РАЗРАБОТКА ПРОГРАММНОГО КОМПЛЕКСА СТАТИСТИЧЕСКОГО ТЕСТИРОВАНИЯ КРИПТОГРАФИЧЕСКИХ ГЕНЕРАТОРОВ НА ОСНОВЕ МАРКОВСКИХ МОДЕЛЕЙ

студента 4 курса
Деркача Максима Юрьевича,
специальности
"Компьютерная безопасность"

научный руководитель:
чл.-корр. НАН Беларуси,
доктор физ.-мат. наук,
профессор
Харин Юрий Семенович

Введение

Основной способ оценки стойкости криптографических генераторов основан на статистическом анализе выходных последовательностей криптографических генераторов. Существующие в настоящее время программные комплексы (так называемые “батареи тестов”) используют простейщие вероятностные модели выходных последовательностей.
Для современных криптографических генераторов этих простейших моделей недостаточно.
Для этих целей был разработан программный комплекс, использующий следующие модели Маркова: порядка $s \geqslant 1$, скрытые цепи Маркова, двойные цепи Маркова.

Математическое описание марковских моделей

Использовались следующие математические модели выходных последовательностей:

M_1: Однородная цепь Маркова.

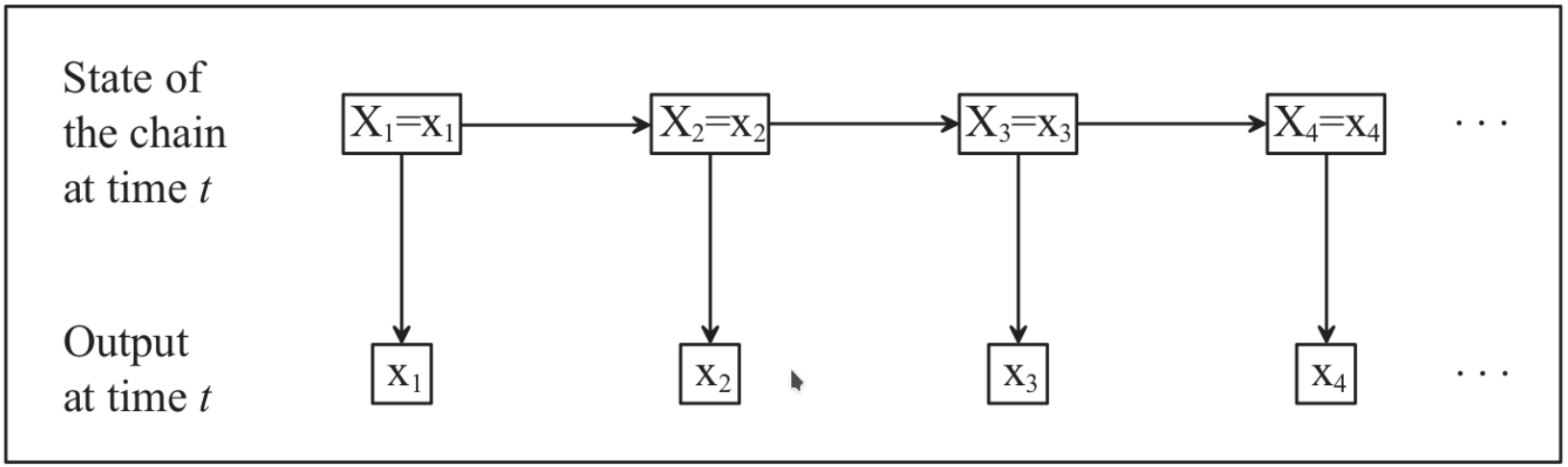
M_2: Однородная цепь Маркова s-го порядка.

M_3: Скрытая марковская модель.

M_4: Двойная марковская модель.

Математическое описание марковских моделей

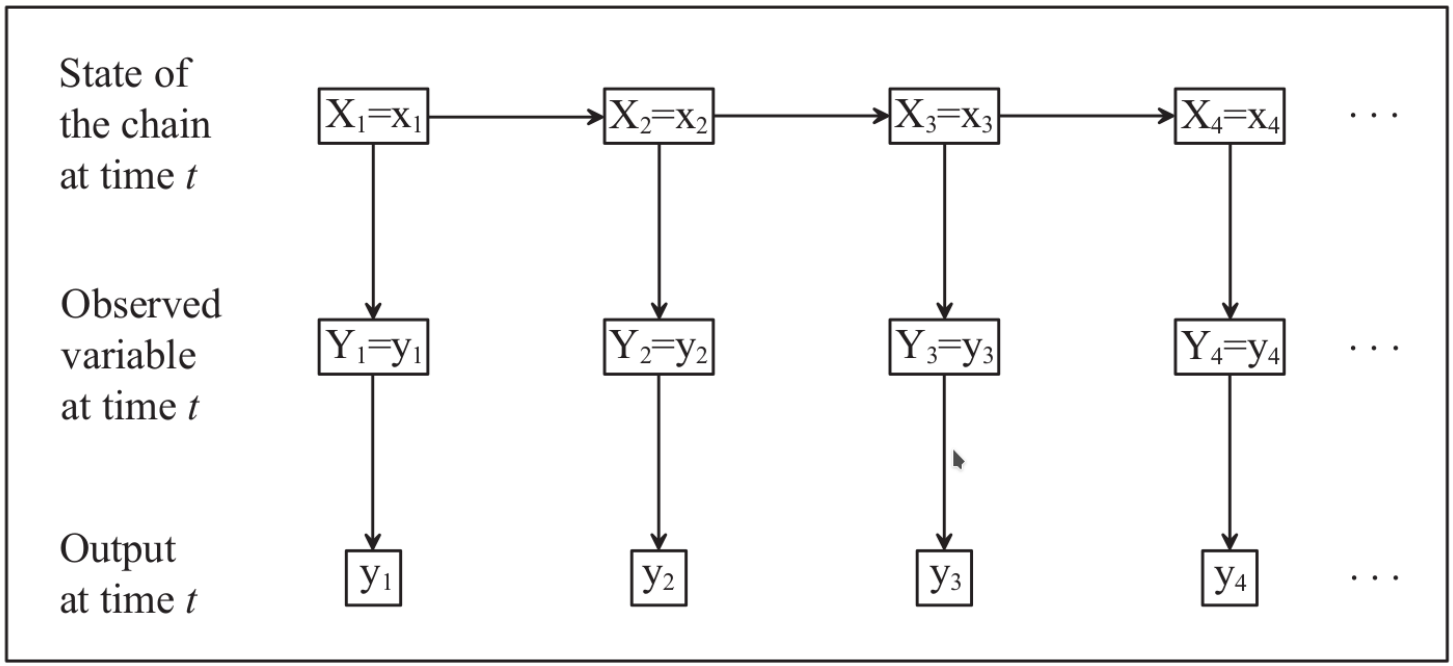
Однородная цепь Маркова:



Однородная цепь Маркова определяется начальным распределением вероятностей π и матрицей вероятностей одношаговых переходов P.

Математическое описание марковских моделей

Скрытая марковская модель:

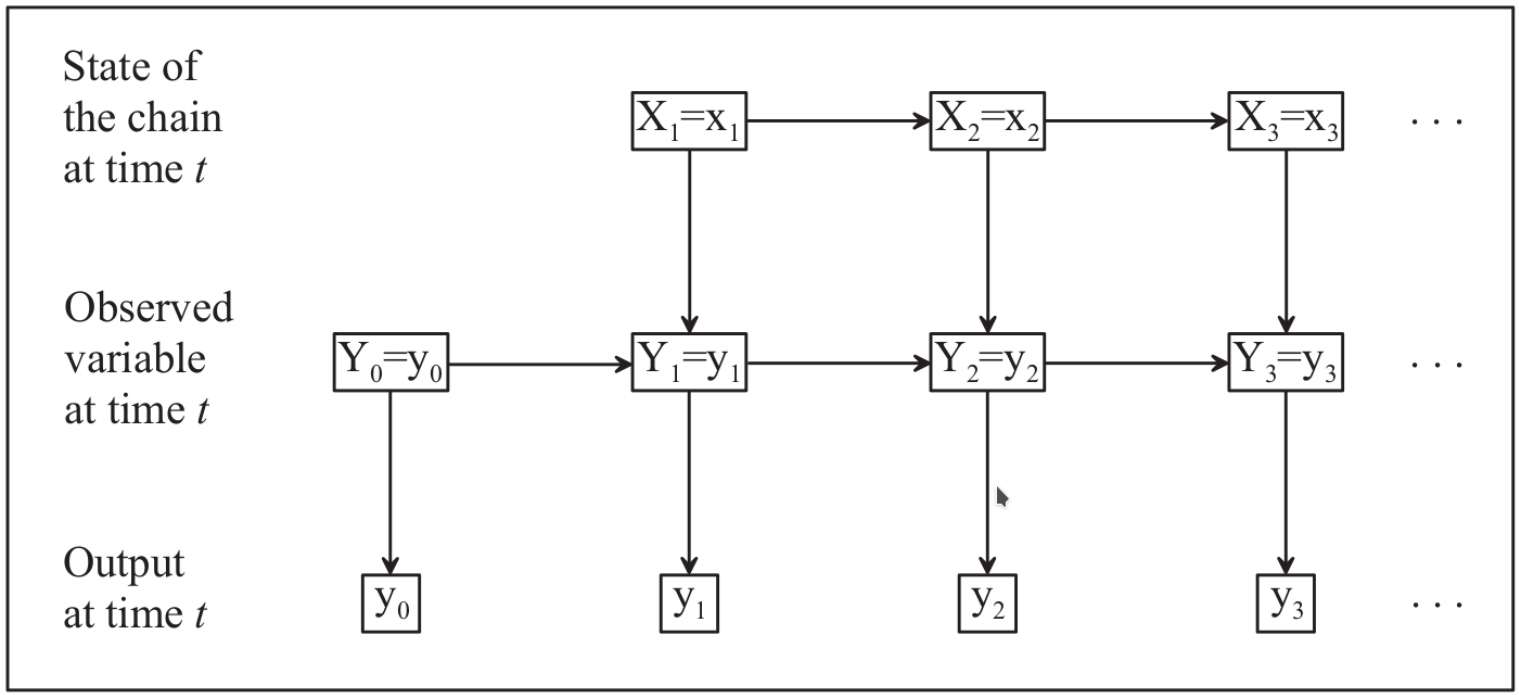


Скрытая марковская модель полностью определяется следующими параметрами:

- начальное распределение вероятностей скрытых состояний π;
- матрица вероятностей одношаговых переходов скрытых состояний P;
- матрица вероятностей переходов из скрытого состояния в обозреваемое C.

Математическое описание марковских моделей

Двойная марковская модель:



Двойная марковская модель полностью определяется следующими параметрами:

- начальное распределение вероятностей скрытых состояний π;
- матрица вероятностей одношаговых переходов скрытых состояний P;
- множество матрица C.

Алгоритмы оценивания параметров ОЦМ и ОЦМ-s

Алгоритмы статистического оценивания параметров однородной цепи Маркова и однородной цепи Маркова s-го порядка:

- Алгоритм построения оценки максимального правдоподобия (Maximum Likehood Estimation);
- Алгоритм статистического бутстрэпа (Bootstrap);
- Алгоритм сглаживания оценки максимального правдоподобия.

Алгоритмы оценивания параметров DCMM и HMM

Алгоритмы статистического оценивания параметров двойной и скрытой марковских моделей:

- Алгоритма прямого-обратного хода (Forward–backward algorithm);
- Алгоритм Баума-Велша (прямого-обратного хода) оценивания параметров модели;
- Алгоритмом Витерби оценивания оптимальной последовательности состояний.

Алгоритмы статистического тестирования криптогенераторов

Введем следующие гипотезы:

$$\begin{aligned} H_0 &: \theta = \theta^0, \\ H_1 = \bar{H}_0 &: \theta \neq \theta^0, \end{aligned}$$

Для M_1, M_2**:**
$\theta = P, \quad \theta^0 = P^0$

Для M_3, M_4**:**
$\theta = (P, C), \quad \theta^0 = (P^0, C^0),$

где P — неизвестная истинная матрица вероятностей однашаговых переходов скрытых состояний,
P^0 — гипотетическая (предпологаемая) матрица вероятностей однашаговых переходов скрытых состояний,
C — неизвестная истинная (множество)матрицы вероятностей переходов из скрытого состояния в обозреваемое,
C^0 — гипотетическая (предпологаемая) (множество) матрица вероятностей переходов переходов из скрытого состояния в обозреваемое.

Алгоритмы статистического тестирования криптогенераторов

В дальнейшем будем рассматривать случай, когда

$H_0 = \{$ гипотеза о том, что наблюдаемая последовательность X - РРСП $\}$.

В параметрическом виде эта гипотеза представима следующим образом:

$$H_0 \quad : \quad P^0 = (p_{ij}^0), p_{ij}^0 \equiv \frac{1}{N}, i, j \in A,$$
$$C^0 = (c_{ij}^0), c_{ij}^0 \equiv \frac{1}{M}, i \in A, j \in B.$$

Алгоритмы статистического тестирования криптогенераторов

χ^2 - Критерий согласия Пирсона для моделей M_1, M_2

Этот критерий имеет следующий вид:

$$\text{принимается} \quad \begin{cases} H_0, & \text{если} \quad \gamma_T \leqslant \Delta; \\ H_1, & \text{если} \quad \gamma_T > \Delta, \end{cases}$$

$$\gamma_T = \sum_{i,j \in A, p_{ij}^0 \neq 0} \frac{(n_{ij} - n_i p_{ij}^0)^2}{n_i p_{ij}^0} = \sum_{i,j \in A, p_{ij}^0 \neq 0} \frac{(p_{ij} - p_{ij}^0)^2}{p_{ij}^0} n_i,$$

где n_i - количество раз, когда состояние $i \in A$ обозревается в последовательности $x_1, x_2, ..., x_{n-1}$, $\Delta = F^{-1}_{\chi^2_{m-N}}(1-\epsilon)$ - порог критерия, определяемый по уровню значимости $\epsilon = P\{H_1|H_0\} \in (0,1)$ через квантиль χ^2 - распределения с $m - N$ степенями свободы, где $m = N^2 - \sum_{i,j \in A} \delta_{p_{ij}^0}$ - число ненулевых вероятностей одношаговых переходов в гипотетической $(N \times N)$ - матрице $P^0 = (p_{ij}^0)_{i,j \in A}$.

Алгоритмы статистического тестирования криптогенераторов

Критерий отношения правдоподобия для моделей M_3, M_4

Проверка гипотез H_0, H_1 основаная на критерии отношения правдоподобия, имеет следующий вид:

$$\text{принимается} \quad \begin{cases} H_0, & \text{если} \quad \lambda_T \leqslant \Delta; \\ H_1, & \text{если} \quad \lambda_T > \Delta, \end{cases}$$

$$\lambda_T = -2(logL(P^0, C^0) - logL(P, C)).$$

где $\Delta = F^{-1}_{\chi^2_m}(1-\epsilon)$ - порог критерия, определяемый по уровню значимости $\epsilon = P\{H_1|H_0\} \in (0,1)$ с m степенями свободы, где $m = N(N-1) + N(M-1)$ - число независимых параметров, задающих $\Theta = P, C$.

Структура программного комплекса

Результаты

Результаты на модельных и реальных данных

Модельные данные:

- $S_1 = \{\pi_1, P_1\} : \pi_1 = (0.5, 0.5)^T, P_1 = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}, T_1 = 20.$
- $S_2 = \{\pi_2, P_2\} : \pi_2 = (0.5, 0.5)^T, P_2 = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}, T_2 = 100.$
- $S_3 = \{\pi_3, P_3\} : \pi_3 = (0.5, 0.5)^T, P_3 = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}, T_3 = 1000.$
- $S_4 = \{\pi_4, P_4\} : \pi_4 = (0.45, 0.55)^T, P_4 = \begin{pmatrix} 0.43 & 0.57 \\ 0.58 & 0.42 \end{pmatrix}, T_4 = 50.$
- $S_5 = \{\pi_5, P_5\} : \pi_4 = (0.45, 0.55)^T, P_5 = \begin{pmatrix} 0.43 & 0.57 \\ 0.58 & 0.42 \end{pmatrix}, T_5 = 1000.$

Результаты

Результаты на модельных данных

№	Данные	Длина последовательности	M1	M2	M3	M4
1	S_1	20	H_0	H_1	H_1	H_1
2	S_2	100	H_0	H_0	H_0	H_0
3	S_3	1000	H_0	H_0	H_0	H_0
4	S_4	50	H_0	H_0	H_0	H_0
5	S_5	1000	H_1	H_1	H_1	H_1

Результаты

Результаты на реальных данных:
Для экспериментов использовались следующие выходные последовательности физического генератора, полученные с сайта Humboldt University of Berlin (http://qrng.physik.hu-berlin.de/download):

- $S_1, N = 2, T_1 = 8388608(1MB)$.
- $S_2, N = 2, T_2 = 125829120(15MB)$.
- $S_3, N = 2, T_3 = 838860800(100MB)$.

Результаты

Результаты на реальных данных

№	Данные	Длина последовательности	M1	M2	M3	M4
1	S_1	8388608	H_0	H_0	H_0	H_0
2	S_2	125829120	H_0	H_0	H_0	H_0
3	S_3	838860800	H_0	H_0	H_0	H_0

Спасибо за внимание