

Е.Г.Василенко

АДАПТАЦИЯ ШУМА К ВХОДНЫМ ДАННЫМ В МОДЕЛИ СКРЫТОЙ ПЕРЕДАЧИ ИНФОРМАЦИИ

Новгородский государственный университет им. Ярослава Мудрого

The simple method improving some properties of the output sequence in the model of the latent digital data transmission is proposed. The dependence of the output sequence volume on the input data is investigated.

Данная статья посвящена одному из аспектов применения модели скрытой передачи информации по каналам связи. Идея модели была впервые предложена Б.Ф.Кирияновым в [1] и разработана Т.В.Жгун [2]. Модель включает в себя источник шума, пару управляющих генераторов разрядности n (расположенных в передатчике и приемнике соответственно) и секретный ключ K разрядности m , $m \leq n$. Перед началом работы оба генератора с помощью синхропосылки переводятся в режим синхронизма. Далее в канал передается шумовой сигнал из источника шума до тех пор, пока m разрядов выхода генератора не совпадут с ключом K . В этом случае передается пакет данных. Схема рассматриваемой модели представлена на рис.1.

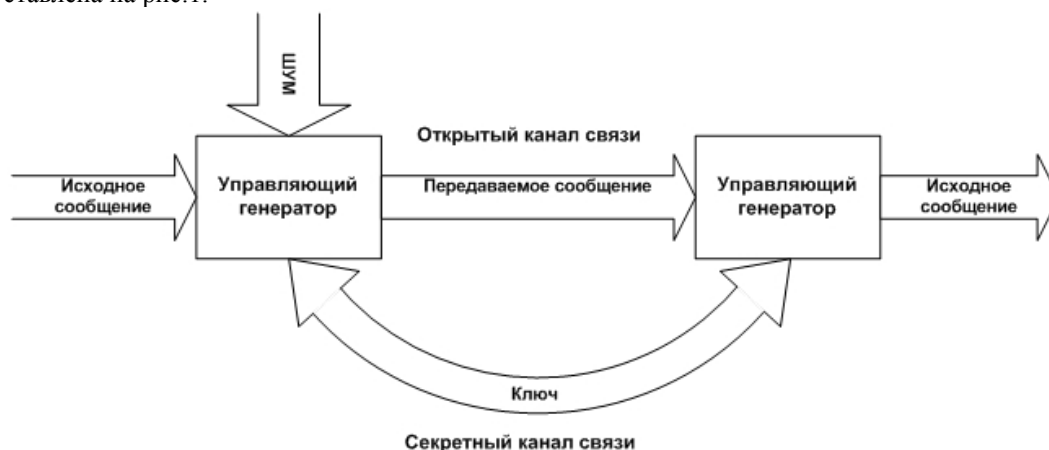


Рис.1. Система скрытой передачи информации

Одним из недостатков модели является то, что параметры генератора псевдослучайных последовательностей (ПСП), вырабатывающего шумовую последовательность, выбираются независимо от исходного сообщения. В статье предлагается простой метод адаптации шума, позволяющий приблизить частотные характеристики передаваемого сообщения к соответствующим характеристикам шума.

Введем необходимые обозначения.

Пусть I — исходное (или входное) сообщение, N — накладываемый шум, O — передаваемое (или выходное) сообщение*. Через n_i , n_o , n_n будем обозначать число символов в исходном, передаваемом сообщении и шуме соответственно. Пусть U — упорядоченное множество всех возможных символов, причем порядок задан функцией f (т.е. $f(\alpha)$ — номер символа α в алфавите U). Естественно, U включает в себя не менее двух

* Далее слова «сообщение», «сигнал», «последовательность» употребляются как идентичные.

символов: $|U| \geq 2$. На практике U часто совпадает с кодовой таблицей ASCII и $|U| = 256$.

Пусть алфавиты сообщений и шума являются некоторыми подмножествами U : $A_i \in U$, $A_o \in U$, $A_n \in U$. Под $P_i(\alpha)$, $P_o(\alpha)$, $P_n(\alpha)$ будем понимать вероятность, что наудачу выбранный символ из соответствующей последовательности совпадет с α . Очевидно, если $\alpha \notin A_i$, то $P_i(\alpha) = 0$, то же самое верно и для других алфавитов. Поскольку на практике эти вероятности будут оцениваться частотами встречаемости соответствующих символов, мы не будем делать различий между этими двумя понятиями.

Будем предполагать в дальнейшем, что в шуме встречаются все возможные символы, и он распределен равномерно:

$$A_n = U \quad \forall \alpha \in A_n, \quad P_n(\alpha) = p_n = \frac{1}{|U|}. \quad (1)$$

Пусть $M_i(f)$, $M_n(f)$ — математические ожидания порядка символа во входном сигнале и шуме, $D_i(f)$, $D_n(f)$ — их дисперсии. Определим характеристики выходного сигнала в данной криптосистеме:

$$M_o(f) = \frac{n_i}{n_i + n_n} \cdot M_i(f) + \frac{n_n}{n_i + n_n} \cdot M_n(f),$$

$$D_o(f) = \frac{n_i}{n_i + n_n} \cdot D_i(f) + \frac{n_n}{n_i + n_n} \cdot D_n(f) + \frac{n_i \cdot n_n}{(n_i + n_n)^2} \cdot (M_i(f) - M_n(f))^2. \quad (2)$$

Мы стремимся, чтобы параметры выходной последовательности, в том числе математическое ожидание и дисперсия, как можно меньше отличались от параметров шума. Поскольку на практике $n_i \ll n_n$, то из (1) следует, что $M_o(f) \approx M_n(f)$. Однако, как следует из (2), при значительном различии свойств шума и исходного сообщения, в частности величин $M_i(f)$ и $M_n(f)$, разница дисперсий $D_i(f)$ и $D_n(f)$ может оказаться заметной.

В силу чего возникает такое отклонение? Алфавит передаваемого сообщения является объединением алфавитов шума и исходного сообщения: $A_o = A_i \cup A_n$. Очевидно, в общем случае алфавиты шума и исходного сообщения могут пересекаться, т.е. $A_n \cap A_i \neq \emptyset$. В силу этого символы из множества $A_n \cap A_i$ будут встречаться в передаваемом сообщении чаще остальных. Понятно, что входной сигнал, вообще говоря, может обладать произвольной структурой, но можно изменить шумовой сигнал таким образом, чтобы частоты встречаемости в передаваемом сигнале почти не отличались.

Потребуем, чтобы передаваемое сообщение было распределено равномерно:

$$P_o(\alpha) = p_n. \quad (3)$$

Определим параметры шума, при которых это условие будет выполнено. Поскольку $A_o = A_i \cup A_n$ и $n_o = n_i + n_n$, то из (3) следует:

$$P_n(\alpha) \cdot \frac{n_n}{n_i + n_n} + P_i(\alpha) \cdot \frac{n_i}{n_i + n_n} = p_n. \quad (4)$$

Отсюда

$$\begin{cases} P_n(\alpha) = p_n \quad \forall \alpha \in A_n \setminus A_i, \\ P_n(\alpha) = \frac{n_i + n_n}{n_n} \cdot p_n - P_i(\alpha) \cdot \frac{n_i}{n_n} \quad \forall \alpha \in A_i. \end{cases} \quad (5)$$

Поскольку вероятности, по определению, лежат в диапазоне $[0...1]$, из (5) получаем требование к входному сигналу:

$$\forall \alpha \in A_i \quad 0 \leq \frac{n_i + n_n}{n_n} \cdot p_n - P_i(\alpha) \cdot \frac{n_i}{n_n} \leq 1,$$

из чего следует:

$$\forall \alpha \in A_i \quad P_i(\alpha) \leq \frac{n_i + n_n}{n_n} \cdot p_n \cdot \frac{n_n}{n_i} = \frac{n_i + n_n}{n_i} \cdot p_n, \quad (6)$$

$$\forall \alpha \in A_i \quad P_i(\alpha) \geq \left(\frac{n_i + n_n}{n_n} \cdot p_n - 1 \right) \cdot \frac{n_n}{n_i} = \frac{(n_i + n_n) \cdot p_n - n_n}{n_i}. \quad (7)$$

Из систем неравенств (6) и (7) выберем по наиболее жесткому ограничению. Тогда решение совокупности этих систем сведется к решению системы

$$\begin{cases} \frac{n_i + n_n}{n_i} \cdot p_n \geq \max_{\alpha} P_i(\alpha), \\ \frac{(n_i + n_n) \cdot p_n - n_n}{n_i} \leq \min_{\alpha} P_i(\alpha). \end{cases}$$

После преобразований получим:

$$\frac{n_n}{n_i} \geq \max \left\{ \frac{\max_{\alpha} P_i(\alpha) - p_n}{p_n}, \frac{p_n - \min_{\alpha} P_i(\alpha)}{1 - p_n} \right\}.$$

Нетрудно видеть, что чем ближе частотные свойства входного сообщения к свойствам шума, тем меньше пороговое отношение «шум-сигнал», при котором достигается поставленное условие. В идеале при $\max_{\alpha} P_i(\alpha) = \min_{\alpha} P_i(\alpha) = p_n$ шум вообще не требуется.

Исследуем, как меняется правая часть (8) в зависимости от мощности алфавита U . С учетом (1) имеем

$$\frac{n_n}{n_i} \geq \max \left\{ |U| \cdot \max_{\alpha} P_i(\alpha) - 1, \frac{1 - |U| \cdot \min_{\alpha} P_i(\alpha)}{|U| - 1} \right\}.$$

Первое из выражений в правой части (8) дает нам нижнюю границу отношения «шум-сигнал», которую мы должны обеспечить, чтобы компенсировать часто встречающиеся в исходном сообщении символы. Она линейно зависит от максимальной частоты встречаемости символов в исходном сообщении $\max_{\alpha} P_i(\alpha)$ и в пределе равна $|U| - 1$. Впрочем, предельный случай имеет место, когда алфавит входного сообщения включает в себя лишь один символ, что не встречается на практике.

Второе выражение в (8) является пороговым значением отношения «шум-сигнал», при котором могут быть компенсированы редко встречающиеся символы. Нетрудно видеть, что оно возрастает с убыванием $\min_{\alpha} P_i(\alpha)$, все же не превосходя величины $\frac{1}{|U| - 1}$. В связи с

этим возникает вопрос: существенно ли второе ограничение? Положим

$$A = \max_{\alpha} P_i(\alpha), \quad B = \min_{\alpha} P_i(\alpha), \quad p_n \leq A \leq 1, \quad 0 < B \leq p_n$$

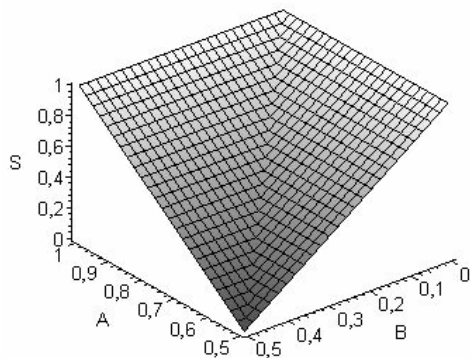
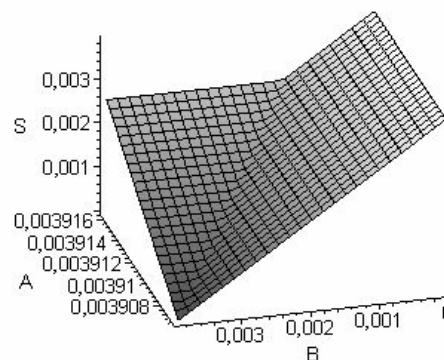
и рассмотрим функцию

$$S(|U|, A, B) = \max \left\{ |U| \cdot A - 1, \frac{1 - |U| \cdot B}{|U| - 1} \right\}.$$

Разрешая неравенство

$$|U| \cdot A - 1 \leq \frac{1 - |U| \cdot B}{|U| - 1}$$

относительно A , получим условие, при котором будет актуально второе ограничение:

Рис.3. График функции $S(2, A, B)$ Рис.2. График функции $S(256, A, B)$

$$A \leq \frac{1-B}{|U|-1}. \quad (9)$$

Построим график функции S при $|U|=2$ (рис.2) и при $|U|=256^*$ (рис.3).

Итак, второе ограничение может быть существенным лишь при малых значениях A , близких к $1/|U|$. Из (9) следует, что при

$$A = \frac{1-B}{|U|-1} \quad (10)$$

достигается равенство выражений в правой части (8), т.е. объем информации, необходимой для того, чтобы компенсировать часто встречающиеся символы, совпадает с объемом информации, который требуется, чтобы «добрать» редко встречающиеся символы во входном сообщении. Сравним величины отклонений максимальной частоты встречаемости A и минимальной частоты встречаемости B от идеального состояния $1/|U|$ при условии (10):

$$\frac{A - \frac{1}{|U|}}{\frac{1}{|U|} - B} = \frac{1}{|U|-1}.$$

Следовательно, «компенсация» отклонения частоты в большую сторону $A - 1/|U|$ эквивалентна с точки зрения требуемой длины шумовой последовательности «компенсации» отклонения частоты встречаемости в меньшую сторону в $|U|-1$ раз большей. Впрочем, это и не удивительно — в первом случае нам необходимо «подтянуть» частоты встречаемости почти всех символов.

Из (8) кроме того следует, что длина шумовой последовательности, которая требуется, чтобы обеспечить одинаковую встречаемость символов в выходном сигнале, растет с возрастанием мощности алфавита $|U|$. Действительно, чем больше мощность алфавита, тем меньше вероятность встречи каждого отдельного символа при их равновероятном распределении, а значит, тем больше потребуется членов последовательности, чтобы «добрать» редко встречающиеся и «компенсировать» часто встречающиеся символы.

Из условия (4) следует, что параметры выходного сигнала при данной адаптации шума совпадут с исходными параметрами самого шума:

$$M'_o(f) = M_n(f), \quad D'_o(f) = D_n(f),$$

* Второй график построен с большей точностью по оси A – при обычном масштабе «перегиб» неразличим.

что и являлось целью проведенной адаптации.

Практическая реализация такого варианта может быть осуществлена следующим образом: если на выходе источника шума получили $\alpha \in A_n \setminus A_i$, то пропускаем его без обработки в выходную последовательность. В противном случае с помощью генератора псевдослучайных чисел моделируем распределение $P_n(\alpha)$ и передаем в выходную последовательность символ, полученный в результате испытания. Очевидно, данный метод адаптации шума требует предварительной обработки исходного сообщения для определения алфавита A_i и частот встречаемости $P_i(\alpha)$. Это делает невозможным поточное шифрование. Кроме того, данная схема требует несколько больше временных ресурсов для осуществления процесса кодирования — ведь теперь нам нужно две ПСП. Заметим, что предложенный метод построен в предположении, что мы располагаем источником идеального шума. В реальных приложениях распределение как шума, так и вспомогательной ПСП будет, скорее всего, отличаться от равномерного. Для определения влияния этих отклонений на свойства выходной последовательности и установления допустимых величин отклонений требуются дополнительные исследования.

Вывод

Предлагаемая методика адаптации шумовой последовательности к исходным данным позволяет приблизить частотные характеристики передаваемого сообщения к соответствующим характеристикам шума. Благодаря этому можно ожидать улучшения криптостойкости передаваемой информации.

-
1. Кирьянов Б.Ф. Микропроцессорные средства в задачах имитации и обработки случайных сигналов. Ч.2. Новгород: НПИ, 1989. 48 с.
 2. Жгун Т.В., Кирьянов Б.Ф. // Вестник НовГУ. Сер.: Математика и информатика. 2002. №22. С.50-53.