СТАТИСТИЧЕСКИЕ ОЦЕНКИ ЭНТРОПИЙНЫХ ФУНКЦИОНАЛОВ И ИХ ИСПОЛЬЗОВАНИЕ В ЗАДАЧАХ СТАТИСТИЧЕСКОГО ТЕСТИРОВАНИЯ КРИПТОГРАФИЧЕСКИХ ГЕНЕРАТОРОВ

В.Ю. ПАЛУХА, Ю.С. ХАРИН
ФПМИ БГУ, НИИ ППМИ БГУ

Введение

Важным структурным элементом средств криптографической защиты информации (криптосистем) являются генераторы случайных и псевдослучайных последовательностей [1]. Стойкость криптосистем зависит от того, насколько близка генерируемая последовательность по своим свойствам к равномерно распределённой случайной последовательности (РРСП). Одним из подходов к оценке качества генератора является статистическое оценивание энтропии и сравнение полученной оценки с ожидаемым значением для РРСП. Существуют различные функционалы энтропии, например, в [2] приводятся формулы 23 функционалов. Наиболее распространены функционалы энтропии Шеннона, Реньи и Тсаллиса, которые рассматриваются в данной статье.

Функционалы информационной энтропии

Пусть на вероятностном пространстве (Ω, F, P) с множеством состояний $\Omega = \{\omega_1, \ldots, \omega_N\}$ и вероятностями $p_i = P\{x = \omega_i\}$, $p_i \geq 0$, $\sum_{i=1}^{N} p_i = 1$ наблюдается последовательность $\{x_t\}$.

Определим функционал энтропии согласно [1]:

$$H_{h,w}^{\varphi_1,\varphi_2}(P) = h\left(\frac{\sum_{i=1}^{N} w_i \varphi_1(p_i)}{\sum_{i=1}^{N} w_i \varphi_2(p_i)}\right), \qquad (1)$$

где $w_i > 0$, $i = 1, \ldots, N$ — веса состояний ω_i, и заданы функции $h: \mathbb{R} \to \mathbb{R}$, $\varphi_1: [0,1) \to \mathbb{R}$, $\varphi_2: [0,1) \to \mathbb{R}$.

В таблице 1 приведены наиболее часто используемые в теории информации функционалы энтропии в соответствии с общим видом (1).

Таблица 1

Основные функционалы энтропии

Тип	Формула	h(x)	$\varphi_1(x)$	$\varphi_2(x)$	w_i
Шеннон	$H(P) = -\sum_{i=1}^{N} p_i \ln p_i$	x	$-x\log x$	x	w
Реньи	$H_r(P) = \frac{1}{1-r} \ln\left(\sum_{i=1}^{N} p_i^r\right),\ r \in \mathbb{N}, r > 1.$	$(1-r)^{-1}\log x$	x^r	x	w
Тсаллис	$S_r(P) = \frac{1}{r-1}\left(1 - \sum_{i=1}^{N} p_i^r\right),\ r \in \mathbb{N}, r > 1.$	$(1-r)^{-1}(x-1)$	x^r	x	w

Статистические оценки энтропии

Будем рассматривать стационарные в узком смысле двоичные последовательности $\{x_t\} \in V = \{0,1\}$ на некотором вероятностном пространстве (Ω, F, P). Пусть $p_{i_1,\dots,i_s} = P\{x_{t+1} = i_1, \dots, x_{t+s} = i_s\}$ — распределение вероятностей s-граммы $(x_{t+1},\dots,x_{t+s}) \in V_s$, которое не зависит от $t \in \mathrm{N}_0 = \mathrm{N} \cup \{0\}$. В предыдущих обозначениях имеем $N = 2^s$.

Обозначим: $x^{\underline{r}} = x(x-1)\dots(x-r+1)$ — нисходящая факториальная степень (при $x < r$ полагают $x^{\underline{r}} ::= 0$) [3]; $H_* = \{\{x_t\}$ есть РРСП$\}$ — гипотеза о том, что наблюдаемая последовательность является «чисто случайной», т.е. РРСП.

Пусть наблюдается n фрагментов длины s: $X^{(k)} = (x_1^{(k)}, \dots, x_s^{(k)}) \in V_s, \quad k = 1, \dots, n$, сформированных из элементов последовательности $\{x_t\}$. Для построения статистических оценок функционалов энтропии необходимо построить оценки распределения вероятностей $p_J(s) = P\{X^{(k)} = J_1^s\}$, где $J_1^s = (j_1, \dots, j_s) \in V_s$ — мультииндекс, по n фрагментам длины $s \geq 1$. Построим оценки вероятностей $\{\hat{p}_J(s)\}$ и r-ых степеней вероятностей $\{\widehat{p_J^r}(s)\}$, используя частотные статистики:

$$\hat{p}_J(s) = \frac{\nu_J}{n}, \quad \widehat{p_J^r}(s) = \frac{\nu_J^{\underline{r}}}{n^r}, \quad \nu_J = \sum_{k=1}^{n} \delta_{X^{(k)},J}, \quad \delta_{X^{(k)},J} = \begin{cases} 1, X^{(k)} = J; \\ 0, X^{(k)} \neq J. \end{cases} \tag{2}$$

На основе частотных статистик (2) построим оценку s-мерной энтропии Шеннона:

$$\hat{H}(n,s) = -\sum_{J \in V_s} \hat{p}_J(s) \ln \hat{p}_J(s), \tag{3}$$

а также оценки s-мерной энтропии Реньи и Тсаллиса

$$\widehat{H_r}(n,s) = \frac{1}{1-r} \ln\left(\sum_{J \in V_s} \widehat{p_J^r}(s) \right), \tag{4}$$

$$\widehat{S_r}(n,s) = \frac{1}{r-1}\left(1 - \sum_{J \in V_s} \widehat{p_J^r}(s) \right). \tag{5}$$

Рассмотрим асимптотику, означающую, что с увеличением количества наблюдений n увеличивается число N возможных исходов полиномиальной схемы:

$$n, N = 2^s \to \infty, n/N \to \lambda, 0 < \lambda < \infty. \tag{6}$$

Распределение вероятностей статистических оценок функционалов энтропии описывается следующей теоремой, доказательство которой базируется на результатах из [4].

Теорема. При истинной гипотезе H_* в асимптотике (6) статистические оценки функционалов энтропии (3)–(5) имеют асимптотически нормальное распределение, причём для асимптотических математических ожиданий и дисперсий справедливы равенства

$$E\{\hat{H}(n,s)\} = \ln n - e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\ln(k+1)\lambda^k}{k!},$$

$$D\{\hat{H}(n,s)\} = \frac{e^{-\lambda}}{n} \sum_{k=1}^{+\infty} \frac{(k+1)\lambda^k}{k!} \ln^2(k+1) - \frac{e^{-2\lambda}}{2^s}\left(\sum_{k=1}^{+\infty} \frac{\ln(k+1)\lambda^k}{k!} \right)^2 - \frac{e^{-2\lambda}}{n}\left(\sum_{k=1}^{+\infty} \ln(k+1) \frac{\lambda^k}{k!}(k+1-\lambda) \right)^2,$$

$$E\{\widehat{H_r}(n,s)\} = s\ln 2, \tag{7}$$

$$D\{\widehat{H_r}(n,s)\} = \frac{\sum_{i=1}^{r} s(r,i)\sum_{j=0}^{i-1} C_i^j r^{i-j}\sum_{k=1}^{j} S(j,k)\lambda^k - r^2\lambda^{r-1} + r!}{(r-1)^2 n\lambda^{r-1}},$$

$$E\{\widehat{S_r}(n,s)\} = \frac{1}{r-1}\left(1 - \frac{1}{2^{s(r-1)}}\right),$$

$$D\{\widehat{S_r}(n,s)\} = \frac{\lambda^{r-1}}{(r-1)^2 n^{2r-1}}\left(\sum_{i=1}^{r} s(r,i)\sum_{j=0}^{i-1} C_i^j r^{i-j}\sum_{k=1}^{j} S(j,k)\lambda^k - r^2\lambda^{r-1} + r!\right),$$

где $s(r, i)$ и $S(r, i)$ — числа Стирлинга первого и второго рода соответственно [3].

Интервальные оценки энтропии и проверка гипотез

Поскольку нам известно асимптотическое распределение точечных оценок энтропии, мы можем построить интервальные оценки. Зададим уровень значимости $\varepsilon \in (0,1)$. Введём обозначения: $\hat{h}(n,s)$ — статистическая оценка энтропии Шеннона, Реньи или Тсаллиса, построенная по n фрагментам длины s; μ и σ^2 — асимптотические математическое ожидание и дисперсия статистической оценки используемого функционала энтропии при истинной гипотезе H_*, вычисленные по формулам (7) из теоремы 1. Тогда в асимптотике (6) с вероятностью $1 - \varepsilon$ $\hat{h}(n,s) \in (t_-, t_+)$, $t_\pm = \mu \pm \sigma\Phi^{-1}(1 - \frac{\varepsilon}{2})$, где $\Phi^{-1}(p)$ — квантиль уровня p стандартного нормального закона [5].

На основе интервальной оценки построим решающее правило для статистического тестирования генераторов, т.е. проверки гипотез о том, является ли наблюдаемая последовательность генератора «чисто случайной»: H_* и $\overline{H_*}$. Вычислим для наблюдаемой последовательности статистику $\hat{h}(n,s)$ и доверительный интервал для заданного уровня значимости. Решающее правило имеет вид:

$$\begin{cases} H_*, & \text{если } t_- < \widehat{h_r}(n,N) < t_+; \\ \overline{H_*}, & \text{в противном случае,} \end{cases} \qquad t_\pm = \mu \pm \sigma\Phi^{-1}\left(1 - \frac{\varepsilon}{2}\right),$$

В случае принятия решения о справедливости гипотезы H_* можно сделать вывод о том, что генератор пригоден для использования в криптосистемах, поскольку по своим свойствам он неотличим от равномерно распределённой случайной последовательности на основе выборки объёма n.

Заключение

Для проверки пригодности генераторов случайных и псевдослучайных последовательностей в криптосистемах предложено использовать статистические оценки функционалов энтропии Шеннона, Реньи и Тсаллиса, построенные по выходным последовательностям генераторов. Найдены асимптотические распределения вероятностей точечных стати-

стических оценок, на основе которых построены интервальные оценки и разработано решающее правило для статистического тестирования генераторов.

СПИСОК ЛИТЕРАТУРЫ

1. Криптология / Ю. С. Харин [и др.]. — Минск: БГУ, 2013. — 512 с.
2. Esteban, M.D. A summary on entropy statistics / M. D. Esteban, D. Morales // Kybernetika. — 1995. — Vol. 31, № 4. — P. 337–346.
3. Энвин, А. Ю. Дискретная математика: Конспект лекций / А. Ю. Энвин. — Челябинск: Издательство ЮУрГУ, 1998. — 176 с.
4. Holst, L. Asymptotic normality and efficiency for certain goodness-of-fit tests / L. Holst // Biometrika. — 1972. — № 59. — P. 137–145.
5. Харин, Ю. С. Теория вероятностей, математическая и прикладная статистика / Ю. С. Харин, Н. М. Зуев, Е. Е. Жук. — Минск: БГУ, 2011. — 463 с.

ОСОБЕННОСТИ ТЕСТИРОВАНИЯ СРЕДСТВ ЗАЩИТЫ ИНФОРМАЦИИ ОТ НЕСАНКЦИОНИРОВАННОГО ДОСТУПА В ВИРТУАЛЬНОЙ ИНФРАСТРУКТУРЕ

Н. В. МОЗОЛИНА, К. А. ЛУГОВЦОВА
МФТИ, Москва, Россия

В статье рассматриваются особенности и проблемы функционального тестирования средств защиты информации от несанкционированного доступа в виртуальной инфраструктуре, а также предлагается способ их решения при помощи средств автоматизации тестирования и разделения процесса тестирования на два независимых этапа.

Ключевые слова: функциональное тестирование, автоматизация процессов тестирования, виртуальная инфраструктура.

Неотъемлемым этапом разработки программного обеспечения, в том числе и средств защиты информации, является тестирование всех его компонентов на соответствие заданным требованиям. Существует множество видов тестирования, в данной статье ограничимся рассмотрением функционального тестирования, направленного на проверку того, какие функции программного обеспечения реализованы, и что они работают верным образом [1].

Процесс функционального тестирования состоит из четырёх стадий: выбор действия, определение ожидаемого результата этого действия, определение фактического результата и сравнение результатов.

Одним из самых простых способов тестирования является выполнение всех этих операций человеком вручную. Такой подход хоть и не требует большого объема подготовительных действий от тестировщика, но вместе с тем имеет ряд существенных минусов: человек может случайно или осознанно пропустить часть тестов, неверно интерпретировать результаты испытания, сделать ложные выводы, а также такое тестирование может растянуться на недели и даже месяцы. Для решения таких проблем используют автоматизацию: IBM Rational Functional Tester (RFT), TestComplete и другие продукты позволяют воспроизводить действия пользователя программного обеспечения, обраба-