

**СВЯЗЬ “ФУНКЦИИ НЕНАДЕЖНОСТИ” И “РАССТОЯНИЯ
ЕДИНСТВЕННОСТИ” КРИПТОСИСТЕМ С БАЗОВЫМИ ПАРАМЕТРАМИ
ШИФРАТОРА В ФОРМЕ МАТЕМАТИЧЕСКОЙ МОДЕЛИ
СИНХРОННОГО АВТОМАТА ХАФФМАНА**

А.А. Горбунов, К.Г. Кирьянов

Найдены связи “функции ненадежности” и “расстояния единственности” К. Шеннона [1] с базовыми параметрами удобных для измерения входных, выходных и взаимных векторных данных [2] шифраторов криптосистем в форме математической модели синхронных автоматов Хаффмана, обобщающие формулы, полученные в работе [3].

1. ВВЕДЕНИЕ

Одними из важнейших теоретических характеристик, определяющих криптостойкость систем шифрования (криптосистем – КС) с секретным ключом, являются функция ненадежности шифра и расстояние единственности. Данные характеристики были предложены Шенноном [1] и определены через вероятностные характеристики открытого и зашифрованного текстов. В [1] (рис.1) предполагается, что входной сигнал $u \equiv u(t) = u(v \cdot \Delta t) = \{u_0, u_1, \dots, u_{M-1}\}$ (открытое сообщение) и выходной сигнал $y \equiv y(t) = y(v \cdot \Delta t) = \{y_0, y_1, \dots, y_{M-1}\}$ (синхронная с открытым сообщением шифрограмма) ($v=0, 1, \dots, M-1$) шифратора

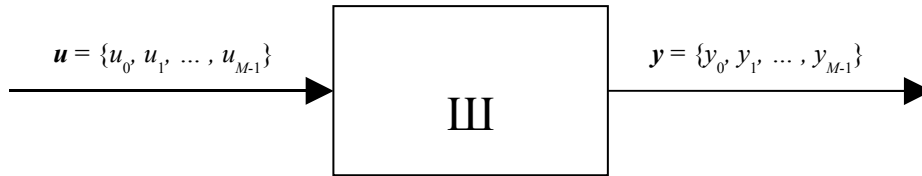


Рис. 1. Схема шифратора (Ш)

состоят из последовательности M дискретных по времени (интервалы времени Δt между итерациями автомата – *такты*, величина которых принимается обычно за единицу измерения времени) и значению отсчетов, каждый из которых выбирается из конечных, соответственно, Q_u - и Q_y - элементных множеств символов (алфавитов). Данные сигналы \mathbf{u} и \mathbf{y} , интерпретируются как тексты с набором символов из алфавитов $[0, Q_u-1]$ и $[0, Q_y-1]$ соответственно.

Энтропия (мера неопределенности) входного сообщения $H(\mathbf{u})$ введена Шенноном через вероятности $p(\mathbf{u}^{(i)})$ появления различных вариантов данного текста $\mathbf{u}^{(i)} = \{u^{(i)}_0, u^{(i)}_1, \dots, u^{(i)}_{M-1}\}$ [1]:

$$H(\mathbf{u}) = - \sum_i p(\mathbf{u}^{(i)}) \cdot \log p(\mathbf{u}^{(i)}), \quad (1)$$

где суммирование осуществляется по всем возможным $(Q_u)^M$ вариантам $\mathbf{u}^{(i)}$ сообщения \mathbf{u} . Аналогичную меру неопределенности имеет и выходное сообщение \mathbf{y} :

$$H(\mathbf{y}) = - \sum_i p(\mathbf{y}^{(i)}) \cdot \log p(\mathbf{y}^{(i)}). \quad (2)$$

Совместная энтропия сигналов $H(\mathbf{u}, \mathbf{y})$ вводится через совместную вероятность появления определенного варианта сообщения \mathbf{u} на входе и определенного варианта сообщения \mathbf{y} на выходе:

$$H(\mathbf{u}, \mathbf{y}) = - \sum_{i,j} p(\mathbf{u}^{(i)}, \mathbf{y}^{(j)}) \cdot \log p(\mathbf{u}^{(i)}, \mathbf{y}^{(j)}). \quad (3)$$

В качестве теоретической меры секретности используется *средняя* условная энтропия исходного сообщения $H(\mathbf{u}/\mathbf{y})$ при условии известной криптограммы \mathbf{y} , названная Шенноном *функцией ненадежности текста сообщения* ([1, стр. 368]):

$$H(\mathbf{u}/\mathbf{y}) = H(\mathbf{u}, \mathbf{y}) - H(\mathbf{y}). \quad (4)$$

Значение длины криптограммы, при которой только у одного из вариантов исходного сообщения вероятность остается близкой к единице, а вероятности всех остальных вариантов сообщений стремятся к нулю (т.е. и значение функции ненадежности $H(\mathbf{u}/\mathbf{y}) \rightarrow 0$), называется *расстоянием единственности*. В работе Шеннона [1] оценка данной величины дается выражением:

$$n_0 = \frac{H(K)}{D} \quad (5)$$

где $D = D_u = \log Q_u - H(\mathbf{u})/M$ – избыточность открытого сообщения, отнесенная к одному символу, $H(K)$ – энтропия ключа шифрования.

В то же время открытый текст и “шифротекст” можно рассматривать как сигналы, порождаемые соответствующими Источниками экспериментальных данных. В [2] введена математическая модель (ММ) нестационарных источников экспериментальных данных – генетическая карта данных (ГК). К *оптимальным базовым параметрам* (БП) ММ источника каждого i -го участка стационарности длиной $m^{(i)}$ изначально продискретизированного по времени с интервалом Δt текста относятся:

$$\text{БП} = \{q, n\}. \quad (6)$$

Здесь $q = q_{opt}$ – число уровней квантования, $n = n_{opt}$ – т.н. “сложность” порождающего источника, т.к. предполагается, что для каждого i -го участка стационарности длиной $m^{(i)}$ существует абстрактный стационарный детерминированный дискретный автомат n -го порядка, являющийся источником

участка текста, который может безошибочно породить по его n начальным символам все оставшиеся $m^{(i)}-n$ символов i -го участка текста. Значения оптимальных по минимуму энтропии БП источника сигнала определяются экспериментально по его реализации [2].

2. ПОНЯТИЕ ЭНТРОПИИ, ВВОДИМОЕ ЧЕРЕЗ БАЗОВЫЕ ПАРАМЕТРЫ ТЕКСТА

Как следует из определения шифратора криптосистемы (“секретной системы”), данного Шенноном, “секретная система” есть семейство *однозначно обратимых* отображений множества возможных сообщений во множество криптограмм” [1, стр. 340]), преобразование, осуществляемое шифратором КС, является *невыврожденным*. Отсюда следует, что шенноновские энтропии входного сообщения и его шифротекста равны [1, стр. 268]. С учетом данного обстоятельства в рамках подхода, основанного на понятии ГК и БП, в работе [3] получена формула для расстояния единственности в терминах базовых параметров шифротекста

$$n_0 = \frac{H(K)}{D}, \quad (7)$$

где за оценку избыточности текста взята величина

$$D = \frac{1}{2} \cdot \log q \quad (8)$$

(см., напр., [2]), в качестве энтропии ключа шифрования

$$H(K) = \log(\bar{m}) \quad (9)$$

взята величина логарифма средней длины участка стационарности текста (средней длины “гена” шифротекста), составляющая

$$\bar{m} = (q^n)^{1/2}, \quad (10)$$

а так же, как и в работе Шеннона [1], предполагалась скалярность входного и выходного сигналов шифратора.

Одним из проявлений универсальности подхода, основанного на ММ ГК данных, в отличие от описанного в [1], является возможность получения оптимальных оценок БП не только скалярных, но также и векторных сигналов (см., например, [2]). Знание оценок БП векторных сигналов даёт, как следует из дальнейшего изложения, возможность установить связь функции ненадежности и расстояния единственности криптосистем с базовыми параметрами шифратора КС для самой общей формы ММ дискретного синхронного автомата (автомата Хаффмана) с наблюдаемыми (доступными для измерения) векторными входами, выходами и ненаблюдаемыми (недоступными для измерения) внутренними состояниями (рис. 2).

Для нахождения функции ненадежности КС через базовые параметры открытого текста, шифротекста, а также векторного текста, полученного объединением их компонент, будем рассматривать шифратор как неавтономный генератор текстового сигнала $y(t)$ (рис. 2). На входе он испытывает воздействие сигнала открытого текста (и, возможно, ключа – как компоненты $u(t)$), а на выходе выдает сигнал шифротекста. Без ограничения общности при нашем желании можно считать, что символы ключа порождаются и воздействуют на передаваемое сообщение *подсистемой*, встроенной в саму систему шифратора и включены в математическую модель (ММ) данного дискретного автомата-шифратора. Для векторных сигналов данные выражения представляются следующим образом:

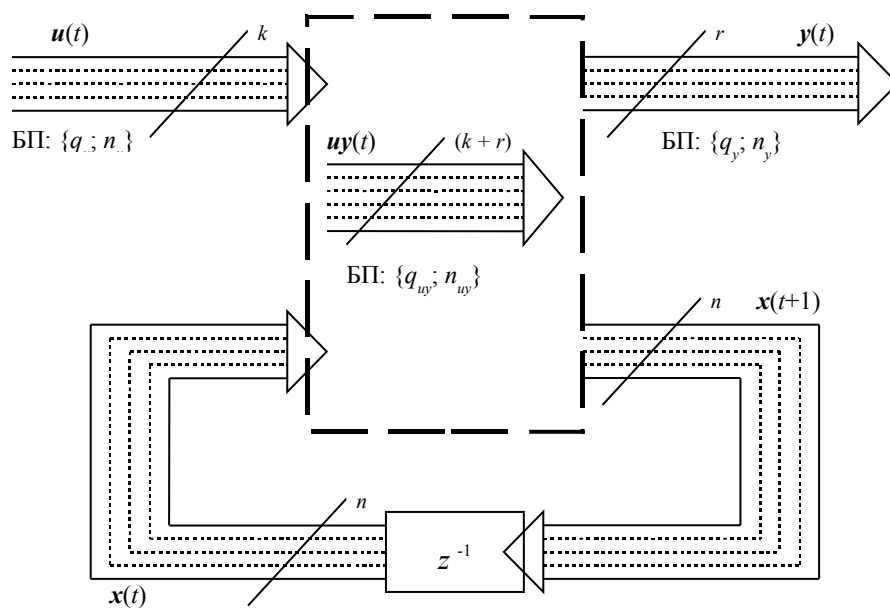


Рис. 2. Шифратор КС в форме ММ синхронного дискретного автомата Хаффмана с указанием доступных для измерения БП его входного ($u(t)$), выходного ($y(t)$) и взаимного ($uy(t)$) (вместо ненаблюдаемого внутреннего $x(t)$) векторных текстов

$$u(t) = \begin{bmatrix} u^{(1)}(t) \\ \vdots \\ u^{(k)}(t) \end{bmatrix} \text{ – векторный входной сигнал,}$$

$$y(t) = \begin{bmatrix} y^{(1)}(t) \\ \vdots \\ y^{(r)}(t) \end{bmatrix} \text{ – векторный выходной сигнал,} \quad (11)$$

$$uy(t) = \begin{bmatrix} u^{(1)}(t) \\ \vdots \\ u^{(k)}(t) \\ y^{(1)}(t) \\ \vdots \\ y^{(r)}(t) \end{bmatrix} \text{ – векторный взаимный сигнал.}$$

Для одной и той же длины M всех компонент векторов $u(t)$, $y(t)$ и $uy(t)$ соответствующие объемы N_u , N_y и N_{uy} фазовых пространств (или числа точек / состояний) дискретных динамических систем (ДДС), порождающих

соответствующие три “гена” (“участка стационарности”) векторных текстовых (т.е. дискретизированных по времени и по уровню) сигналов, равняются:

$$N_u = \left(q_u^k\right)^{n_u} = \left(q_u\right)^{n_u \cdot k}, \quad N_y = \left(q_y^r\right)^{n_y} = \left(q_y\right)^{n_y \cdot r},$$

$$N_{uy} = \left(q_{uy}^{(k+r)}\right)^{n_{uy}} = \left(q_{uy}\right)^{n_{uy} \cdot (k+r)}, \quad (12)$$

где k – число компонент во входном сигнале $u(t)$, r – число компонент в выходном сигнале $y(t)$, $(k+r)$ – число компонент во взаимном сигнале (все компоненты сигнала $u(t)$ совместно со всеми компонентами сигнала $y(t)$).

Выражения для энтропии в терминах базовых параметров записываются в виде [2]:

$$E(\mathbf{u}) = \log(N_u) = n_u \cdot k \cdot \log(q_u), \quad E(\mathbf{y}) = \log(N_y) = n_y \cdot r \cdot \log(q_y), \quad (13)$$

$$E(\mathbf{uy}) = \log(N_{uy}) = n_{uy} \cdot (k+r) \cdot \log(q_{uy}).$$

По аналогии с условной энтропией в терминах Шеннона записываем в терминах базовых параметров следующее выражение для условной энтропии “с выхода на вход \leftarrow ” (функции ненадежности):

$$E(\mathbf{u}/\mathbf{y}) = E(\mathbf{uy}) - E(\mathbf{y}) = n_{uy} \cdot (k+r) \cdot \log q_{uy} - n_y \cdot r \cdot \log q_y$$

$$E(\mathbf{u}/\mathbf{y}) = n_{uy} \cdot k \cdot \log q_{uy} + r \cdot \left(n_{uy} \cdot \log q_{uy} - n_y \cdot \log q_y\right). \quad (14)$$

Видим, что вклад в условную энтропию на “входе с выхода (через систему)” определяется членами

$$k \cdot n_{uy} \cdot \log(q_{uy}) = \frac{k}{k+r} \cdot \log(N_{uy}) \text{ и } r \cdot \left(n_{uy} \cdot \log(q_{uy}) - n_y \cdot \log(q_y)\right) = \left(\left(\frac{r}{k+r} \cdot \log(N_{uy})\right) - \log(N_y)\right), \quad (15)$$

т.е. “уравнение регрессии” \equiv “уравнению влияния выхода на вход”:

$$E(\mathbf{u}/\mathbf{y}) = \log(N_{uy}) - \log(N_y). \quad (16)$$

3. СВЯЗЬ ШЕННОНОВСКОЙ ЭНТРОПИИ n -ДЕТЕРМИНИРОВАННЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ И ПОНЯТИЯ ЭНТРОПИИ НА ОСНОВЕ БП

Для последовательности символов текста \mathbf{u} , составляющих i -й участок стационарности длины $m^{(i)}$, вся данная последовательность однозначно воспроизводится Источником данных (абстрактным *детерминированным* автоматом) по начальной n_u -ке символов (следовательно, однозначно определяется ею). Поэтому при известных начальных n_u отсчетах i -го “гена” неопределенность относительно остальных $m^{(i)} - n_u$ символов настоящего “гена” отсутствует и в этом случае мера неопределенности (условная энтропия) становится равной нулю:

$$H\left(\frac{\{u_{n_u}, u_{n_u+1}, \dots, u_{M-1}\}}{\{u_0, \dots, u_{n_u-1}\}}\right) = 0 \quad (17)$$

Следовательно, полагая всю текстовую последовательность \mathbf{u} ограниченной одним участком стационарности процесса, для энтропии данного процесса получаем:

$$\begin{aligned} H(\mathbf{u}) &= H(\{u_0, u_1, \dots, u_{M-1}\}) = H(\{u_0, \dots, u_{n_u-1}\}, \{u_{n_u}, u_{n_u+1}, \dots, u_{M-1}\}) = \\ &= H(\{u_0, \dots, u_{n_u-1}\}) + H\left(\frac{\{u_{n_u}, u_{n_u+1}, \dots, u_{M-1}\}}{\{u_0, \dots, u_{n_u-1}\}}\right) = \\ &= H(\{u_0, \dots, u_{n_u-1}\}) + 0 = H(\{u_0, \dots, u_{n_u-1}\}) \end{aligned} \quad (18)$$

Таким образом, неопределенность всей последовательности \mathbf{u} связана лишь с неопределенностью выбора самой начальной n_u -ки. Поэтому мера неопределенности (шенноновская энтропия) всего текста, относительно которого известно, что он является участком стационарности (“геном”) в рамках ММ ГК, совпадает с энтропией начальной (служащей начальным условием для детерминированного автомата) n_u -ки:

$$H(\mathbf{u}) = H(\{u_0, u_1, \dots, u_{M-1}\}) = H(\{u_0, \dots, u_{n_u-1}\}). \quad (19)$$

Если же при этом все варианты из $(q_u^k)^{n_u}$ возможных начальных n_u -ок равновероятны, то шенноновская энтропия

$$\begin{aligned} H(\mathbf{u}) &= H(\{u_0, \dots, u_{n_u-1}\}) = - \sum_j^{q^n} p(\{u_0, \dots, u_{n_u-1}\}) \cdot \log p(\{u_0, \dots, u_{n_u-1}\}) = \\ &= - q^n \cdot \frac{1}{q^n} \cdot \log \frac{1}{q^n} = n \log q = E(\mathbf{u}) \end{aligned} \quad (20)$$

становится равной энтропии $E(\mathbf{u})$, выражающейся через БП текста \mathbf{u} .

К аналогичному выражению можно прийти и при рассмотрении совместной текстовой последовательности $\mathbf{uy} \equiv \{uy_0, uy_1, \dots, uy_{M-1}\} =$

$$= \left\{ \begin{bmatrix} u_0 \\ y_0 \end{bmatrix}, \begin{bmatrix} u_1 \\ y_1 \end{bmatrix}, \dots, \begin{bmatrix} u_{M-1} \\ y_{M-1} \end{bmatrix} \right\} \text{ двух процессов } \mathbf{u} \text{ и } \mathbf{y} \text{ (11), для которой БП равны } \{q_{uy},$$

$n_{uy}\}$, и, следовательно, существует Источник данных в виде детерминированного дискретного автомата, однозначным образом воспроизводящего сигнал \mathbf{uy} по его n_{uy} начальным векторным отсчетам. Отсюда:

$$\begin{aligned} H(\mathbf{uy}) &= H(\{uy_0, uy_1, \dots, uy_{M-1}\}) = H(\{uy_0, \dots, uy_{n_{uy}-1}\}, \{uy_{n_{uy}}, uy_{n_{uy}+1}, \dots, uy_{M-1}\}) = \\ &= H(\{uy_0, \dots, uy_{n_{uy}-1}\}) + H\left(\{uy_{n_{uy}}, uy_{n_{uy}+1}, \dots, uy_{M-1}\} \middle/ \{uy_0, \dots, uy_{n_{uy}-1}\}\right) = \\ &= H(\{uy_0, \dots, uy_{n_{uy}-1}\}) + 0 = H(\{uy_0, \dots, uy_{n_{uy}-1}\}) \end{aligned} \quad (21)$$

В предположении же о равновероятном распределении вероятностей всех

$$(q_{uy}^{(k+r)})^{n_{uy}} \text{ возможных вариантов начальных векторных } n_{uy}\text{-ок } \{uy_0, \dots, uy_{n_{uy}-1}\}$$

опять приходим к равенству шенноновской энтропии и энтропии, выражающейся через БП

$$\begin{aligned}
H(\mathbf{u}\mathbf{y}) &= H(\{u_0, \dots, u_{n_{uy}-1}\}) = - \sum_j^{q_{uy}^{n_{uy}}} p(\{u_0, \dots, u_{n_{uy}-1}\}) \cdot \log p(\{u_0, \dots, u_{n_{uy}-1}\}) = \\
&= - q_{uy}^{n_{uy}(k+r)} \cdot \frac{1}{q_{uy}^{n_{uy}(k+r)}} \cdot \log \frac{1}{q_{uy}^{n_{uy}(k+r)}} = n_{uy}(k+r) \log q_{uy} = E(\mathbf{u}\mathbf{y})
\end{aligned} \tag{22}$$

Средняя условная энтропия Шеннона для тех же последовательностей \mathbf{u} и \mathbf{y} , при тех же предположениях равняется

$$\begin{aligned}
H(\mathbf{u}/\mathbf{y}) &= H(\mathbf{u}\mathbf{y}) - H(\mathbf{y}) = E(\mathbf{u}\mathbf{y}) - E(\mathbf{y}) \equiv \\
&\equiv E(\mathbf{u}/\mathbf{y}) = n_{uy}(k+r) \log q_{uy} - n_y r \log q_y,
\end{aligned} \tag{23}$$

т.е. совпадает с выражением для функции ненадежности (14), записываемой в терминах БП сигналов шифратора, для которого \mathbf{u} и \mathbf{y} являются входом и выходом соответственно.

4. ВЫРАЖЕНИЕ “РАССТОЯНИЯ ЕДИНСТВЕННОСТИ” ШИФРАТОРА ЧЕРЕЗ БАЗОВЫЕ ПАРАМЕТРЫ ЕГО ТЕКСТОВЫХ СИГНАЛОВ

В соответствии с выведенными выше выражениями произведем оценку расстояния единственности на основе БП входного \mathbf{u} и выходного \mathbf{y} текстов шифратора. По аналогии с [1] распишем условие $E(\mathbf{u}/\mathbf{y}) = 0$, при котором неопределенность открытого текста обращается в нуль:

$$\begin{aligned}
0 &= E(\mathbf{u}/\mathbf{y}) = E(\mathbf{u}\mathbf{y}) - E(\mathbf{y}) = n_{uy} \cdot (k+r) \cdot \log q_{uy} - n_y^* \cdot r \cdot \log q_y \\
n_y^* \cdot r \cdot \log q_y &= n_{uy} \cdot (k+r) \cdot \log q_{uy}.
\end{aligned} \tag{24}$$

Теперь найдём размерность n_y^* Источника шифротекста, при помощи которого однозначным образом восстанавливается оставшийся шифротекст на участке одного “гена” (по построению параметров (6))

$$n_y^* = \frac{n_{uy} \cdot (k + r) \cdot \log(q_{uy})}{r \cdot \log(q_y)} = \frac{E(uy)}{r \cdot \log(q_y)}. \quad (25)$$

В результате, приходим к формуле для “расстояния единственности” в терминах базовых параметров, внешне имеющей ту же запись, что и формула, полученная ранее в [3]:

$$n_y^* = \frac{\log(\overline{m_{uy}})}{D_y}, \quad (26)$$

где, как и в [2], в качестве оценки средней длины участка стационарности векторного текстового сигнала **uy** берётся выражение:

$$\overline{m_{uy}} = \left(q_{uy}^{(k+r) \cdot n_{uy}} \right)^{1/2}, \quad (27)$$

а в качестве оценки избыточности r -компонентного векторного сигнала **y** выражение:

$$D_y = 1/2 \cdot \log q_y^r. \quad (28)$$

В соответствии с этими выражениями общая формула для произвольного вида шифратора (рис.2) с размерностью n_y^* теперь переписывается как

$$n_y^* = \frac{1/2 \cdot n_{uy} \cdot (k + r) \cdot \log(q_{uy})}{1/2 \cdot r \cdot \log(q_y)} = \frac{\log\left((q_{uy}^{(k+r)})^{n_{uy}}\right)^{1/2}}{1/2 \cdot \log q_y^r}. \quad (29)$$

При использовании формул (25)–(29) не трудно получить их полезную модификацию в виде неравенства:

$$n_y^* \geq \frac{\log\left((q_{uy}^{(k+r)})^{n_{uy}}\right) - \log\left((q_u^k)^{n_u}\right)}{\log q_y^r}. \quad (30)$$

Ещё раз укажем, что с точки зрения ММ нестационарного процесса в формализме ГК “переключающихся Источников данных” [2] настоящие рассуждения относятся только к одному гену векторного текста, так как оценка

БП производится именно для источника, порождающего текст в пределах своего “векторного” гена. Таким образом, утверждение о существовании порождающего текст Источника, могущего выступить его прогнозирующим оператором по первым n_y символам, справедливо только в пределах одного гена.

5. ВЫВОДЫ

1. Предложен подход (разделы 2, 4) к оценке параметров “функции ненадежности” и “расстояния единственности” для шифратора КС в форме ММ синхронного детерминированного дискретного автомата через измеряемые практически БП его входного, выходного и взаимного (вместо ненаблюдаемого внутреннего) векторных текстов.
2. Данный подход распространяется на случай как скалярных, так и векторных (“многоканальных”, в отличие, например, от [1]) ММ шифратора КС в виде детерминированного дискретного синхронного автомата Хаффмана.
3. Получены соотношения (14), (25), (29), связывающие указанные функции ненадежности и расстояния единственности шифратора КС с БП только наблюдаемых и измеряемых текстов шифратора КС.
4. Предложенный подход к оценке параметров одного шифратора КС не зависит от других блоков и структуры КС в целом и, поэтому, полученные соотношения справедливы для КС как с “закрытыми” [1] так и с “открытыми” [4] ключами.

5. Для практического использования формул при оценке исходные данные в несинхронных вариантах КС следует приводить к синхронной форме.

6. Условие отсутствия “идеальной секретности КС” по Шеннону [1] (наличие

решения $n_y^* < \infty$ уравнения (24) $H(u/y; n_y^*, \dots) = 0$) в данной модели КС

фактически выполняется всегда, даже для протяжённых нестационарных

процессов, если обрабатываемый участок данных содержит один “ген”

(или, что то же, когда мы по тем или иным причинам хотим и можем

аппроксимировать весь анализируемый процесс одним прогнозирующим

оператором с одним набором БП $= \{q_{opt}, n_{opt}\}$ и $N = N_{opt} = (q_{opt})^{n_{opt}}$ [2]).

При этом стационарные процессы с исходным $N = N_{opt}$ при любых

разумных размерах объёма фазового пространства $N \leq N_{opt}$ могут

превратиться в нестационарные, так как будут интерпретированы как

состоящие из многих стационарных участков (генов), и условие

“идеальной секретности КС” тем более не будет выполняться.

6. ЛИТЕРАТУРА

1. Шеннон К. Работы по теории информации и кибернетике. Пер. В.Ф. Писаренко – М.: Иностранная литература, 1963.

2. Кирьянов К.Г.// Труды III Межд. конф. «Идентификация систем и задачи управления» SICPRO'04. Москва, 28-30 января 2004 г. – М.: ИПУ РАН, 2004. с.187-208.

3. Горбунов А.А., Кирьянов К.Г.// Труды VIII Научной конференции по радиофизике, ННГУ, 7 мая 2004 г. – Нижний Новгород: ТАЛАН, 2004, с. 258-259.
4. Диффи У., Хеллман М.Э.// ТИИЭР. Март 1979. Т. 67, №3, с. 71-109.