложнения может быть представлена в виде некоторой нелинейной функции усложнения F. Часто является оправданным введение в состав блока усложнения элементов памяти, что превращает его в конечный автомат с входом. Это эквивалентно значительному увеличению количества аргументов функции усложнения F.

3. Требования к криптографическому преобразованию.

Криптографические преобразования, которые используются для шифрования, должны обладать рядом специальных качеств. Одно из них, известное как эффект «лавинности», сформулировано К. Шенноном в [1]. Суть его состоит в том, что применение шифрующего преобразования к наборам аргументов, отличающихся в незначительном числе позиций, должно приводить к существенному изменению результата. Для повышения криптографической стойкости шифра часто используется

способ, основанный на последовательном применении к шифруемому символу различных преобразований. В одной из работ Месси [3] доказано существование совершенно стойкого поточного шифра над $GF(q)$, в котором для шифрования каждого знака открытого текста используется k случайных независимых между собой знаков гаммы. Существует предположение, что шифр будет совершенно стойким уже при использовании 2 знаков гаммы на каждый знак открытого текста.

От криптографических преобразований, которые реализуются блочными шифрами, обычно требуется, чтобы длина двоичного представления блока шифртекста равнялась длине исходного блока открытого текста. Алфавитом, на котором действует блочный шифр, служит множество двоичных векторов-блоков одинаковой длины. Построение преобразований больших алфавитов является сложной задачей и вступает в противоречие с принципом простоты реализации шифра. Один из выходов из этой ситуации указал К. Шеннон [1], предложив реализовывать сложные преобразования в виде суперпозиции нескольких базовых некоммутируемых отображений. Обычно используются базовые преобразования двух типов: сложные в криптографическом отношении локальные преобразования над отдельными частями шифруемых блоков и простые преобразования, переставляющие части шифруемых блоков. Первые преобразования получили название *перемешивающих*, а вторые – *рассеивающих*. Простейшим примером реализации преобразования перемешивания является подстановка. Перемешивание усложняет восстановление взаимосвязи статистических и аналитических свойств открытого и шифрованного текстов, а рассеивание распространяет влияние одного знака открытого текста на большое число знаков шифртекста, что позволяет сгладить влияние статистических свойств открытого текста на свойства шифртекста.

4. Подходы к оценке стойкости поточных шифрсистем.

Центральным понятием современной криптографии является понятие *стойкости*, под которым понимают способность шифрсистем противостоять всевозможным видам нападений (криптоаналитических атак с применением тех или иных методов криптоанализа) [1,3]. Известны два подхода к анализу стойкости существующих шифрсистем [1,3,4], основанные на следующих положениях:

– доказательстве *теоретической* (*безусловной* или *совершенной*) *стойкости* шифрсистемы, то есть справедливости утверждения, что вскрытие ШС противником, обладающим неограниченными вычислительными ресурсами, принципиально невозможно;

– исследовании *практической* (*вычислительной*) *стойкости* шифрсистем, которая определяется вычислительной сложностью решения криптоаналитических задач по вскрытию ключей шифрования.

В соответствии определением К. Шеннона [1], шифрсистема является совершенно стойкой или безусловно стойкой (unconditionally secure), если взаимная информация между открытым текстом сообщения и соответствующим шифртекстом равна нулю независимо от длины сообщения. Этому условию, к примеру, удовлетворяют шифрсистемы на основе одноразового шифра Вернама [1,3]. Однако на практике одноразовые шифрсистемы сложны в использовании, так как требуется физическая рассылка больших объемов идентичной шифрующей гаммы на все узлы связи.

Практическая невозможность широкого использования одноразовых шифров привела к развитию поточных систем шифрования, которые зашифровывают открытые тексты таким же образом, как и одноразовые системы, но применяют для этого детерминировано сгенерированную псевдослучайную последовательность гаммы. Именно такие шифрсистемы служат предметом криптоаналитических исследований при решении

задач дешифрования. По этой причине наиболее существенным является оценивание их практической стойкости.

Стойкость используемых на практике ШС характеризуется [illegible], то есть определяется трудоемкостью лучшего из известных на сегодняшний день алгоритмов криптоанализа, которую принимают в качестве оценки практической стойкости шифрсистемы [1,3,4]. Один из подходов к анализу стойкости состоит в представлении криптоалгоритма ШС в виде детерминированного шифрующего автомата с конечным числом состояний. На вход автомата поступает последовательность знаков открытого текста, а с выхода снимаются знаки шифрованного текста. Всегда можно считать, что входная и выходная последовательности являются двоичными, поскольку символы открытого и шифрованного текстов представляются с помощью соответствующих двоичных кодов, то есть шифрование выполняется двоичными блоками. Функцию перехода и функцию выхода такого автомата можно представить в виде многомерных булевых функций от $(N+M)$ переменных, где N – длина двоичного блока открытого текста, а M – объем памяти автомата в битах. Стойкость криптографического алгоритма шифрсистемы определяется свойствами этих булевых функций. Поскольку число переменных $(N+M)$ достаточно большое (>100), то проведение непосредственного анализа функции перехода и функции выхода является невозможным. Выход заключается в применении метода декомпозиции к шифрующему автомату. В результате можно перейти к более простым логически завершенным узлам преобразований, проанализировать свойства этих узлов, их взаимосвязи и, затем, сделать вывод о свойствах и стойкости всей ШС. Выделение «логических» узлов криптоалгоритма и исследование этих узлов является необходимым этапом анализа стойкости любой шифрсистемы.

Таким образом, основными узлами рассматриваемых шифрсистем является узел шифрования, генератор гаммы и подсистема формирования рабочего ключа, характеристики и стойкость которых относительно методов опробования ключей, аналитических и смешанных методов криптоанализа определяют практическую стойкость всей шифрсистемы.

Показателями практической стойкости шифрсистем относительно методов опробования ключей является эффективная мощность ключевой установки и коэффициенты снижения стойкости элементов рабочего ключа относительно метода направленного перебора в соответствие с вероятностями их построения.

Практическая стойкость шифрсистем относительно алгебраических методов криптоанализа определяется длиной периода и рангом (линейной сложностью) гаммы шифрования.

Показателем стойкости поточной шифрсистемы относительно вероятностно-статистических методов криптоанализа является трудоемкость перебора тех элементов рабочего ключа, которые необходимо вскрывать первыми, а стойкость реальных шифрсистем в целом определяется трудоемкостью лучшего из известных алгоритмов криптоанализа, которая и принимается за оценку практической стойкости шифрсистемы.

Литература

1. *Shannon C.E. Communication theory of secrecy systems // Bell System Techn. J. - 28 (1949) - No.4. - P. 656-715. (Имеется перевод: в кн. К. Шеннон Работы по теории информации и кибернетике. - М.: ИЛ, 1963. Теория связи в секретных системах. - С. 333-402.)*

2. *Rueppel R.A. Analysis and Design of Stream Ciphers. Communications and Control Engineering Series. - Springer-Verlag, Berlin Heidelberg, 1986.*

3. *Massey. J. L. An Introdution to Contemporary Cryptology // Proc. IEEE. - 1988. - V. 76. - № 5. - P. 533 - 549.*

4. Молдовян Н. А. Проблематика и методы криптографии. - СПб.: Изд-во СПбГУ, 1998. - 245 с.

Рецензент: Корнейчук М.Т.
Поступила 21.12.2011

УДК 004.021

Петров А.О.
Східноукр. Національ. Унів. Ім..В.Даля

МОДЕЛІ ТА МЕТОДИ РОЗПІЗНАВАННЯ МОВИ

1. Приховані Марківські моделі

Найшвидша та ефективна взаємодія між людьми відбувається за допомогою усної мови. За допомогою мови можуть бути передані різні почуття й емоції, а головне — корисна інформація. Необхідність створення комп'ютерних інтерфейсів звукового введення-виводу не викликає сумнівів, оскільки їх ефективність заснована на практично необмежених можливостях формулювання у всіляких областях людської діяльності.

Перша електронна машина, що синтезує англійську мову, була представлена в Нью-Йорку на торговельній виставці в 1939 році та називалася voder, але звук, який вона відтворювала, був украй нечітким. Перше ж пристрій для розпізнавання мови було створено у 1952 році та був здатен розпізнавати цифри [3].

У процесі розпізнавання мови можна виділити наступні складності: довільний, наївний користувач; спонтанна мова, супроводжувана аграматизмами й мовним «сміттям»; наявність акустичних перешкод і викривлень; наявність мовних перешкод [20].

Із усього різноманіття методів у даній статті ми розглянемо можливість створення статистичної моделі за допомогою прихованих Марківських моделей (ПММ) [1].

Прихована Марківська модель — статистична модель, що імітує роботу процесу, схожого на марківський процес з невідомими параметрами. Завданням ПММ ставиться *визначення невідомих параметрів на основі спостережуваних*. Отримані параметри може бути використано в подальшому аналізі, наприклад, для розпізнавання образів. ПММ може розглядатися як найпростіша Байєсова мережа довіри [1].

При аналізі природньої мови першим кроком необхідно визначити: до якої частини мови ставиться кожне зі слів у пропозиції. В англійській мові завдання на цьому етапі називається Part-Of-Speech tagging. Яким образом ми можемо визначити частину мови окремого члена пропозиції? Розглянемо речення англійською мовою: «The can will rust». Отже, the –певний артикль або частка «тем»; can – може одночасно бути й модальним дієсловом, і іменником, і дієсловом; will – модальне дієслово, іменник і дієслово; rust – іменник або дієслово. У статистичному підході необхідно побудувати таблицю ймовірностей використання слів у кожному граматичному значенні. Це завдання можна розв'язати на основі тестових текстів, проаналізованих вручну. І відразу можна виділити одну із проблем: слово «can» у більшості випадків використовується як дієслова, але іноді воно може бути й іменником. Враховуючи цей недолік, була створена модель, що ухвалює в увагу той факт, що після артикля піде прикметник або іменник:

$$\underset{t1..tn}{\arg\max} \prod_{i=1}^{n} p(w_i|t_i)p(t_i|t_{i-1}) \quad ,(1)$$