УДК 681.3.067:681.3.016

В.Н. Куценко, Т.В. Левченко, В.В. Мясоедов

КОДЫ ПЕРЕМЕННОЙ ДЛИНЫ В ПРОБЛЕМЕ ПОМЕХОУСТОЙЧИВОСТИ СВЯЗИ

Введение

Ключевым понятием реальной передачи дискретных сообщений является различимость элементов, опирающаяся на синхронизацию и измерение/оцифровку (т.е. квалификацию) сигналов с учётом возможных ошибок "невключение" и "ложное включение" [1]. Упрощение формальных аспектов передачи данных, достигаемое применением двоичного алфавита, в зашумленных каналах порождает ряд инженерных трудностей, разрешаемых многообразными способами, зачастую приводящими к взаимоисключающим решениям. Это свидетельствует о том, что возможности методов помехоустойчивого кодирования исчерпаны.

Двоично-байтовый формат данных, хорошо зарекомендовавший себя в компьютерах, сильно ограничивает поле видения разработчиков аппаратуры передачи данных. В данной работе делается попытка переформулировать проблему помехоустойчивости за счет переноса акцентов с формата данных на синтаксис, в котором лексемы представляют формализованное выражение передачи данных азбукой кодов переменной длины.

Изменение видения проблемы помехоустойчивости и слабая изученность азбук с числом элементов, большим двух, определяют конкретный и конструктивный характер изложения. При этом основное внимание уделено выявлению пределов помехоустойчивой передачи данных по отношению к развиваемым средствам диагностики и автоматической коррекции вместо абстрактной защиты данных от искажений.

Очевидно, что существуют пределы возможностей автоматической коррекции ошибок передачи данных без избытка. Поэтому имеет смысл прагматическое исследование проблемы передачи текстовых данных. Имеется различие в семантике текстов естественного языка и оцифрованных аналоговых данных, т.к. последние практически не допускают статистически устойчивых алфавитов (таблиц) преобразования данных в коды переменной длины, что приводит к необходимости включения эмпирического алфавита в сообщение.

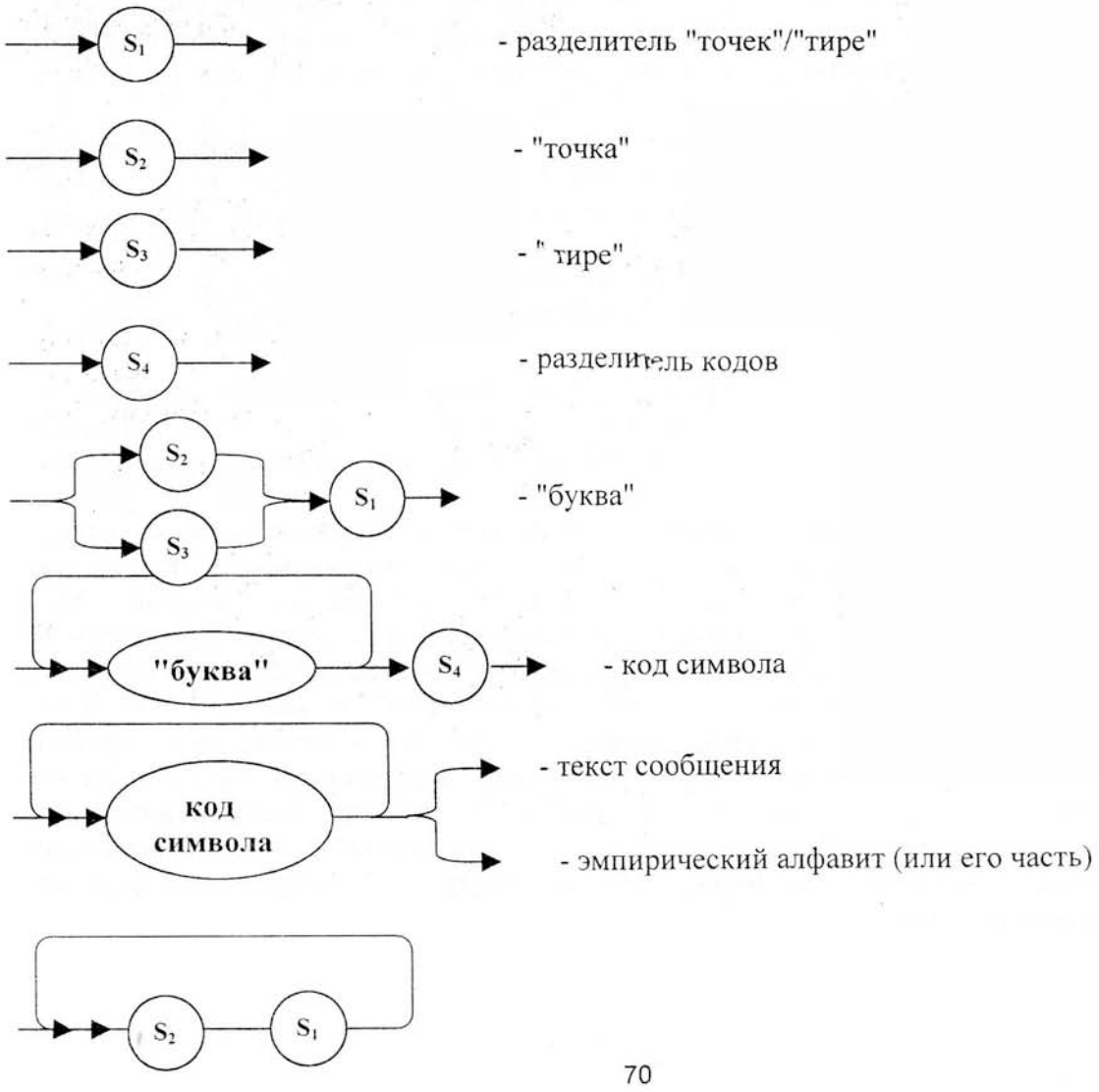
Временно́е разделение процедур подготовки данных и передачи сообщений позволяет уменьшить время использования физического канала. Это является существенным в сильно зашумленных каналах, особенно при передаче данных без обратной связи.

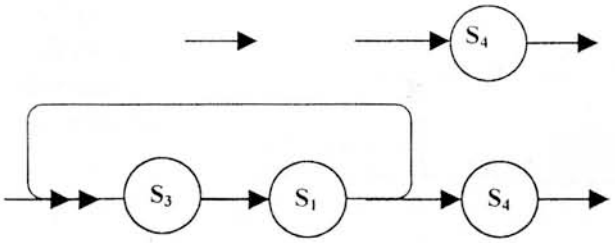
Классическим примером в текстовой семантике передачи данных является азбука Морзе, в которой семантический (эмпирический) алфавит, относящийся к подготовке данных, представлен кодами переменной длины. Азбука кодов переменной длины - это "точка", "тире", разделитель между "буквами", разделитель между кодами. Допустимы и собственные имена ("иероглифы"), длина которых больше максимальной длины кодового представления знаков эмпирического алфавита. "Иероглифом", в частности, является эмпирический алфавит, позволяющий восстанавливать формальную семантику сообщения для содержательной интерпретации. Другим примером собственного имени является последовательность "букв" азбуки, в той или иной степени идентифицирующая сообщение. Таким образом, собственные имена - это общий случай кодов переменной длины Интерпретация собственных имён, например, электронной подписи, определяется конвенциями передачи данных.

При недвусмысленном разделении содержательных и формальных аспектов передачи данных задача идентификации собственных имён не рассматривается, т.к. может быть решена дополнительными средствами подготовки данных вне режима передачи. Начало сообщения может быть представлено, например, длинной последовательностью "точек", конец - длинной последовательностью "тире". Предполагается, что эти последовательности достаточно длинны, чтобы фиксировать начало и конец сообщения независимо от числа ошибок, вызванных помехами в канале. "Начало сообщения" может предварять текст сообщения и/или эмпирический алфавит или его содержательную часть (в некоторой принятой кодировке), упорядоченный по возрастанию используемых для передачи кодов переменной длины. Тип соответствующего фрагмента сообщения должен быть указан при обращении к модему, либо допускать определение по самим данным. "Начало сообщения" может содержать и средства, облегчающие физическую синхронизацию данных.

Предметом рассмотрения, теперь не зависящим от типа передаваемых данных, является формальный синтаксис сообщений, использующих азбуку кодов переменной длины при передаче подготовленных данных в канале связи, рассматриваемый главным образом с точки зрения помехоустойчивости.

Самым примитивным способом представления элементов азбуки таких кодов является использование длительностей t_1, t_2, t_3, t_4, трактуемых как разделители s_1, s_2, s_3, s_4 между импульсами. Формальный синтаксис сообщений очень прост и в скрытом виде использует пятый элемент азбуки - физический импульс s_5, имеющий длину t_5, не отображённый в языке схем порождения:





Различие собственных имён ("эмпирический алфавит", "начало сообщения", "конец сообщения") и кодов переменной длины в языке схем порождения невыразимо и становится явным после введения ограничений реализации.

Постановка задачи

Семантический алфавит представляет собой таблицу, смещение от начала которой есть передаваемый байт (символ текста). Эта таблица при естественных ограничениях реализации содержит пары байтов: первый байт указывает длину кода, а второй – последовательность битов (0 – "точка", 1 – "тире"). Получаемая по этой таблице заготовка кода поступает в передающее устройство, где она оформляется в код в соответствии с формальным синтаксисом. Код в аналоговом виде отправляется в канал связи, пока счётчик остатка длины кода остаётся ненулевым.

Такая схема, вообще говоря, может быть реализована без счётчика длины кодов при использовании готовых кодов переменной длины с представлением разделителей парами битов, например, '0','1' ↔ s_1, '1','0' ↔ s_2, '1','1' ↔ s_3, '0','0' ↔ s_4. Тогда разделителям соответствуют целые числа t_1, t_2, t_3, t_4 пустых тактов генератора сигналов, разделённых целым числом t_5 тактов передачи сигнала в канал связи. Однако отслеживание интервалов по числу тактов предполагает их счёт. Кроме того, снижение удельного веса аналоговой части в устройстве за счёт усиления цифровой части делает схему негибкой, отклоняющейся от стандарта побитной обработки, и, что хуже всего, ограничивает возможности диагностики и коррекции ошибок передачи при наличии помех в канале связи.

Отмеченные выше обстоятельства в полной мере должны быть учтены при приёме сигналов из канала связи, оцифровываемых принимающим устройством таким образом, что s_1 ↔ '0', s_2 ↔ '1', s_3 ↔ '1','1', s_4 ↔ '0','0'. В этом случае код символа текста формируется цифровой частью, что разгружает принимающее устройство для более качественного решения задач квалификации и синхронизации сигналов.

Таким образом, требуется синтез устройства, выполняющего правило непосредственной обработки последовательностей битов и некоторые правила начала и окончания обработки. Формальная постановка задачи для конечного автомата, допускающего потенциально бесконечное состояние <коррекция>, имеет следующий вид:

Входной алфавит: '0', '1' .

Состояния: "бланк", "начало", "пусто", "точка", "тире".

Здесь "точка" и "тире" соответствуют решению о добавлении бита в формируемый код. "Начало" и "пусто" соответствуют выжиданию. "Бланк" соответствует передаче управления (коммутации).

Таблица переходов:

<бланк> + '1'	<начало>	R_1
<начало> + '1'	<тире>	R_2
<тире> + '1'	<?начало? коррекция>	R_3 (ошибка оцифровки)
<пусто> + '1'	<начало>	R_4
<начало> + '0'	<точка>	R_5
<точка> + '0'	<бланк>	R_6
<бланк> + '0'	<?пусто? коррекция>	R_7 (ошибка оцифровки)
<пусто> + '0'	<бланк>	R_8
<тире> + '0'	<пусто>	R_9
<точка> + '1'	<бланк>	R_{10}

В сформированный код при переходе в состояние "бланк" добавляют слева бит 1, представляющий длину кода l как 2^l, получая смещение в таблице символов текста. Нулевой и первый адрес в этой таблице резервируются. При максимальной длине кода 7 таблица содержит до 254 байтов текстовых символов. Например, при безошибочном приеме "· — " будет сформирован код '01' и адрес байта символа "А" имеет вид '00000101'.

При наличии ошибок оцифровки принимаемых сигналов достигается потенциально бесконечное состояние < коррекция >. Абстрактному логическому обнаружению не поддаются пакеты ошибок оцифровки ("невключение", "ложное включение"), переводящие, например код "· — " в код " — ·". Не всегда обнаружимы и ошибки "дробления"/"слияния" кодов - даже с учетом ограниченной реализации, а обнаруженные ошибки такого типа не всегда могут быть исправлены на основании формального синтаксиса. Поэтому логически необходимым является поэлементное сопровождение передачи кодов длиной кода, например, стробом в оплётке коаксиального кабеля, используемой в качестве дополнительного подканала. (Ошибки передачи длины кода не играют особой роли, т.к. длина кода может быть установлена и синтаксически.) Этим достигается высокая точность локализации и диагностики ошибок передачи дискретных данных в сочетании физического приемного устройства с распознающим принятый код конечным автоматом.

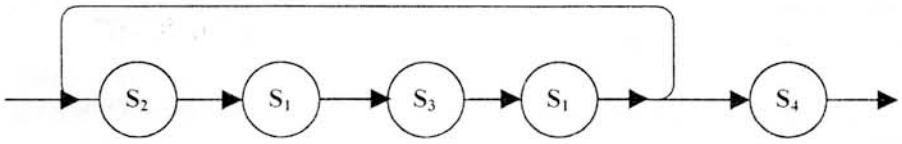
Возможности автоматической коррекции ошибок передачи ограничены качеством и стоимостью применяемых каналов связи. Эти возможности могут быть расширены тонкими формальными методами анализа оцифрованных сигналов, а также повышением качества сенсора в отношении квалификации сигналов и синхронизации их восприятия в зависимости от требуемой скорости приема/передачи и заданной различимости сигналов на фоне помех. При этом не последнюю роль играют параметры сигналов.

Синтез воспринимающего устройства и канала должен быть рассмотрен при реализации изложенного протокола кодов переменной длины (VaryingCodeLength-протокол, VCLP).

Основная часть

Основная цель решения задачи on-line анализа принимаемых кодов и сигналов - принятие решения о повторении передачи данных в случае, когда исчерпаны возможности автоматической коррекции. Наиболее достоверными критериями такого решения являются критерии, выводимые из самого процесса передачи с помощью виртуальной (в т.ч. локальной) обратной связи. Протокол приема кодов (оцифрованных сигналов) записывается в память, например, на флеш-карту для последующего анализа (при

необходимости) и/или для текущего выяснении идентификации сообщений по словарю собственных имен (каталог протокола приема). Идентификация повтора сообщения, начиная с некоторого обусловленного конвенцией места, - это достаточно длинная последовательность перемежающихся "точек" и "тире":



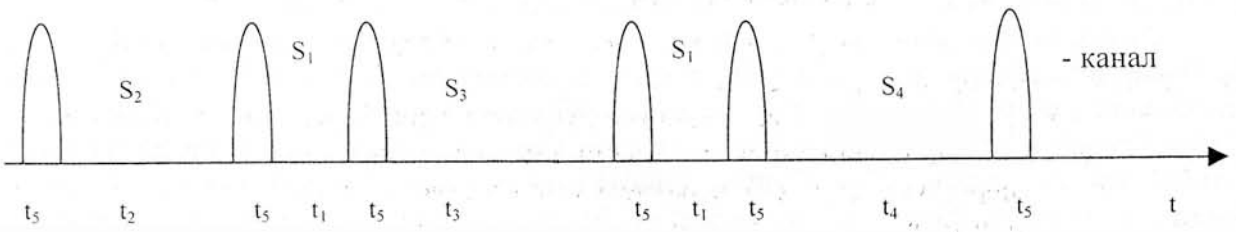
- "повтор сообщения".

Обычный критерий проверки совпадения переданных и принятых данных и повторение передачи данных до точного совпадения приводят к слабой сходимости ряда вероятностей безошибочной передачи, увеличивая среднее время передачи кода. При отсутствии обратной связи помехоустойчивость обычно обеспечивается избыточными данными - контрольными суммами и семантической избыточностью самого текста. Например, в известном методе случайной перестановки символов текста пакетные ошибки передачи "размазываются" по тексту, не затемняя его смысл.

При использовании виртуальной, т.е. локальной, обратной связи для решения вопроса о повторении передачи имеется возможность применить динамическую оценку качества передачи данных с помощью преобразования Фурье скользящего "окна" сигналов. Равноправие помех и сигналов в канале связи по отношению к виртуальной обратной связи позволяет игнорировать помехи, спектр которых отличается от динамически размытого спектра полезных сигналов. По-видимому, имеется также различие динамической спектральной интенсивности помехи и сигнала. Критерием решения о повторении передачи является неразличимость помех и сигналов по динамическому спектру и интенсивности при ненулевом динамическом образе разности ожидаемого и виртуального потока сигналов, независимо от синтаксиса передаваемых кодов.

Более эффективными являются критерии, учитывающие синтаксис кодов переменной длины и направленные на ограничение потенциальной бесконечности состояния <коррекция>. Такие критерии с некоторой оцененной погрешностью позволяют сконструировать конечный подавтомат, выводящий VCLP из потенциально бесконечного состояния. Поэтому ниже рассматривается возможности подавления помех в самом сенсоре.

В последовательном канале с независимым подканалом ошибки слияния /дробления кодов могут быть исключены синхронной передачей в подканале импульса, длительность которого меньше времени передачи кода на величину $t_1 + (t_4 + t_5)$. Различимость интервалов времени t_1, t_2, t_3, t_4, t_5 в таком канале является необходимым условием успешной передачи кодов. Случайные помехи при прецизионном отслеживании этих интервалов не имеют значения до тех пор, пока они не совпадают с возможным положением физического сигнала s_5 на временной диаграмме передачи кода. Ниже на рисунке приведена примерная временная диаграмма передачи кода "· — ":



Время передачи этого кода в сообщении есть

$$T_{\text{"}\cdot\_\text{"}} = 2t_1 + t_2 + t_3 + t_4 + 5t_5.$$

Выбор длительностей $t_1 \div t_5$ определяется настройкой на фактическое состояние канала при его предварительном прослушивании, либо после передачи обусловленной последовательности "точек"/"тире" с анализом в виртуальной обратной связи. Возможна и жесткая фиксация этих длительностей. В частности, при одиночных помехах начальные значения могут удовлетворять неравенствам треугольника $t_2 + t_3 > t_4$, $t_3 + t_4 > t_2$, $t_4 + t_2 > t_3$ и т.п. Мощность сигнала и его длительность t_5 также могут варьироваться в зависимости от уровня помех в канале. После адаптации длительностей становится возможным тонкое измерение (синхронизация) интервала кода.

Для передачи кода "· — " требуется такое же время, как для передачи кода " — ·". При прецизионной передаче выбор длительностей может быть уточнен таким образом, что коды разной длины не могут быть переданы как два кода меньшей длины. Это снизило бы требования к подканалу передачи длины кода. Вообще, время передачи некоторого символа из семантического алфавита имеет вид:

$$T_c = c_1 t_1 + c_2 t_2 + c_3 t_3 + t_4 + c_5 t_5 .$$

После восстановления кода из канала связи эта формула, вообще говоря, изменяется:

$$T_c^{'} = c_1^{'} t_1 + c_2^{'} t_2 + c_3^{'} t_3 + c_4^{'} t_4 + c_5^{'} t_5 .$$

Желательно, чтобы суммарное время передачи кодов различной длины занимало различное время независимо от возможностей дробления/слияния кодов, что приводит к недостижимому в последовательном канале соотношению

$$\left(c_1 - c_1^{'}\right)t_1 + \left(c_2 - c_2^{'}\right)t_2 + \left(c_3 - c_3^{'}\right)t_3 + \left(c_4 - c_4^{'}\right)t_4 + \left(c_5 - c_5^{'}\right)t_5 = 0 ,$$

тогда и только тогда, когда $c_i = c_i^{'}$, $i = 1,2,3,4,5$.

Другими словами, элементы кода должны быть линейно независимыми:

$$c_1 s_1 + c_2 s_2 + c_3 s_3 + c_4 s_4 + c_5 s_5 = 0 ,$$

если и только если $c_1 = c_2 = c_3 = c_4 = c_5 = 0$.

Это легко достигается в параллельно-последовательном канале, например, в пятипроводной линии связи, дополненной средствами передачи длины кода.

Скорость передачи данных в параллельно-последовательном канале может быть увеличена в несколько раз, если применить код постоянной длины МТК-2 вместо кода переменной длины. Однако такое решение фактически ухудшает различимость кодов при наличии помех и должно быть соотнесено с качеством связи, поскольку не предполагается наличие средств диагностики ошибок оцифровки сигналов. Увеличение числа линий (частот, цветов) в таком канале также не позволяет кардинально решить проблему автоматической коррекции ошибок [2] без диалога по обратной связи.

Раздельная оцифровка сигналов в параллельно-последовательном канале с VCLP позволяет наращивать средства автоматической коррекции воспринимаемых данных при сведении и оценке результатов независимой оцифровки сигналов (5 битов, из которых в точности два должны быть ненулевыми - 5-й и один из $s_1 \div s_4$). Главным остается исключение ошибок слияния/дробления кодов, достигаемое передачей длины кода в аналоговом (6-ой "провод") или цифровом виде (на каждый элемент кода).

Независимое поэлементное подавление помех методом автокорреляции [3] может быть дополнено оценкой корреляции цуга волн s_5 с цугами $s_1 \div s_4$. Коэффициент корреляции даёт оценку вероятности правильности оцифровки одного из четырех битов, преобразуемых в адрес символов входного алфавита в конечном автомате.

Если конечный автомат диагностирует ошибку оцифровки, то результат оцифровки корректируется по коэффициенту корреляции, следующему по величине. При этом состояние коррекции не прекращается, в памяти автомата накапливается последовательность четверок результатов оцифровки и рационального (целочисленного) представления четырех коэффициентов корреляции до обнаружения конца кода (это еще один подавтомат).

Алгоритм автоматической коррекции максимизирует вероятность синтаксически правильного декодирования, записывает всю ситуацию в долговременную память параллельно с работой основного конечного автомата, переведенного в состояние <бланк> по условному концу кода. Существенное улучшение обработки помех независимо от алгоритма автоматической коррекции может быть достигнуто за счет временно́го разделения параллельно-последовательного канала для передачи кодов (чёт) и импульсов синхронизации (нечет), реконструируемых с высокой точностью за счет выбора частот $s_1 \div s_5$ и амплитуд цугов волн. Это позволяет повысить доверие к оценкам автокорреляции и корреляции сигналов.

Заметим a priori [5], что при шифровании текстовых данных также проявляется необходимость передачи эмпирического алфавита для сжатия сообщения. Однако битовый стандарт предоставляет более тонкие возможности шифрования/дешифрования путём случайной [6] перестановки битов исходного байта, битов кода переменной длины и случайной коммутации линий связи (частот).

Случайная коммутация линий связи доставила бы наибольшие затруднения при взломе данных за счёт неопределённости выделения кода из потоков битов оцифрованных сигналов стандартной длительности в параллельно-последовательном канале.

Проверка предложенных схем решений до синтеза устройства приема/передачи может быть произведена методами имитационного моделирования [4], использующими датчик случайных чисел [6], усиленный возможностью генерировать псевдослучайные помехи с требуемой вероятностью. Расчет технических параметров устройств, выделение параллельных процессов, построение алгоритма автоматической коррекции, оценка времени передачи сообщений выходят за рамки анализа формальных свойств кодов переменной длины и за рамки настоящей статьи.

Выводы

Рассмотрение целостной проблемы передачи дискретных данных приводит к определенной классификации предметной области. Это достигается систематическим разделением содержательного и формального аспектов, цифрового и сенсорного восприятия, семантического и абстрактного выражений, последовательного и параллельно-последовательного способов, полезных и калибрующих сигналов.

Результатам являются развитые средства выявления пределов помехоустойчивости передачи данных, допускающие техническую реализацию.

Список литературы

1. Протоколы и методы управления в сетях передачи данных: Пер.с англ./Под ред. **Ф.Ф.Куо**.-М.: Радио и связь.-1985.-480с.

2. Кодирование информации (двоичные коды).//**Березюк Н.Т., Андрущенко А.Г., Мощицкий С.С. и др**.- Харьков: Вища школа.- 1978. – 252с.

3. **Кузьмин И.В., Кедрус В.А.** Основы теории передачи и кодирования.-К.: Вища школа.-1977.-280с.

4. **Семишин Ю.А.** Методические вопросы непроцедурного подхода к конструированию имитационных алгоритмов управления // УСИМ.- 1988.- №2.- с.31-35.

5. **В.Куценко, Т. Левченко, Н. Миронов, В. Мясоедов**. Основная проблема тестирования генераторов случайных чисел. - В сб.: Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні.-Науково-технічний збірник.- Випуск 5.-К.:НДЦ "Тезіс" НТУУ "КПІ".-2002.-213с.- С.130-133.

6. **В.Мясоедов**. Золотое сечение в шифровании данных. - В сб.: Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні.-Науково-технічний збірник.-Випуск 4.-К.:НДЦ "Тезіс" НТУУ "КПІ".-2002.-213с.- С.105.

Поступила 25.02.2003г.