

Девиз: "Нейрон"

*Тема: "Биномиальное сжатие информации в
сенсорных сетях"*

СОДЕРЖАНИЕ

Введение	3
1. Сенсорные сети	4
2. Сжатие информации без потерь	9
3. Двоичная биномиальная система счисления	13
3.1 Нумерационная функция биномиальной системы счисления	13
3.2 Биномиальное нумерационное сжатие информации	14
4. Биномиальное сжатие информации в сенсорных сетях	17
4.1 Модель передачи данных в сенсорной сети	17
4.2 Метод биномиального сжатия информации в сенсорных сетях	22
Выводы	28
Литература	29

ВВЕДЕНИЕ

Современные тенденции в развитии цифровой и микропроцессорной техники, ее миниатюризация, снижение стоимости и потребляемой мощности привели к возникновению распределенных измерительных, вычислительных и управляющих систем. Ярким примером подобных систем являются сенсорные сети. Благодаря их простой установке, экономичности, надежности и эффективности использования, сенсорные сети получили широкое развитие во всех областях жизнедеятельности человека. [1]

Специфические особенности построения и функционирования сенсорных сетей привели к необходимости создания и развития новых подходов при передаче и обработке передаваемой в сети информации. Современные исследования нацелены на разработку протоколов обмена данными между элементами сети, выбора экономичного маршрута передачи сообщений, способов повышения помехоустойчивости сенсорных сетей. [2-4] Однако проблеме сжатия информации в сенсорных сетях уделяется достаточно мало внимания.

В данной работе предлагается одно из решений проблемы сжатия информации в сенсорных сетях. За основу был взят метод биномиального нумерационного сжатия. [12,13] Предлагаемый метод биномиального сжатия информации в сенсорных сетях довольно прост в реализации, не требует высоких аппаратурных затрат и весьма эффективен при сжатии распределенной между элементами сети информации с неизвестной статистикой появления сообщений.

1. СЕНСОРНЫЕ СЕТИ

Беспроводная сенсорная сеть – это беспроводная система, которая представляет собой распределенную, самоорганизующуюся и устойчивую к отказам отдельных элементов сеть, узлами которой являются специальные устройства – моты.

Мот (от англ. motes - пылинки) – миниатюрное вычислительно-коммуникационное устройство. Это небольшая плата, которая содержит процессор, оперативную память и флэш-память, ЦАП и АЦП, радиочастотный приемопередатчик, источник питания и датчики. Типы датчиков могут быть совершенно произвольными. Чаще всего используются датчики температуры, давления, влажности, освещения, вибрации и т.д. [2]

Работа сенсорных сетей основана на совместной работе большого числа крошечных узлов, которые состоят из модулей сбора и обработки данных, а так же приемо-передающего модуля. Такие сети имеют значительные преимущества перед набором традиционных датчиков.

Сенсорная сеть состоит из большого числа узлов (мотов), которые густо расположены близко к наблюдаемому явлению. Положение мотов не нужно предварительно рассчитывать. Это позволяет случайным образом располагать их в труднодоступных местностях или использовать для операций, которые требуют быстрого реагирования. С другой стороны, это означает, что сетевые протоколы и алгоритмы работы мотов должны обладать возможностью самоорганизации. Еще одной уникальной особенностью сенсорных сетей является совместная работа отдельных узлов. Моты оснащены процессором. Поэтому вместо передачи исходных данных, они могут их обрабатывать, выполняя простые вычисления и передавать далее только необходимые и частично обработанные данные. Описанные выше особенности обеспечивают широкий спектр применения сенсорных сетей.

Многие протоколы и алгоритмы были разработаны для традиционных беспроводных одноранговых сетей, поэтому они не очень хорошо подходит для уникальных особенностей и требований сенсорных сетей. Основные различия сенсорных и одноранговых сетей:

- Количество узлов сенсорной сети может быть на несколько порядков выше, чем узлов в одноранговой сети.
- Узлы плотно расположены.
- Узлы подвержены сбоям.
- Топология сенсорных сетей может часто изменяться.
- Узлы в основном используют широковещательные сообщения, в то время как большинство одноранговых сетей основаны на связи "точка-точка".
- Узлы ограничены в питании, вычислительных мощностях, и памяти.
- Узлы не могут иметь глобальный идентификационный номер из-за большого количества накладных расходов и большого количества датчиков.

Так как узлы в сети расположены плотно, соседние узлы могут оказаться очень близко друг к другу. Следовательно, multi-hop связи в сенсорных сетях будут потреблять меньше энергии, чем прямые связи. Кроме того, можно использовать низкую мощность сигнала передачи данных, что полезно в скрытых наблюдениях. Multi-hop связи могут эффективно преодолевать некоторые трудности при распространении сигнала на дальние расстояния в беспроводной связи. Одним из наиболее важных ограничений для узлов является малое потребление энергии. Моты имеют ограниченные источники энергии. Итак, в то время как традиционные сети направлены на достижение высокого качества сигнала, сетевые протоколы мотов должны сосредоточиться главным образом на сохранение энергии. Они должны обладать механизмами, которые дают пользователю возможность продления времени жизни мота за счет либо снижения пропускной способности, либо увеличения времени задержки передачи данных. [1]

Основные проблемы современных беспроводных сенсорных сетей:

- Назначение новой задачи по контролю объекта, исходя из его местоположения и требования минимальной нагрузки сети;
- Сжатие данных и контроль передаваемого трафика;
- Выбор стратегии при вычислениях, в которых задействовано несколько узлов (с одной стороны требуется обмен данными, с другой – требуется затраты батарей на пересылку).

Структура сенсорных сетей. Сенсорная сеть состоит из программного и аппаратного компонента. Аппаратный компонент представляет собой множество оснащенных приемопередатчиками микропроцессорных устройств, с помощью которых возможна передача информации между данными устройствами в сети. Программный компонент агрегирует канальный, сетевой уровень и уровень приложений, позволяющие поддерживать передачу данных от всех узлов сети к некоторым узлам-«стокам» (базовым станциям), которые соединены с промышленными SCADA-системами. Сенсорная сеть – самоорганизующаяся структура: в случае выхода узлов из строя или добавления новых узлов в сеть, поток данных от других узлов к базовой станции перераспределяется с учётом изменения конфигурации сети. На рисунке 1 показана сенсорная сеть и возможные изменения ее конфигурации. [3]

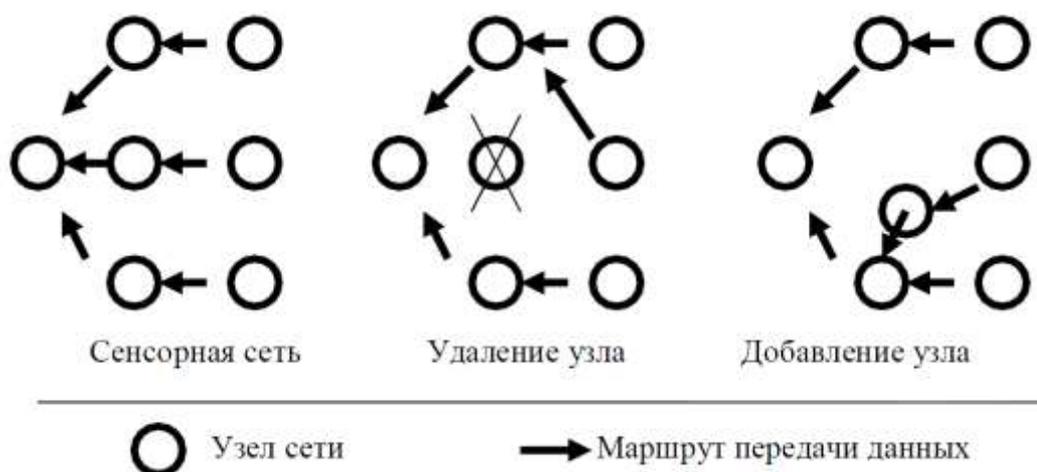


Рисунок 1 - Возможные варианты изменения конфигурации сенсорной сети

Архитектура сенсорных сетей. Узлы, как правило, расположены случайным образом по всей территории наблюдения. Каждый из них может осуществлять сбор данных и знает маршрут передачи данных обратно в центральный узел, конечному пользователю. Данные передаются с помощью multi-hop архитектуре сети. Центральный узел может общаться с менеджером задач через Интернет или спутник. Стек протоколов включают в себя информацию о мощности и информации о маршрутах, содержит данные о сетевых протоколах, помогает эффективно общаться посредством беспроводной среды, и содействует совместной работе узлов. Слой управления мобильностью (MAC) определяет и регистрирует передвижение узлов, поэтому всегда существует маршрут для передачи данных в центральный узел и узлы могут определять своих соседей. А зная своих соседей узел может сбалансировать энергопотребление, работая совместно с ними. Менеджер задач планирует и составляет расписания сбора информации для каждого региона отдельно. Не все узлы в одном регионе необходимы для выполнения задач зондирования в одно и то же время. Как результат, некоторые узлы выполняют больше задач, чем другие, это зависит от их мощности. Эти все слои и модули необходимы для того чтобы узлы работали вместе и стремились к максимальной энергоэффективности, оптимизации маршрута передачи данных в сети, а также совместно использовали ресурсы друг друга. Без них, каждый узел будет работать индивидуально. С точки зрения всей сенсорной сети эффективнее, если узлы будут работать совместно друг с другом, что способствует продлению времени жизни самой сетей. [1]

Применение беспроводных сенсорных сетей. Сенсорные сети могут состоять из различных типов датчиков: сейсмических, датчиков определения магнитного поля, тепловых, инфракрасных, акустических, и т.д. Используя сенсорные сети можно осуществлять самые разнообразные измерения условий окружающей среды. Моты могут использоваться для непрерывного

зондирования, обнаружения и идентификации событий. Концепция микрозондирования и беспроводное соединение открывают много новых областей применения для таких сетей. Классификация сенсорных сетей по основным направлениям: военное применение, исследование окружающей среды, здравоохранение, использование в домах и других коммерческих областях. Но можно расширить эту классификацию и добавить больше категорий, например исследование космического пространства, химическая обработка и ликвидации последствий стихийных бедствий.

Военное применение. Беспроводные сенсорные сети могут быть неотъемлемой частью военного управления, связи, разведки, наблюдения и систем ориентирования. Быстрое развертывание, самоорганизации и отказоустойчивость – это характеристики сенсорных сетей, которые делают их перспективным инструментом для решения поставленных задач. Поскольку сенсорные сети могут быть основаны на плотном развертывании одноразовых и дешевых узлов, то уничтожение некоторых из них во время военных действий не повлияет на военную операцию так, как уничтожение традиционных датчик. Поэтому использование сенсорных сетей лучше подходит для сражений.

Экологическое применение. Некоторые из направлений в экологии, где применяют сенсорные сети: отслеживание движения птиц, мелких животных и насекомых; мониторинг состояния окружающей среды, с целью выявления ее влияния на сельскохозяйственные культуры и скота; орошения; широкомасштабный мониторинга земли и исследования планет; химическое и биологическое обнаружение; обнаружение лесных пожаров; метеорологические или геофизические исследования; обнаружение наводнений; и исследование загрязнения.

Применение в медицине. Одним из применений в медицине является устройства для инвалидов; мониторинг пациентов; диагностика; мониторинг использования медикаментов в больницах; сбор физиологических данных человека; и мониторинга врачей и пациентов в больницах. [1]

2. СЖАТИЕ ИНФОРМАЦИИ БЕЗ ПОТЕРЬ

Существует достаточно много различных алгоритмов сжатия информации без потерь. [5] Наиболее распространенными и широко известными являются:

- Метод Хаффмана;
- Метод арифметического сжатия;
- Метод Лемпеля-Зива (LZ метод).

Метод Хаффмана.

Под префиксным кодированием понимается кодирование событий посредством кодов, состоящих из целого числа бит. Под префиксностью подразумевается, что никакой из кодов при просмотре слева направо не является частью другого, более длинного кода. [6]

Для получения кодов Хаффмана события также выписываются в порядке убывания их вероятностей, после чего берутся два наименее вероятные события и удаляются из списка. На их место помещается событие с суммарной вероятностью. К коду менее вероятного события слева добавляется бит “0”, к коду более вероятного – “1”. Эта операция выполняется последовательно, пока коды не будут полностью сформированы. Если наименее вероятное событие уже было обработано ранее, т.е. представляет собой группу событий, то бит добавляется слева к коду каждого из событий в группе. [7]

Чем больше длина символов входного алфавита (для которого строится код Хаффмана), тем меньше избыточность выходного текста и тем выше степень сжатия. Однако при этом значительно возрастают требования к памяти данных и к быстродействию программы, поскольку количество кодов равно количеству символов исходных букв. Избыточность кода Хаффмана в значительной мере зависит от того, на сколько вероятности появления символов близки к отрицательным степеням числа 2. Основной недостаток

алгоритма Хаффмана заключается в сложности процесса построения системы кодов. [6]

Характеристики канонического алгоритма Хаффмана:

- Степень сжатия: 8, 1.5, 1 (лучшая, средняя, худшая степени).
- Симметричность по времени: 2:1 (за счет того, что требует двух проходов по массиву сжимаемых данных).
- Характерные особенности: один из немногих алгоритмов, который не увеличивает размера исходных данных в худшем случае (если не считать необходимости хранить таблицу перекодировки вместе с файлом). [5]

Арифметическое кодирование

Арифметическое кодирование является методом, позволяющим упаковывать символы входного алфавита без потерь при условии, что известно распределение частот этих символов. Арифметическое кодирование является оптимальным, достигая теоретической границы степени сжатия.

Текст, сжатый арифметическим кодером, рассматривается как некоторая двоичная дробь из интервала $[0, 1)$. Результат сжатия можно представить как последовательность двоичных цифр из записи этой дроби. Каждый символ исходного текста представляется отрезком на числовой оси с длиной, равной вероятности его появления и началом, совпадающим с концом отрезка символа, предшествующего ему в алфавите. Сумма всех отрезков, очевидно, должна равняться единице. Если рассматривать на каждом шаге текущий интервал как целое, то каждый вновь поступивший входной символ “вырезает” из него подинтервал пропорционально своей длине и положению. [7,8]

При рассмотрении этого метода возникают две проблемы: во-первых, необходима вещественная арифметика неограниченной точности, и во-вторых, результат кодирования становится известен лишь при окончании входного потока. Дальнейшие исследования, однако, показали, что можно практически без потерь обойтись целочисленной арифметикой небольшой

точности (16-32 разряда), а также добиться инкрементальной работы алгоритма: цифры кода могут выдаваться последовательно по мере чтения входного потока. [6]

Подобно алгоритму Хаффмана, арифметический кодер также является двухпроходным и требует передачи вместе с закодированным текстом еще и таблицы частот символов.

Характеристики арифметического алгоритма:

- Лучшая и худшая степень сжатия: лучшая > 8 (возможно кодирование менее бита на символ), худшая – 1.
- Плюсы алгоритма: обеспечивает лучшую степень сжатия, чем алгоритм Хаффмана (на типичных данных на 1-10%).
- Характерные особенности: так же как кодирование по Хаффману, не увеличивает размера исходных данных в худшем случае. [5]

Алгоритмы семейства LZ

Алгоритм LZ был описан А. Лемпелем и Я. Зивом и получил свое название по первым буквам фамилий авторов и году публикации. С тех пор традиционно аббревиатуры почти всех словарных алгоритмов начинаются с «LZ», вследствие чего словарные алгоритмы и методы часто называют алгоритмами и методами группы LZ. В данном пункте рассматриваются алгоритмы семейства LZ77. [5]

В алгоритме LZ77 в качестве словаря выступают N последних обработанных символов. Каждая кодовая комбинация состоит из трех компонентов (кодовая триада). Первые два компонента указывают на строку словаря, идентичную начальной части кодируемой символьной последовательности: один из этих компонентов является смещением строки в словаре (смещение определяется относительно текущей обрабатываемой позиции), а второй – длиной этой строки. Третий компонент представляет собой первый символ кодируемой последовательности, не принадлежащий строке, описываемой первыми двумя компонентами (символ записывается в кодовой триаде в своем исходном представлении). Длина совпадения

ограничивается сверху некоторым числом M . В процессе работы алгоритма на каждом шаге в словаре осуществляется поиск строки, имеющей наибольшее совпадение с кодируемой символьной последовательностью. В случае, когда совпадение в словаре отсутствует, обрабатывается только первый символ последовательности, для чего в кодовой триаде указывается нулевое смещение (обрабатываемый символ является третьим компонентом триады).

Выбор параметров скользящего буфера самым непосредственным образом влияет на эффективность кодирования. С одной стороны, увеличение размера буфера поиска означает расширение словаря и ведет к повышению вероятности нахождения идентичных строк, а увеличение размера буфера просмотра позволяет учитывать при кодировании более длинные совпадения. С другой стороны, размер кодового пространства, необходимого для хранения смещений строк и длин совпадений, возрастает с увеличением соответствующих составляющих скользящего окна по логарифмическому закону. [6-8]

Характеристики алгоритмов семейства LZ77:

- Степени сжатия: определяются данными, обычно 2-4.
- Типы данных: алгоритмы универсальны, но лучше всего подходят для сжатия разнородных данных, например файлов ресурсов.
- Симметричность по скорости: примерно 10:1; если алгоритм обеспечивает хорошее сжатие, то декодер обычно гораздо быстрее кодера.
- Характерные особенности: обычно медленно сжимают высокоизбыточные данные; из-за высокой скорости восстановления идеально подходят для создания дистрибутивов программного обеспечения. [5]

3. ДВОИЧНАЯ БИНОМИАЛЬНАЯ СИСТЕМА СЧИСЛЕНИЯ

3.1 Нумерационная функция биномиальной системы счисления

Биномиальными системами счисления называются позиционные системы счисления с биномиальными коэффициентами в качестве оснований [9,10]. В данной работе используется линейная биномиальная система счисления с двоичным алфавитом, числовая функция которой имеет вид

$$F = x_{r-1}C_{n-1}^{k-q_{r-1}} + \dots + x_i C_{n-r+i}^{k-q_i} + \dots + x_1 C_{n-r+1}^{k-q_1} + x_0 C_{n-r}^{k-q_0}, \quad (1)$$

где r – количество разрядов биномиального числа (длина), $r \in 1, 2, \dots$;

k – максимальное количество единиц q_{max} в биномиальном числе;

i – порядковый номер разряда, $i = 0, 1, \dots, r-1$;

x_i – биномиальная двоичная цифра – 0 или 1;

n – целочисленный параметр системы счисления;

q – число единиц в биномиальном числе;

q_i – сумма единичных значений цифр x_i от $(r-1)$ -го разряда до $(i+1)$ -го включительно:

$$q_i = \sum_{j=i+1}^r x_j; \quad (2)$$

$$i = 0, 1, \dots, r-1; x_r = 0.$$

В качестве весового коэффициента i -го разряда в числовой функции (1) выступает биномиальный коэффициент $C_{n-r+i}^{k-q_i}$. Он зависит как от позиции $i = 0, 1, \dots, r-1$ рассматриваемого разряда, так и от суммы q_i предшествующих этому разряду двоичных значений цифр x_i . Последняя зависимость характерна только для структурных систем счисления и придает им помехоустойчивость и структурообразующие свойства. [9]

Существует возможность преобразования кода с постоянным весом в неравномерное биномиальное число. Суть преобразования состоит в том, что из n -разрядной равновесной комбинации удаляются младшие подряд стоящие единичные разряды до первого появления нуля или младшие подряд

стоящие нулевые разряды до первого появления единицы. Аналогичным образом строится алгоритм обратного преобразования неравномерного биномиального числа в n -разрядную равновесную комбинацию. К биномиальной комбинации дописываются со стороны младших разрядов одинаковые подряд стоящие нулевые или единичные разряды до длины n . [10] Алгоритмы прямого и обратного преобразований используются при построении методов сжатия и восстановления информации на основе биномиальной системы счисления.

3.2 Биномиальное нумерационное сжатие информации

Под нумерационным кодированием принято понимать такое кодовое отображение $f: A \rightarrow B$, которое является отображением исходного множества A на множество B целых неотрицательных чисел. Таким образом, каждому сообщению, порожденному источником A , может быть поставлен в соответствие некоторый номер, то есть осуществляется нумерация. [11]

В работах [12,13] был рассмотрен метод биномиального нумерационного сжатия и восстановления информации. Данный метод строится на основе нумерационной функции (1) двоичной биномиальной системы счисления. Биномиальное нумерационное сжатие позволяет сжимать двоичные сообщения, присваивая им номера соответствующих биномиальных чисел. Данный метод ставит в соответствие двоичной n -разрядной последовательности A блок длиной $\lceil \log_2(n+1) \rceil$ двоичных разрядов, хранящий число k единиц в исходной двоичной последовательности, и блок переменной длины $\lceil \log_2(C_n^k) \rceil$, хранящий номер F биномиального числа для биномиальной системы счисления с параметрами n и k .

Основные этапы работы метода биномиального нумерационного сжатия двоичной информации следующие:

1. [Ввод двоичной последовательности] Считать из входного потока данных n -разрядную двоичную последовательность;
2. [Подсчет числа единиц] В исходной двоичной последовательности подсчитать количество единичных разрядов (после выполнения данной операции двоичную последовательность можно рассматривать как равновесную кодовую комбинацию длиной n с количеством k единиц (весом равновесной комбинации));
3. [Переход от равновесной комбинации к биномиальному числу] Данная операция выполняется в соответствии с алгоритмом, описанным ранее;
4. [Вычисление номера биномиального числа] Для полученного на предыдущем шаге биномиального числа вычислить номер в соответствии с числовой функцией (1) для биномиальной системы счисления. Так же на данном этапе вычисляется количество разрядов, необходимых для хранения номера биномиального числа в соответствии с формулой $\lceil \log_2(C_n^k) \rceil$;
5. [Вывод числа единиц] Вывести в выходной поток число k единиц, состоящее из $\lceil \log_2(n+1) \rceil$ двоичных разрядов;
6. [Вывод номера биномиального числа] Вывести в выходной поток номер F биномиального числа, состоящий из $\lceil \log_2(C_n^k) \rceil$ двоичных разрядов.

Восстановление сжатой двоичной последовательности осуществляется в обратном порядке. При этом переход от номера к неравномерному биномиальному числу реализуется на основе последовательного (начиная со старшего разряда) подбора и вычитания из номера значений биномиальных коэффициентов. Вычисление значений разрядов $x_i \in \overline{0,1}$ неравномерного биномиального числа длины r осуществляется в соответствии с рекуррентным соотношением:

$$x_i = H(F(i) - \rho_i), \quad (3)$$

где $F(r-1) = F_j$, $F(i-1) = F(i) - x_i \rho_i$, $\rho_i = C_{n-r+i}^{k-q_i}$,

$$H(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} - \text{функция Хэвисайда.}$$

Этапы выполнения метода биномиального нумерационного восстановления информации следующие:

1. [Ввод числа единиц] Считать из входного потока данных число k единиц, состоящее из $\lceil \log_2(n+1) \rceil$ двоичных разрядов.
2. [Ввод номера биномиального числа] Считать из входного потока данных номер F биномиального числа, состоящий из $\lceil \log_2 C_n^k \rceil$ разрядов.
3. [Преобразование номера в биномиальное число] В соответствии с методом биномиального нумерационного декодирования при помощи выражения (3) вычислить значение каждого из разрядов биномиального числа для заданного номера F .
4. [Преобразование биномиального числа в равновесный код] Выполнить преобразование биномиального числа в равновесный код. Суть преобразования состоит в конкатенации к младшим разрядам биномиального числа определенного количества двоичных нулей или единиц с целью получить равновесную кодовую комбинацию, состоящую из n двоичных разрядов.
5. [Вывод двоичной последовательности] Вывести в выходной поток n разрядную двоичную последовательность.

4. БИНОМИАЛЬНОЕ СЖАТИЕ ИНФОРМАЦИИ В СЕНСОРНЫХ СЕТЯХ

4.1 Модель передачи данных в сенсорной сети

Исходя из проведенного обзора, можно выделить ряд уникальных требований, свойственных сенсорным сетям: малое энергопотребление при работе мотов и ограниченные аппаратно-программные затраты. Большинство мотов реализуется с использованием микропроцессоров, следовательно уровень аппаратно-программных затрат определяется исходя из объема используемой памяти и размера исполняемой программы. Большая часть энергетического ресурса в моте расходуется на передачу данных с использованием приемопередающей аппаратуры, а меньшая часть энергии потребляется микропроцессорным блоком. Причем, потребление электроэнергии каждым устройством мота в значительной степени зависит от времени работы данного устройства. В результате, сократив время работы приемопередающей аппаратуры за счет некоторого увеличения времени работы микропроцессорного блока, можно в значительной степени уменьшить энергопотребление каждого отдельного узла сенсорной сети.

Общее время работы T мота сенсорной сети состоит из времени работы приемо-передающей аппаратуры на прием t_r и на передачу t_t , а так же от времени t_p выполнения управляющей программы на микроконтроллере:

$$T = t_r + t_t + t_p. \quad (4)$$

Уменьшить время t_t работы передатчика можно за счет сокращения объемов передаваемых данных в сенсорной сети. Универсальным методом сокращения объемов передаваемых данных является сжатие информации. Использование сжатия внутри каждого из мотов сети позволяет не только снизить энергетические затраты данного мота, но и уменьшить потребление энергии цепочки всех последующих мотов, через которые будет передаваться информация, за счет сокращения времени t_t . Кроме того, сжатие информации позволяет уменьшить общее время T работы мота, повышая,

тем самым, быстродействие и оперативность получения данных из сенсорной сети.

При выборе определенного метода сжатия в сенсорной сети следует учитывать характеристики передаваемой по сети информации, а так же особенности построения и специфику работы самой сенсорной сети. В проведенном обзоре были рассмотрены принципы организации и передачи данных в сенсорной сети. Используя данные сведения, построим модель передачи данных в сенсорной сети. Данная модель представлена на рисунке 2.

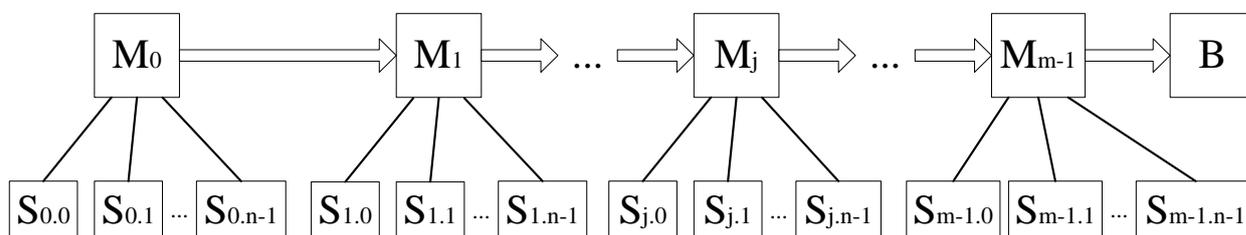


Рисунок 2 – Модель передачи данных в сенсорной сети

M_0 - является первым мотом в последовательности мотов M_j , через которые передается сообщение на базовую станцию B . Общее число мотов принимающих участие в передаче данных на базовую станцию равно m .

К каждому из мотов подключено ряд датчиков (сенсоров) S . Количество датчиков S , подключенных к моту M_j , будем считать величиной постоянной и равной n . Сигналы от сенсоров являются основным источником информации, передаваемой между узлами сенсорной сети. Наиболее распространенной и чаще всего используемой формой представления информации в современных устройствах является двоичная форма записи. Данные представляются в виде последовательностей нулей и единиц. В предлагаемой модели так же будем использовать двоичную форму представления информации. В качестве сенсоров для модели передачи данных в сенсорной сети были выбраны наиболее простые для описания и

моделирования бинарные датчики. Особенность таких датчиков состоит в том, что каждый из датчиков порождает всего один бит информации. Когда сенсор S_j находится в активном состоянии – он выдает двоичную единицу, а когда условие срабатывания датчика не выполняется – он выдает двоичный ноль. Используя множество бинарных датчиков можно выполнять моделирование более сложных источников информации (например аналого-цифровых преобразователей), преобразующих измеряемую величину в двоичный вид.

Объем информации, порождаемой каждым из мотов сенсорной сети, определяется количеством подключенных к данному моту сенсоров. Для разработанной модели он составляет n бит. К информации, порождаемой самим мотом, добавляется еще и информация от всех предыдущих мотов в цепочке передачи данных. В результате, общий объем V передаваемой через последовательность мотов информации будет равен:

$$V = \sum_{j=0}^{m-1} \sum_{i=0}^j n = \frac{n \cdot m \cdot (m+1)}{2} \text{ бит.} \quad (5)$$

Из приведенной оценки (5) видно, что сокращение объема передаваемой информации уже на первом моте M_0 в значительной степени уменьшает временные и энергетические затраты для всех последующих мотов в сенсорной сети.

Приведенная модель передачи информации в сенсорной сети показывает ряд специфических ограничений, накладываемых на передаваемую информацию:

1. Отсутствие потерь при получении, обработке и передаче информации от сенсоров. Данное ограничение обусловлено отсутствием у мотов необходимого количества ресурсов и информации для анализа и принятия решений об исключении ненужных данных. Анализ, обработка и преобразование информации, полученной от всех датчиков, выполняется только на базовой станции или же непосредственно на конечном устройстве с участием человека.

2. Отсутствие статистических и вероятностных характеристик принимаемой и передаваемой информации между мотами сенсорной сети.
3. Распределенный характер передаваемых данных. Каждый из мотов представляет собой отдельный и независимый от остальных мотов источник информации. Доступ ко всей информации в сенсорной сети можно получить только на базовой станции.
4. Односторонняя направленность передачи данных в сенсорной сети. Моту доступна информация только от предыдущих мотов в цепочке передаче данных.
5. Отсутствие обмена статистическими характеристиками информации между мотами сенсорной сети.

Наличие указанных ограничений в значительной степени снижает эффективность работы наиболее распространенных методов сжатия информации в сенсорной сети. С учетом ограничения на отсутствие потерь при обработке и передаче данных, а так же ввиду отсутствия определенного типа получаемой информации (например текст, графика, видео и т.д.), в обзоре были рассмотрены универсальные и наиболее часто используемые методы сжатия без потерь. К таким методам относятся: сжатие Хаффмана, LZ сжатие и арифметическое сжатие. Помимо специфических особенностей реализации указанных методов, существует еще две основных подхода при моделировании источника сообщений: статический и динамический. Статическая модель подразумевает наличие заранее известных статистических характеристик источника информации или же возможность предварительного просмотра всей информации с целью получения данных характеристик. Динамическая модель подразумевает постепенное накопление статистической информации параллельно с ее сжатием. При этом метод сжатия перестраивается в соответствии с вновь полученной информацией. Рассмотрим ограничения и недостатки использования в

сенсорных сетях указанных методов сжатия с учетом динамического и статического подходов при их реализации:

1. Реализация статического алгоритма Хаффмана, статического LZ сжатия и статического арифметического сжатия невозможна ввиду отсутствия необходимых значений вероятности появления сообщений для построения кодовых деревьев, создания словарей и кодовых таблиц при реализации указанных методов сжатия. Кроме того, наличие большого числа мотов в сенсорной сети приводит к возникновению большого числа источников информации. Поскольку источники не связаны между собой, то необходимо создавать кодовые структуры для каждого мота в отдельности, что приводит к значительному росту аппаратно-программных затрат и сложности реализации методов сжатия.
2. Использование динамического алгоритма Хаффмана, LZ сжатия и арифметического сжатия в сенсорных сетях является малоэффективным. Динамический подход к сжатию информации рассчитан на достаточно большие объемы передаваемой информации. Чем больше размер передаваемого сообщения – тем точнее будет построена модель источника и лучше выполнится сжатие. В сенсорных сетях источником информации являются датчики, подключенные к моту. Количество этих датчиков не велико, следовательно, и количество передаваемой информации является недостаточным для эффективного использования адаптивных методов сжатия. Сложности возникают и при использовании информации для сжатия от предыдущих мотов. В этом случае, для добавления к полученной информации в сжатом виде новых данных необходимо выполнить восстановление принятой информации с целью получения статистической характеристики источника. Процесс восстановления информации в каждом из последовательности мотов

при передаче сообщения приводит к значительному замедлению процесса передачи информации в сенсорной сети.

4.2 Метод биномиального сжатия информации в сенсорных сетях

Исходя из перечисленных выше недостатков применения стандартных методов сжатия информации в сенсорных сетях, возникает задача разработки метода сжатия распределенной информации, способного сжимать двоичные данные без использования статистических свойств передаваемой информации. При этом в алгоритме должна быть возможность добавлять в сжатый массив новую информацию без восстановления данного массива.

Для сжатия информации в сенсорных сетях с требуемыми характеристиками предлагается использовать метод биномиального нумерационного сжатия. Описание и основные характеристики указанного метода были приведены ранее. Метод биномиального нумерационного сжатия позволяет сжимать двоичные последовательности без знания статистических характеристик источника информации. Для применения метода биномиального сжатия информации в сенсорных сетях, необходимо его модифицировать. Процесс сжатия информации указанным методом необходимо разбить на независимые этапы. Новая информация будет добавляться к уже имеющейся, позволяя тем самым выполнять сжатие распределенной между мотами информации в сенсорной сети.

Нумерационная функция (1) биномиальной системы счисления для модели сенсорной сети будет иметь вид:

$$F_j = F_{j-1} + \sum_{i=0}^{n-1} (S_{j,i} \cdot C_{n-j+i}^{k-q_i}), \quad (6)$$

где F_j – номер двоичной последовательности, получаемый в j -ом моте M_j сенсорной сети;

$S_{j,i}$ - значение i -го бинарного датчика в j -ом моте M_j ;

n - количество бинарных датчиков в одном моте (одинаковое для всех мотов сети);

k - количество всех активных датчиков (число единиц) от 0-го до j -го мота включительно;

q_i – количество активных датчиков $S_{j,i}$ от $(n-1)$ -го разряда до $(i+1)$ -го включительно:

$$q_i = \sum_{l=i+1}^{n-1} S_{j,l}. \quad (7)$$

Как уже указывалось ранее, информацию получаемую от бинарных датчиков в каждом моте можно представить в виде n -разрядной двоичной последовательности. Активному сенсору будет соответствовать единичное значение в данной последовательности, а не активному – нулевое значение. В дальнейшем, для формализации, вместо количества активных датчиков будем использовать число двоичных единиц.

Сжатие в 0-м моте M_0 сенсорной сети выполняется с использованием обычного метода биномиального нумерационного сжатия, приведенного ранее. Сжатие информации во всех последующих мотах M_j , $j > 0$, выполняется с использованием метода биномиального сжатия в сенсорных сетях. Основные этапы выполнения биномиального нумерационного сжатия для j -го мота следующие:

1. [Ввод числа единиц] Считать из входного потока данных число k единиц от предыдущих мотов сенсорной сети, состоящее из $\lceil \log_2(j \cdot n + 1) \rceil$ двоичных разрядов;
2. [Ввод номера биномиального числа] Считать из входного потока данных текущий номер F_{j-1} биномиального числа, состоящий из $\lceil \log_2 C_{j-n}^k \rceil$ разрядов;
3. [Ввод двоичной последовательности] Считать информацию от датчиков в виде n -разрядной двоичной последовательности;

4. [Подсчет числа единиц] В двоичной последовательности подсчитать количество k_j единичных разрядов (после выполнения данной операции двоичную последовательность можно рассматривать как равновесную кодовую комбинацию длиной n с количеством k_j единиц);
5. [Получение нового числа единиц] К текущему значению k единиц, полученному от предыдущих мотов прибавить число k_j единиц, полученное на данном этапе ($k = k + k_j$). В результате величина k будет хранить число единиц от 0-го до j -го мота включительно;
6. [Вычисление номера биномиального числа] Используя полученное значение F_{j-1} номера от предыдущих мотов, а так же двоичную последовательность текущего мота вычислить новый номер биномиального числа в соответствии с нумерационной функцией (6) для сенсорной сети. Так же на данном этапе вычисляется количество разрядов, необходимых для хранения нового номера биномиального числа по формуле $\lceil \log_2 \left(C_{(j+1)n}^k \right) \rceil$;
7. [Вывод числа единиц] Вывести в выходной поток число k единиц, состоящее из $\lceil \log_2 \left((j+1) \cdot n + 1 \right) \rceil$ двоичных разрядов;
8. [Вывод номера биномиального числа] Вывести в выходной поток номер F_j биномиального числа, состоящий из $\lceil \log_2 \left(C_{(j+1)n}^k \right) \rceil$ двоичных разрядов.

Восстановление сжатой информации выполняется на базовой станции сенсорной сети. При этом распаковщику доступна информация о количестве мотов m , принимавших участие в передаче информации и информация о количестве n находящихся в каждом моте датчиков. Метод биномиального нумерационного восстановления для сенсорных сетей состоит из тех же

операций, что и обычный метод биномиального нумерационного восстановления информации, описанный ранее.

Размер L_j сжатого сообщения, полученного в j -м моте M_j , $0 \leq j \leq m-1$, сенсорной сети, можно рассчитать по формуле:

$$L_j = \left\lceil \log_2((j+1) \cdot n + 1) \right\rceil + \left\lceil \log_2 C_{(j+1) \cdot n}^{\Sigma k_j} \right\rceil, \quad (8)$$

где Σk_j - количество активных сенсоров (число единиц) во всех мотах от 0-го до j -го включительно:

$$\Sigma k_j = \sum_{t=0}^j k_t, \quad (9)$$

k_t - количество активных датчиков в t -м моте сенсорной сети.

Общий размер L передаваемых в сенсорной сети сообщений, сжатых по методу биномиального нумерационного сжатия будет равен сумме всех сообщений, порождаемых каждым из мотов при передаче информации от 0-го мота до базовой станции. Величина L рассчитывается по формуле:

$$L = \sum_j L_j = \sum_{j=0}^{m-1} \left\lceil \log_2((j+1) \cdot n + 1) \right\rceil + \sum_{j=0}^{m-1} \left\lceil \log_2 C_{(j+1) \cdot n}^{\Sigma k_j} \right\rceil. \quad (10)$$

Для оценки качества сжатия используют коэффициент сжатия K . Значение K равно отношению исходного объема информации к объему информации, полученному в результате сжатия. По величине коэффициента сжатия можно оценить во сколько раз уменьшились затраты времени на передачу сообщения в сенсорной сети, а, следовательно, и уменьшились энергетические затраты каждого из мотов сети. С учетом соотношений (5) и (10), получим

$$K = \frac{n \cdot m \cdot (m+1)}{2 \cdot \sum_{j=0}^{m-1} \left(\left\lceil \log_2((j+1) \cdot n + 1) \right\rceil + \left\lceil \log_2 C_{(j+1) \cdot n}^{\Sigma k_j} \right\rceil \right)}. \quad (11)$$

Рассмотрим пример работы метода биномиального сжатия в сенсорной сети. Пусть имеется часть сенсорной сети, состоящая из $m=4$ мотов. К

каждому моту подключено $n=8$ бинарных датчиков. Значения в виде двоичных n -разрядных последовательностей представлены в таблице 1.

Таблица 1 – Значения бинарных датчиков в мотах сенсорной сети

M ₃								M ₂							
S ₇	S ₆	S ₅	S ₄	S ₃	S ₂	S ₁	S ₀	S ₇	S ₆	S ₅	S ₄	S ₃	S ₂	S ₁	S ₀
0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0
M ₁								M ₀							
S ₇	S ₆	S ₅	S ₄	S ₃	S ₂	S ₁	S ₀	S ₇	S ₆	S ₅	S ₄	S ₃	S ₂	S ₁	S ₀
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0

Вначале выполняется биномиальное сжатие в 0-м моте:

1. Считать информацию от сенсоров в виде 8-разрядной двоичной последовательности: 00001000;
2. Подсчитать количество единиц: $k_0 = 1$;
3. Поскольку последний разряд двоичной последовательности равен 0, из последовательности убираются все младшие подряд стоящие нулевые разряды до появления первой 1. В результате будет сформировано биномиальное число 00001, состоящее из $r = 5$ разрядов;
4. По числовой функции (1) вычислить порядковый номер биномиального числа: $F_0 = 0 \cdot C_{8-5+4}^{1-0} + 0 \cdot C_{8-5+3}^{1-0} + 0 \cdot C_{8-5+2}^{1-0} + 0 \cdot C_{8-5+1}^{1-0} + 1 \cdot C_{8-5+0}^{1-0} = 3_{\langle 10 \rangle}$.

Количество двоичных разрядов, необходимых для хранения порядкового номера F_0 : $\lceil \log_2 C_8^1 \rceil = \lceil \log_2(8) \rceil = 3$;

5. Сжатое в моте M_0 сообщение будет иметь вид: 0001 011_{<2>}. Размер сжатого сообщения равен: $L_0 = 7$ бит.

Полученное сообщение передается в мот M_1 . Далее выполняется добавление сжатой информации к уже имеющейся в соответствии с методом биномиального сжатия в сенсорных сетях и кодирующей функцией (6). Поскольку все значения сенсоров равны нулю, значения номера и числа единиц не изменятся. Увеличится только число разрядов для хранения указанных параметров. В результате, от мота M_1 в мот M_2 будет передано

сообщение 00001 0011_{<2>}, состоящее из $L_1 = \lceil \log_2(2 \cdot 8 + 1) \rceil + \lceil \log_2 C_{2,8}^{1+0} \rceil = 9$ двоичных разрядов.

Результаты работы метода биномиального сжатия в сенсорных сетях для мота M_2 следующие:

1. Получить из входного потока сообщение 00001 0011_{<2>} от предыдущего мота;
2. Считать информацию от сенсоров в виде 8-разрядной двоичной последовательности: 00010100;
3. Подсчитать количество единиц: $k_2 = 2$;
4. Прибавить к принятому числу $k = 1$ единиц полученное на предыдущем этапе число: $k = 1 + 2 = 3$;
5. По функции (6) вычислить новый порядковый номер биномиального числа: $F_2 = F_1 + 1 \cdot C_{8,2+4}^{3-0} + 1 \cdot C_{8,2+2}^{3-1} = 3 + 1140 + 153 = 1296_{<10>}$. Количество двоичных разрядов, необходимых для хранения порядкового номера F_2 : $\lceil \log_2 C_{8,(2+1)}^3 \rceil = \lceil \log_2(2024) \rceil = 11$;
6. Сжатое в моте M_2 сообщение будет иметь вид: 00011 1010010000_{<2>}.
Размер сжатого сообщения равен: $L_2 = 5 + 11 = 16$ бит.

Аналогичным образом выполняется сжатие в последнем моте M_3 . В результате сжатия на базовую станцию будет передана двоичная последовательность 000100 0110000111010111_{<2>}, состоящая из $L_3 = 6 + 16 = 22$ двоичных разрядов.

Общее число переданных по сети двоичных разрядов равно сумме двоичных разрядов переданных каждым из мотов сенсорной сети:

$$L = \sum_j L_j = 7 + 9 + 16 + 22 = 54.$$

Коэффициент сжатия информации рассчитываем по формуле (11):

$$K = \frac{n \cdot m \cdot (m + 1)}{2 \cdot L} = \frac{80}{54} \approx 1,48.$$

ВЫВОДЫ

В данной работе был рассмотрен метод биномиального сжатия информации в сенсорных сетях, основанный на принципах нумерационного кодирования. В качестве нумерационной функции была выбрана числовая функция двоичной неравномерной биномиальной системы счисления. Существующие на сегодняшний момент методы сжатия данных в большинстве случаев имеют достаточно сложные алгоритмы, реализация которых требует значительных вычислительных ресурсов и существенных аппаратных затрат при схемотехнической реализации. Кроме того, специфические особенности построения и функционирования сенсорных сетей в значительной степени снижают эффективность использования наиболее распространенных методов сжатия.

Рассмотренный метод биномиального сжатия информации обладает рядом достоинств по сравнению с существующими методами сжатия при использовании их в сенсорных сетях:

1. Предложенный метод сжатия не требует предварительной оценки сжимаемых сообщений с целью набора статистики. Указанное положительное свойство приобретает особую важность в случае сжатия сообщений порожденных мотами сенсорной сети, когда получение статистических данных затруднено или не представляется возможным.
2. Биномиальный нумерационный код имеет равномерную длину и, следовательно, не требует выравнивания скоростей передачи и приема при передаче по каналу связи. Кроме того, случайное ошибочное изменение разряда в биномиальном номере не является критичным для приема всего потока данных.
3. Биномиальное нумерационное кодирование, которое лежит в основе предложенного метода сжатия информации, преобразует двоичные последовательности в числа, что позволяет производить обработку исходных данных (фильтрацию, поиск и т.д.) в сжатом виде.

ЛИТЕРАТУРА

1. Akyildiz I.F., Su W., Sankarasubramaniam Y., Cayirci E. Wireless sensor networks: a survey // *Computer Networks, Volume 38*, 2002. – pp. 393–422.
2. Karl H, Willig A. *Protocols and Architecture for Wireless Sensor Networks*, ISBN 0-470-09510-5. John Wiley & Sons: Chichester, 2005, pp. 60–62.
3. Рагозин Д.В. Моделирование синхронизированных сенсорных сетей // *Проблеми програмування, № 2-3. Спеціальний випуск, 2008.* – 721-729 с.
4. Яцків В.В., Яцків Н.Г., Боднар Д.І. Завадозахищений метод представлення даних в бездротових сенсорних мережах // *Науковий вісник Чернівецького університету, Випуск 446, 2009.* С. 117-120.
5. Ватолин Д., Ратушняк А., Смирнов М., Юкин В. *Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео.* – М.: ДИАЛОГ-МИФИ, 2003. – 384 с.
6. Фомин А.А. *Основы сжатия информации.* - Санкт-Петербург: СПбГПУ, 1998г.
7. Балашов К.Ю. *Сжатие информации: анализ методов и подходов.* – Минск, 2000. – 42 с.
8. Сэломон Д. *Сжатие данных, изображений и звука.* Москва: Техносфера, 2004. – 368 с.
9. Борисенко А.А. *Биномиальный счет. Теория и практика: Монография.* – Сумы: ИДТ "Университетская книга", 2004. – 170 с.
10. Борисенко А. А., Кулик И. А. *Биномиальное кодирование: монография* – Сумы: СумГУ, 2010. – 206 с.
11. Кулик И.А., Костель С.В. Быстродействующий метод биномиального нумерационного кодирования // *"Автоматизированные системы управления и приборы автоматики"*, Выпуск 149, - Харьков: ХНУРЕ, 2009. – с. 66-76

12. Кулик И.А., Костель С.В., Скордина Е.М. Использование биномиальных чисел для сжатия бинарных изображений // Вісник Сумського державного університету. Технічні науки. – 2009. – № 2. – С. 29-36.
13. Кулик И.А., Костель С.В. Система сжатия информации на основе биномиальных чисел // Международная научно-практическая конференция "Информационные технологии и компьютерная инженерия" – Винница: ВНТУ, 2010. – с. 100-101.