УДК 629.391

А.К. ЮДИН[1], В.В. БАРАННИК[2]

[1]***Национальный авиационный университет, Киев, Украина***
[2]***Харьковский университет Воздушных Сил им. И. Кожедуба, Украина***

УСЕЧЕННОЕ ПРЕДСТАВЛЕНИЕ ДВОИЧНЫХ ДАННЫХ С ОГРАНИЧЕННЫМ ЧИСЛОМ СЕРИЙ В ПОЛИАДИЧЕСКОМ ПРОСТРАНСТВЕ

Излагается компактное представление двоичных данных, осуществляемое на борту беспилотного летательного аппарата (БПЛА). Приводится формулировка и доказательство теоремы об определении значения кода-номера для двоичных последовательностей, удовлетворяющих ограничениям на число двоичных серий и на допустимые позиции единичных элементов. Обосновывается эффективность такого представления.

полиадическое пространство, ограниченное число единичных серий, двумерная структурная избыточность

Введение

Одним из направлений совершенствования информационных и телекоммуникационных систем является организация доведения информации на основе использования беспилотных летающих аппаратов (БПЛА). Это позволит доводить информацию с труднодоступных районов и снизить нагрузку на существующие линии связи. В этом случае в качестве передаваемой информации могут быть как видеоданные, так и телеметрические данные. Данный элемент информационных систем может быть включен в цифровую сеть интегрального обслуживания. При этом вместо многочисленных стандартных сетей связи, предлагается построить единую сеть с учетом единых методов транспортирования всех видов информации по технологии асинхронного режима переноса пакетов фиксированной длины [1]. При этом одна из важнейших составляющих эффективности представления данных в информационных и телекоммуникационных системах, в том числе с использованием БПЛА, заключается в снижении времени доведения информации.

Для уменьшения объема исходных данных и соответственно снижения времени доведения данных предлагается организовывать компактное представление на основе исключения избыточности в двоичных данных.

Формулирование проблемы. Среди различных методов компрессии [2 – 5] двоичных данных наибольшая степень сжатия достигается для методов, исключающих двумерную структурную избыточность [6]. В работе [6] предложен метод компактного представления двоичных данных, полученных от источников информации различных видов (аудио, текст, статические и динамические изображения). В тоже время не были рассмотрены направления дополнительного сокращения времени обмена информацией в ТС на основе двоичного структурного кодирования в двухпризнаковом пространстве.

В связи с этим **цель статьи** состоит в разработке нумерации, обеспечивающей дополнительное повышение степени сжатии и снижение времени на обработку.

1. Обоснование возможности повышения эффективности двухпризнакового структурного кодирования

Для обоснования и выбора направлений дополнительного повышения степени сжатия рассмотрим выражения для двоичного кодирования в двухпризанковом пространстве [6]:

$$N(m,\Lambda,\vartheta)_j = \sum_{g=0}^{Z-1} \sum_{\xi=1}^{\Theta^{(\xi_g)}} \sum_{z=1}^{g+1} \sum_{i=1}^{m_z} a_{izj}\, p_{izj}^{(\xi)} \times$$

$$\times \prod_{\phi=z+1}^{Z} V\left(\vartheta_{\phi}^{(\xi)}, \Theta^{(\xi_g)}\right); \quad (1)$$

$$r_{i,zj}^{(\xi)} = \frac{(m_z - i + 1)\,!}{\left(\beta_{izj}^{(\xi)}\right)!\ \left(m_z - i + 1 - \beta_{izj}^{(\xi)}\right)!}; \quad (2)$$

$$\beta_{0zj}^{(\xi)} = 2\,\vartheta_z^{(\xi)}\,;\ a_{0jz}^{(\xi)} = 0\,, \quad (3)$$

где g – количество зон u-й двоичной последовательности, для которых число серий ϑ_{zu} совпадает с числом серий единиц ϑ_{zj} в соответствующих зонах j-й обрабатываемой последовательности;

$\Theta^{(\xi_g)}$ – количество векторов $\Theta^{(\xi)}$, удовлетворяющих условиям допустимости на число серий единиц и позиции с запретом появления единиц;

Ω_g – множество допустимых двоичных последовательностей, предшествующих обрабатываемые последовательности;

$p_{izj}^{(\xi)} = \left(r_{i-1,zj}^{(\xi)} - r_{izj}^{(\xi)}\right)$ – весовой коэффициент ij-го элемента z-й допустимой зоны обрабатываемой последовательности, зависящий от значений m_z и $\vartheta_z^{(\xi)}$;

a_{izj} – ij-й элемент z-й допустимой зоны обрабатываемой двоичной последовательности, при этом двоичными данными являются: представленные в цифровом виде видеоданные или телеметрическая информация;

$\beta_{izj}^{(\xi)}$ – рекуррентный параметр, равный количеству двоичных перепадов (переходов между «0» и «1») для последовательности, состоящей из $(m_z - i + 1)$ необработанных элементов;

$\beta_{0zj}^{(\xi)}$ – начальное значение параметра $\beta_{izj}^{(\xi)}$, равно $\beta_{0zj}^{(\xi)} = 2\,\vartheta_z^{(\xi)}$;

$V\left(\vartheta_z^{(\xi)}, \Theta^{(\xi_g)}\right)$ – количество двоичных подпоследовательностей, полученных для z-й допустимой зоны по количеству серий единиц, равных $\vartheta_z^{(\xi)}$ для вектора $\Theta^{(\xi_g)}$;

$N\left(\vartheta_z^{(\xi)}, \Theta^{(\xi_g)}\right)$ – количество комбинаций длиной m_z элементов с числом серий единиц $\vartheta_z^{(\xi)}$ вектора $\Theta^{(\xi_g)}$, предшествующих z-й зоне обрабатываемой последовательности.

Анализ выражений (1) – (3) показывает, что для дополнительного повышение степени сжатия и для уменьшения времени на обработку требуется выполнить следующие действия.

1. Перейти от нумерации двоичных данных, учитывающей все подмножества Ω_g, к нумерации в пределах одного подмножества, зависящего от одной комбинации величин $\vartheta_z^{(\xi)}$.

2. Обеспечить возможность организации параллельного вычисления кода-номера, т.е. разработать нумерацию, для которой процесс вычисления весовых коэффициентов будет взаимонезависим от процессов вычисления весовых коэффициентов для других элементов обрабатываемой двоичной последовательности.

3. Организовать взаимооднозначное представление данных с учетом отсутствия информации об их статистических характеристиках.

2. Разработка усеченного структурного кодирования двоичных данных в двухпризнаковом пространстве

Сформулируем определение основных понятий усеченного структурного представления данных с учетом ограниченного количества единичных серий.

Определение 1. Множество, составленное из двоичных последовательностей называется множеством ***усеченным (фрагментарным)*** двоичным полиадическим кодированием с учетом ограниченного числа серий единиц, если оно одновременно удовлетворяет ограничениям на:

– позиции с допустимым появлением единичных элементов (задается вектором ограничений на диапазон значений обрабатываемых элементов Λ);

– число серий единиц $\vartheta_z^{(k)}$ в каждой допустимой зоне. Задается вектором $\Theta^{(k)}$ значений величин $\vartheta_z^{(k)}$.

Из определения следуют различия множеств $\Psi(m,\Lambda,\,\vartheta)$ и $\Psi(m,\Lambda,\,\Theta^{(k)})$, обусловленные различными ограничениями, накладываемыми на допустимые последовательности.

В первом случае для нас важно знать суммарное количество серий единиц по всем допустимым зонам обрабатываемых последовательностей

$$\vartheta=\sum_{z=1}^{Z}\vartheta_z^{(k)}\,, \tag{4}$$

где k – индекс комбинации величин $\vartheta_z^{(k)}$, сумма которых равна ϑ .

Тогда множество $\Psi(m,\Lambda,\,\vartheta)$ образуется из двоичных последовательностей, удовлетворяющих ограничениям:

$$\vartheta=\sum_{z=1}^{Z}\vartheta_{z,\,j}\;. \tag{5}$$

Причем значения величин $\vartheta_z^{(k)}$ изменяются в пределах

$$0\le\vartheta_z^{(k)}\le\min\left\{\vartheta;\;\left[\left[\frac{m_z+1}{2}\right]\right]\right\}. \tag{6}$$

Отсюда следует, что в общем случае $k\ge 1$.

В этом случае объем допустимого множества $V(m,\Lambda,\,\vartheta)$ двоичных последовательностей формируется как сумма объемов допустимых множеств двоичных последовательностей, удовлетворяющих вектору ограничений на число серий единиц $\Theta^{(k)}$:

$$V(m,\Lambda,\vartheta)=\sum_{k=1}^{K}V\left(\Theta^{(k)}\right). \tag{7}$$

Во втором случае ограничения накладываются не на сумму величин $\vartheta_z^{(k)}$ по всем допустимым зонам, а на значение числа серий единиц в каждой допустимой зоне. В этом случае ограничения задаются в виде вектора $\Theta^{(k)}$, компонентами которого являются величины $\vartheta_z^{(k)}$: $\Theta^{(k)}=\left\{\vartheta_1^{(k)},...,\vartheta_z^{(k)},...,\vartheta_Z^{(k)}\right\}$. Тогда множество $\Psi(m,\Lambda,\,\Theta^{(k)})$ формируется из двоичных последовательностей, удовлетворяющих следующим ограничениям:

$$\vartheta_{z,\,j}=\vartheta_z^{(k)}\,, \tag{8}$$

где $\vartheta_{z,\,j}$ – количество серий единиц в z-й допустимой зоне j-й двоичной последовательности.

Из анализа выражений (5) и (7) вытекает, что множество $\Psi(m,\Lambda,\,\Theta^{(k)})$ является подмножеством множества $\Psi(m,\Lambda,\,\vartheta)$.

Следовательно, основное отличие заключается в том, что в множестве $V(m,\Lambda,\,\vartheta)$ учитываются ограничения на число серий единиц для всей двоичной последовательности, т.е. на суммарное количество серий единиц во всех зонах, а в множестве $V(m,\Lambda,\,\Theta^{(k)})$ учитываются ограничения на число серий единиц $\vartheta_z^{(k)}$ в каждой допустимой зоне $z=\overline{1,Z}$.

Примечание. В дальнейшем, поскольку будет анализироваться только один вектор величин $\vartheta_z^{(k)}$, то в его обозначение индекс k не вносится, т.е. вместо $\Theta^{(k)}$ используется Θ

Для определения объема $V(m,\Lambda,\,\Theta)$ множества $\Psi(m,\Lambda,\,\Theta)$ пронумерованных двоичных последовательностей, содержащих заданное количество серий единиц, в каждой допустимой зоне докажем следующую теорему.

Теорема об объеме усеченного множества двоичных полиадических чисел по количеству серий единиц. Количество двоичных последовательностей, удовлетворяющих ограничениям (4) – (6) равно

$$V(m,\Lambda,\Theta)=V(\Theta)=\prod_{z=1}^{Z}V(\vartheta_z,\Theta);\qquad(9)$$

$$V(\vartheta_z,\Theta)=\binom{m_z+1}{2\vartheta_z}=\frac{(m_z+1)!}{(2\vartheta_z)!\ (m_z+1-2\upsilon_z)},\qquad(10)$$

где ϑ_z – значение числа серий для z-й допустимой зоны двоичной последовательности A;

Θ – вектор, элементами которого является комбинация количеств серий единиц ϑ_z в допустимых зонах $\Theta=\{\vartheta_1,...,\vartheta_z,...,\vartheta_Z\}$;

Z – количество допустимых зон в двоичной последовательности;

m_z – количество двоичных элементов в z-й допустимой зоне;

$V(\vartheta_z,\Theta)$ – количество допустимых двоичных последовательностей, полученных для z-й допустимой зоны по количеству серий единиц, равному ϑ_z для вектора Θ;

$V(\Theta)$ – количество допустимых двоичных последовательностей, полученное с учетом обработки всех Z допустимых зон для вектора Θ значений величин ϑ_z.

Доказательство. Доказательство теоремы вытекает из доказательства теоремы для определения объема множества $V(m,\Lambda,\vartheta)$. Величина $V(m,\Lambda,\vartheta)$ находится по формуле [4]:

$$V(m,\Lambda,\vartheta)=\sum_{k=1}^{K}V\left(\Theta^{(k)}\right)=$$

$$=\sum_{k=1}^{K}\prod_{z=1}^{Z}\frac{(m_z+1)!}{\left(2\vartheta_z^{(k)}\right)!\ \left(m_z+1-2\vartheta_z^{(k)}\right)!},\qquad(11)$$

где Z – количество допустимых зон в двоичной последовательности;

K – количество векторов $\Theta^{(k)}$ (количество комбинаций длиной Z, составленных из элементов $\vartheta_z^{(k)}$;

m_z – количество двоичных элементов в z-й допустимой зоне.

Выражение (11) учитывает двоичные комбинации для различных вариантов векторов $\Theta^{(k)}$. Поскольку по условию теоремы объем $V(m,\Lambda,\Theta)$ равен количеству двоичных последовательностей, удовлетворяющих ограничению (6), т.е. учитывается только одна комбинация вектора Θ. Поэтому величина $V(m,\Lambda,\Theta)$ равна

$$V(m,\Lambda,\Theta)=\prod_{z=1}^{Z}\frac{(m_z+1)!}{(2\vartheta_z)!\ (m_z+1-2\upsilon_z)}.$$

Теорема доказана.

Доказанная теорема позволяет определить для заданных параметров m, Λ и Θ объем $V(m,\Lambda,\Theta)$ множества двоичных полиадических чисел по количеству серий единиц. Для формирования кода-номера конкретной двоичной последовательности необходимо разработать соответствующий процесс нумерации двоичных полиадических чисел по количеству серий единиц.

Теорема о усеченной нумерации двоичных полиадических чисел по количеству серий единиц. Для двоичной последовательности $A^{(j)}=\{a_{ij}\}_{i=\overline{1,m}}$ с выявленными ограничениями на позиции единиц $\Lambda=\{\lambda_i\}_{i=\overline{1,m}}$ и на количество серий единиц ϑ можно сформировать код-номер $N(m,\Lambda,\Theta)_j$, вычисляемый по формулам:

$$N(m,\Lambda,\Theta)_j=\sum_{z=0}^{Z-1}\sum_{i=1}^{m_z}a_{izj}\left(r_{i-1,zj}-r_{i,zj}\right)\prod_{\phi=z+1}^{Z}V(\vartheta_\phi);$$

$$V(\vartheta_\phi) = \frac{(m_\phi + 1)!}{(2\vartheta_\phi)!\ (m_\phi + 1 - 2\vartheta_\phi)!}; \qquad (12)$$

$$r_{i,zj} = \frac{(m_z - i + 1)!}{(\beta_{izj})!\ (m_z - i + 1 - \beta_{izj})!};\ \beta_{0zj} = 2\vartheta_z;$$

$$a_{0jz} = 0, \qquad (13)$$

где $p_{izj} = \left(r_{i-1,zj} - r_{izj}\right)$ – весовой коэффициент ij-го элемента z-й допустимой зоны обрабатываемой последовательности, зависящий от значений m_z и ϑ_z;

a_{izj} – ij-й элемент z-й допустимой зоны обрабатываемой последовательности;

β_{izj} – рекуррентный параметр, равный количеству двоичных перепадов (переходов между «0» и «1») для последовательности, состоящей из $(m_z - i + 1)$ необработанных элементов:

$$\beta_{izj} = \beta_{i-1,zj} - \left| a_{i-1,zj} - a_{izj} \right|;$$

β_{0zj} – начальное значение параметра β_{izj}, равно $\beta_{0zj} = 2\vartheta_z$. При этом считается, что $a_{0jz} = 0$. Тогда, если первый элемент двоичной последовательности равен единице, то он образует первый перепад с «0» на «1»;

$V(\vartheta_z)$ – количество двоичных подпоследовательностей, полученных для z-й допустимой зоны по количеству серий единиц, равных ϑ_z для вектора Θ;

$N(\vartheta_z)$ – количество комбинаций длиной m_z элементов с числом серий единиц ϑ_z вектора Θ, предшествующих z-й зоне обрабатываемой последовательности.

Доказательство. Доказательство теоремы будет проводиться в два этапа.

На первом необходимо вывести выражение для нахождения количества двоичных комбинаций множества $\Psi(\Theta)$, предшествующих последовательности $A^{(j)}$.

Второй этап посвящается определению величины кода-номера $N(\vartheta_z, \Theta)$ двоичной подпоследовательности, принадлежащей z-й зоне Θ-го вектора значений величин ϑ_z.

Рассмотрим первый этап доказательства, состоящего в получении выражения для формирования кода-номера $N(A^{(j)})$ в пределах множества $\Psi(\Theta)$. В соответствии с выражением (10) количество допустимых последовательностей содержащих все зоны вектора Θ равно количеству перестановок с повторениями, составленных из Z укрупненных элементов, на динамический диапазон которых наложены ограничения, равные соответственно величинам $V(\vartheta_z)$. Величина $V(\vartheta_z)$ равна количеству двоичных подпоследовательностей, образующих соответствующую z-ю зону, у которых число серий единиц равно ϑ_z. В связи с этим z-м элементом множества $\Psi(\Theta)$ является величина кода-номера $N(\vartheta_z)$, для которого выполняется неравенство

$$N(\vartheta_z) \le V(\vartheta_z) - 1.$$

Тогда последовательности, составленные из элементов $N(\vartheta_z)$, $z = \overline{1, Z}$, являются полиадическими числами. Отсюда следует, что количество $N(A^{(j)})$ комбинаций множества $\Psi(\Theta)$, предшествующих последовательности $A^{(j)}$ равно коду-номеру соответствующего полиадического числа, т.е.

$$N(A^{(j)}) = \sum_{z=1}^{Z} N(\vartheta_z) \prod_{\phi=z+1}^{Z} V(\vartheta_\phi). \qquad (14)$$

На втором этапе доказательства теоремы требуется доказать, что величины $N(\vartheta_z)$ являются структурными кодами двоичных последовательностей с ограниченным числом серий единиц. Согласно определению двоичного полиадического кодирования по числу серий единиц и правила нумерации 2.1 величина $N(\vartheta_z)$ равна количеству двоичных комбинаций длиной m_z элементов с числом серий единиц ϑ_z вектора Θ, предшествующих z-й зоне обрабатываемой последовательности. Поэтому значение величины $N(\vartheta_z)$ находится по формуле

$$N(\vartheta_z)=\sum_{i=1}^{m_z} a_{izj}\left(r_{i-1,zj}-r_{i,zj}\right). \qquad (15)$$

Подставив выражения (15) в формулу (14) получим искомую систему соотношение (12) и (13).

Теорема доказана.

Таким образом, система выражений (12), (13) обеспечивает формирование кода-номера двоичным последовательностям в полиадическом пространстве, удовлетворяющих ограничениям (4) – (6) без внесения погрешности.

Оценка характеристик процесса компактного представления данных с заранее неизвестными свойствами для разработанного усеченного представления показала, что относительно двумерного структурного представления двоичных данных:

– степень сжатия дополнительно увеличивается в среднем в 1,8 раза;

– время обработки снижается на 65%.

Заключение

Таким образом, можно сделать выводы:

1. Доказано, что дополнительное увеличение степени сжатия двоичных данных возможно в результате учета дополнительных ограничений на комбинацию числа серий единиц в допустимых зонах.

2. На основе сформулированной и доказанной теоремы разработано усеченное структурное кодирование двоичных данных в двухпризнаковом пространстве. Полученное представление обеспечивает безпогрешностное сжатие двоичных данных с заранее неизвестными статистическими свойствами. При этом динамика изменения статистических свойств в соседних фрагментах обрабатываемых массивах данных не влияет существенным образом на степень сжатия (степень сжатия изменяется в пределах **10%**).

3. Разработанное кодирование обеспечивает относительно двумерного структурного представления двоичных данных увеличение степени сжатия в среднем в **1,8** раза и снижение времени обработки на **65%**.

Литература

1. Богданов В.Н., Вихлянцев П.С., Симонов М.В. Построение корпоративной сети на базе сети АТМ общего пользования // "ИНФОРМОСТ" – "Радиоэлектроника и Телекоммуникации". – 2003. – № 5 (29). – С. 22-33.

2. Ватолин В.И., Ратушняк А., Смирнов М., Юкин В. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео. – М.: ДИАЛОГ – МИФИ, 2002. – 384 с.

3. Зубарев Ю.В., Дворкович В.П. Цифровая обработка телевизионных и компьютерных изображений. – М.: Международный центр научной и технической информации, 1997. – 212 с.

4. Рябко Б.Я. Сжатие данных с помощью стопки книг // Проблемы передачи информации. – 1980. – № 3. – С. 7-13.

5. Рябко Б.Я. Дважды универсальное кодирование // Проблемы передачи информации. – 1984. – № 3. – С. 9-15.

6. Баранник В.В., Юдин А.К. Двоичное полиадическое кодирование по количеству серий единиц // Радиоэлектроника и информатика. – 2005. – № 2. – С. 56-63.

Поступила в редакцию 9.03.2006

Рецензент: д-р техн. наук, проф. П.Ф. Поляков, Украинская государственная академия железнодорожного транспорта, Харьков.