Авторами проведений самостійний цикл класифікаційних випробувань. У доповіді представлені деякі результати щодо числа класів та типів при невеликих значеннях n, а також про потужність типів та вказані криптографічні властивості на множині бульових функцій від 3-х аргументів.

Література: 1. Поваров Г. Н. О групповой инвариантности булевых функций. – В сб.: Применение логики в науке и технике. Москва, 1960. 2. Никонов В. Г. Классификация минимальных базисных представлений всех булевых функций от четырех переменных. – В сб.: Обозрение прикладной промышленной математики, серия дискретная математика (1994) 1, выпуск 3. 3. W. Meier and O. Staffelbach, "Nonlinearity criteria for cryptographic functions", LNCS 434; Proc. Eurocrypt'89.

УДК 681.511:3

ЧАСТНЫЙ ПОДХОД К ВОПРОСУ ИДЕНТИФИКАЦИИ ПРЕОБРАЗУЮЩЕГО АЛГОРИТМА

Иван Четвериков, Александр Манухин, Светлана Паламарчук
Военный институт Национального технического университета Украины "КПИ"

Анотація: **Частковий підхід до питання ідентифікації перетворюючого алгоритму. Аналіз закритих потоків даних припускає наявність математичної моделі ідентифікації шифратора. Проблема вирішена за допомогою використання спектрального аналізу потоку і побудови профілю потоку.**
Summary: **The private approach to a problem of identification of conversing algorithm. The analysis of the enclosed data-flow guesses presence of the mathematical model of identification of an encoder. The problem is solved by means of a spectrum analysis of a stream and introduction of a profile of a stream enclosed by the relevant algorithm.**
Ключові слова: **захист даних, шаблон, спектральний аналіз.**

I Постановка задачи исследования

Проблема диагностики символьных потоков различных форм (звукового, литерного, изображения) крайне актуальна на различных уровнях управления процессом их обработки. Учитывая направленность деятельности некоторых ведомственных структур, интерес представляет идентификация закрытых потоков {T`}, представляющих собой результат применения криптографических алгоритмов (A):

$$T \to T`, \quad T`= A(T); \qquad \{T`\}_n = \{A(T)\}_n. \tag{1}$$

Закрытый поток данных достаточно просто локализуется методами математической статистики, однако идентификация почерка криптографического алгоритма представляет собой некоторые сложности.

Данная работа посвящена идентификации преобразующего алгоритма (вида специальной аппаратуры), порождающего закрытые потоки {T`} методом их спектрального анализа [2-6].

II Схема информационной обработки потоков данных

Открытая передача закрытых потоков данных по каналам связи осуществляется соответственно функциональной схеме обработки сообщения, тракт прохождения которого представлен на рисунке 1. Он представляет собой набор узлов обработки (1-7), имеющих входные и выходные параметры, а также условия принятия решения. Узлы имеют 1 и 2 входа, 1, 2 и 3 выхода.

Узел № 1 – преобразующий алгоритм источника; узел № 2 – информационный ключ; узел № 3 – стандартизатор потока; узел № 4 – спектральный формирователь; узлы № 5, 7 – спектральные фильтры; узел № 6 – спектральный анализатор. Среду формирования составляют узлы 3, 4; среду исследования – 6 и 7.

Диагностический контур реализован спектральным формирователем объекта исследования, математическим сопроцессором и узлами принятия решения и рассматривается человеко-машинным комплексом "обучение с учителем", обучаемым параметром которого являются характеристика потока.

Работа схемы представляет собой два цикла. Первый цикл осуществляется при нулевых начальных условиях и служит для первичного заполнения спектральных фильтров (условия принятия решения). Цикл № 2 реализует полномасштабную систему анализа и идентификации выбранного потока данных. Рассматриваются раздельно.

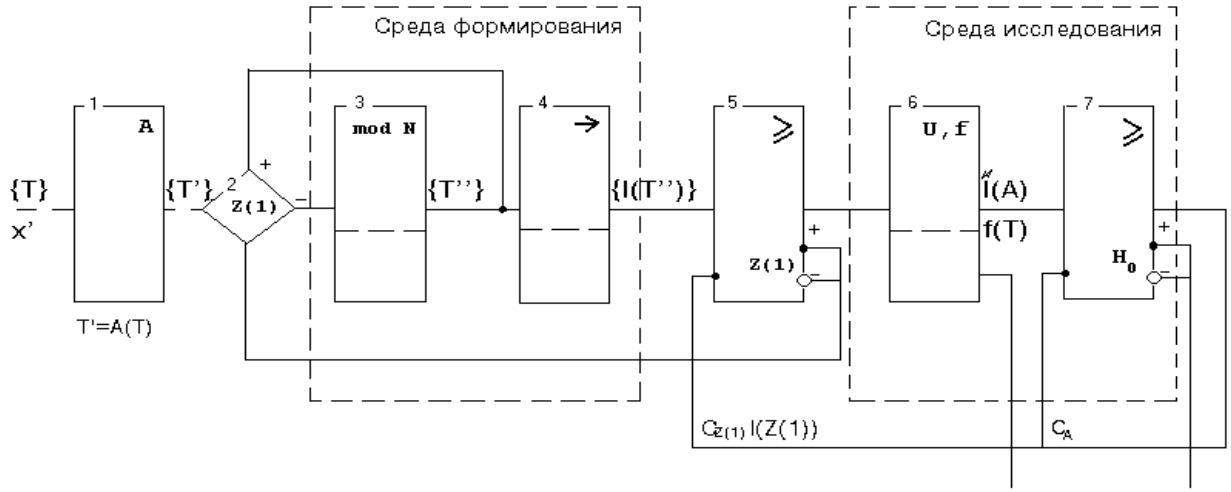


Рисунок 1 – Тракт прохождения информационного потока в контуре диагностической структуры

III Работа схемы в режиме предварительной обработки

Работа схемы в цикле № 1 (предварительная фаза обработки).

Исходное состояние задается гипотезой H_1 или $\overline{H_o}$ (поток алгоритма не идентифицирован), что влечет за собой гипотезу $\overline{Z(1)}$ - алфавит источника потока не детерминирован. Тракт прохождения всего потока будет выглядеть как последовательное движения от узла обработки № 1 к узлу обработки № 7.

Узел № 1 представляет собой процесс получения закрытого потока T` выбранного источника (1), иными словами – его языка. Открытость каналов связи предполагает поступление потока непосредственно на вход среды формирования.

Узел № 4 символизирует получение абстрактного образа *I(T)* – спектра потока, по признакам которого будет принято решение о его поведении. В основе его формирования положен метод частотного анализа символьной повторяемости с учетом взаимозависимости этих символов. С точки зрения математической статистики образ I(T) – это область упорядоченного векторного пространства множества гармоник (g):

$$I(T) = \{g_h\}, \quad h = [1,..,H], \quad H = 360. \tag{2}$$

Здесь $\{g_h\}$ – выборка, объем которой H диктует точность анализа.

Поскольку узел № 4 выполняет преобразования над ограниченным множеством символов алфавита, узел № 3 представляет собой устройство согласования исходного и требуемого множества алфавитов. Корректная работа данного устройства использует стандартный набор правил упаковки данных без потери данных с последующим переходом к используемому множеству обрабатываемого узлом № 4 алфавитом.

Узлы № 2, 5 на данном этапе транзитные.

Узел № 6 является ключевым на первом этапе, этапе предварительной обработки и именно им эксплуатируется раздел математической статистики в приложении к сформированному образу *I(T)*.

Процесс интерполяции подразумевает получение приемлемой функции в промежутках между узловыми точками i_h. Промежуток *[0, l]* включает H опорных точек: $[0 \le i_0 < i_1 < i_2 < ... < i_h \le l]$. Пусть, кроме того, заданы H действительных значений $|g_h|$ функции *I(T)* в узловых точках. Необходимо найти $I(T)=|g_h|$, где $0 \le j \le h$. Интерполяционная функция сопоставляет функции *I(T)* функцию известного класса $I`(T)= I(T; g_1(t),...)$, зависящую от параметров $g_i(t)$ выбранных так, чтобы значения *I(T)* совпадали со значениями *I`(T)*:

$$I`(t) = I(T). \tag{3}$$

При известном отклонении (ошибке) Δ:

$$\Delta = \int |I`(T) - I(T)|^2 \, \partial T,$$

ее выбирают таким образом, чтобы она была минимальной.

В применении к спектральному анализу, процесс аппроксимации заключается в том, что численно не выраженная функция $I(T)=\{g_h(t)\}$ заменяется вычислимой элементарной функцией *I`(T),* по возможности более точно, т.е. с наименьшим отклонением Δ от реальной кривой *I(T).* Рассматривая наличие *I`(T)*, элементами которой являются гармоники $g_h`(t)$, аппроксимацию можно выразить как:

$$I`(T`)=a_1\, g_1`(t)+a_2\, g_2`(t)+\ldots+a_h\, g_h`(t); \quad h=1,\ldots,H. \tag{4}$$

Сглаживание данных осуществляется процедурой, реализующей метод k-наименьших квадратов с адаптивным выбором k.

При исходном задании множества N потоков одного источника $\{T`\}, \{I(T`)\}$, формируется шаблон источника $\tilde{I}(A)$ в виде его аппроксимированной формулы (4) методом вычисления средней статистики потока:

$$\tilde{I}(A)=\sum_{N} I(T`)/N. \tag{5}$$

Кроме шаблона исследуемого источника потока вырабатывается порог принятия решения о принадлежности потока данному источнику, C_A. Естественно, что увеличение порога увеличивает вероятность правильного определения алгоритма, уменьшение – вероятности ложного срабатывания.

Данные результаты заполняют соответственные фильтры (т.н. процесс настройки спектральных фильтров), что устанавливает схему в рабочее состояние и переход в фазу 2.

IV Работа схемы в режиме идентификации

Работа схемы в цикле № 2 (основная обработка).

Алфавит источника детерминирован (условие $Z(1)$), шаблон источника (5) и порог принятия решения известны. Следовательно, процесс информационного прохождения неизвестного потока $x`$ в контуре диагностической структуры выглядит как 1, 2, 4-7.

Работа спектрального фильтра заключается в анализе близости образа потока к исследованному шаблону (выражается в %). Результат – гипотеза о принадлежности потока исследованному источнику (H_0).

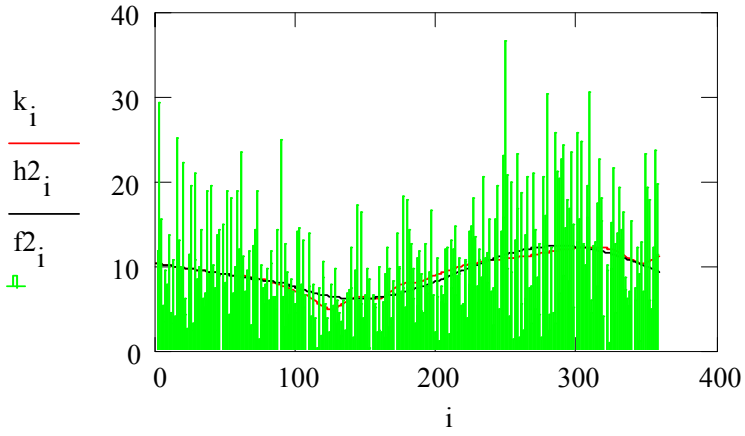
На рисунке 2 представлен пример процесса идентификации алгоритма преобразования ГОСТ 28147-89 с позиции спектрального фильтра.

$$l_i := \frac{2.5\cdot\pi\cdot v_i}{360} \qquad k_i := d2_i \qquad r := \text{mean}(d2) \qquad e_i := \frac{r}{3}\cdot\sin\left(\frac{\pi}{2}-l_i\right)+r$$

$$u := 0,1..89 \qquad l0_u := u \qquad l1_u := k_u \qquad s1 := \text{intercept}(l0,l1)$$

$$s2 := \text{slope}(l0,l1) \qquad h_i := s1 + s2\cdot i \qquad h2_i := \text{if}\left(i<90, h_i, e_i\right)$$

$$a1 := \sum_{i=0}^{359} |k_i| \qquad a2 := \sum_{i=0}^{359} |k_i - h2_i| \qquad B := 1 - \frac{a2}{a1} \qquad B = 0.955$$



$$Gost28147\_89 := \text{if}(B>0.85, \text{"Yes"}, \text{"No"})$$

$$Gost28147\_89 = \text{"Yes"}$$

Рисунок 2 – Идентификация алгоритма ГОСТ 28147-89

Здесь:
$f2_i$ – множество гармоник g_h,
k_i – огибающая спектра, полученная путем интерполяции и сглаживания $f2_i$;
$h2_i$ – огибающая спектра, полученная путем аппроксимации и сглаживания $f2_i$;
Mean (d2) – возвращает среднее значение *(r)* элементов вектора *d2;*
B – близость аппроксимированной кривой к кривой интерполированной.

В процессе аппроксимации результат идентификации потока достигается путем приведения n–множества гармоник к величине [0, 2.5*π]. Аппроксимированная кривая в данном случае имеет вид:

$$f(Gost) = \frac{r}{3} * \sin(\frac{\pi}{2} - g_h) + r.$$

Порог срабатывания – 85% близости к шаблону.

В таблице 1 приведены аппроксимирующие кривые некоторых распространенных алгоритмов [7].

Таблица 1

Формульные выражения огибающих распространенных алгоритмов

Алгоритм	Период	Число отрезков	Функция 1, 3 отрезков	Функция 2 отрезка	Интервал 2 отрезка
ГОСТ 28147-89	2.5π	2	a*x+b, a<0	$\frac{r}{3} * \sin(\frac{\pi}{2} - g_h) + r$	90…360
DES	2π	1	$\frac{r}{3.5} * (\sin(\frac{\pi}{1.9} - g_h) + r)$	–	–
TwoFish	2π	1	$\frac{r}{3} * (\sin(\frac{\pi}{5} - g_h) + r)$	–	–

V Выводы

Отметим некоторые области приложения поточных исследований:
- построение поисковых систем и информационных серверов;
- анализ инструментария, закрывающего поток;
- кластеризация и введение иерархической структуры информационного потока по выбранному целевому признаку.

Литература: 1. Дьяконов В.П. Справочник по MathCAD PLUS 7.0 PRO. – М.: СК Пресс, 1998. – 345 с. 2. Манухин А.В. Спектральный анализ сообщений. Критерий защищенности // Защита информации: Сборник научных трудов. – Киев, КМУГА, 1999. – с. 106–110. 3. Манухін О.В. Спектральний аналіз алгоритму гамування // Праці КВІУЗ. Випуск № 4. – К.: КВІУЗ, 1999. – с. 78–85. 4. Осинський Л.М., Манухін О.В. Спектральний аналіз алгоритму DES // Збірник наукових праць. Випуск № 1. – К.: КВІУЗ, 2000. – с. 103–110. 5. Четвериков И.А., Манухин А.В. Спектральный анализ текстовой информации и ее кластеризация // Защита информации: Сборник научных трудов. – Киев: КМУГА, 1998. – с. 224–229. 6. Четвериков И.А., Манухин А.В. Принципы криптоанализа с использованием портретно–образного метода // Материалы международной научно–технической конференции "Повышение эффективности систем защиты информации". "Защита–97". – Киев: КМУГА, 1997. – С. 66–68. 7. Data Encryption Standard. Federal Information Processing Standard (FIPS). Publication 46, National Bureau of Standards, U.S. Department of Commerce. – January 1977.