

**О. Г. ВАСИЛЬЧЕНКОВ**, к.т.н., доцент НТУ “ХПИ”

**А. А. ЗУЕВ**, к.т.н., доцент НТУ “ХПИ”;

**Р. А. ДАВЫДОВ**, магистр НТУ “ХПИ”;

## **КОДИРОВАНИЕ И ОПТИМИЗАЦИЯ ПРОЦЕССА ПЕРЕДАЧИ ДАННЫХ ЧЕРЕЗ ЛОКАЛЬНЫЕ И ГЛОБАЛЬНЫЕ СЕТИ**

В статье рассматриваются вопросы оптимального кодирования данных при передаче по локальным и глобальным сетям. Описан алгоритм работы и структурная схема программно-аппаратной системы, которая осуществляет сбор и передачу изображений по запросу пользователя. Приводится пример использования данной системы для обработки биомедицинских данных.

**Ключевые слова:** локальная сеть, глобальная сеть, сервер, клиент, алгоритм, избыточность, ПЭВМ.

**Введение.** В настоящее время информационные технологии (ИТ) получили широкое распространение практически во всех областях человеческой деятельности.

ИТ с использованием автономной ПЭВМ позволяют значительно расширить возможности пользователя, еще более значительный эффект можно получить при объединении отдельных ПЭВМ в локальную компьютерную сеть, которая обеспечивает функционирование нескольких ПЭВМ как единую систему. Вычислительные сети объединяют различные службы, ускоряют документооборот, позволяют хранить необходимую информацию и предоставляют доступ к ней по запросу. Естественным приложением ИТ являются компьютерные телекоммуникации и глобальные сети, обеспечивающие доступ пользователей к глобальным информационным ресурсам и выход в мировое информационное пространство.

Количество информации которую необходимо передавать и обрабатывать растет от года к году. Одним из возможных решений, которое позволяет компенсировать такой рост, является кодирование данных, которое позволяет сократить требования к объему устройств хранения данных и главное, к пропускной способности каналов связи. Условиями его применимости является избыточность информации и возможность установки специального программного обеспечения, либо аппаратуры, как вблизи источника, так и вблизи приемника информации. Оба эти условия легко выполнимы.

**Цель работы.** Показать особенности использования алгоритмов сжатия и передачи биомедицинских данных в распределенной программно-аппаратной системе, через локальные и глобальные сети.

**Анализ литературы.** Вычислительные сети (ВС) обычно используют специально проложенные высококачественные и хорошо защищенные от помех линии связи [1]. При этом среднескоростной считается сеть, имеющая пропускную способность 100 Мбит/с. При помощи ВС можно реализовать распределенную обработку и хранение данных. Обработка данных в этом случае распределена между несколькими объектами: сервером и клиентами или несколькими клиентами [1]. Клиенты формируют запросы на сервер (или другой клиент) для выполнения сложных процедур, чтения файлов, поиска информации в базе данных и т.д. Запрос выполняется, результаты выполнения передаются запрашивавшему их клиенту.

При организации обмена данными между двумя клиентами, в ряде сетей в которых клиенты подключены с использованием технологии NAT (преобразование сетевых адресов) различных типов, возникает задача определения адреса нужного компьютера в сети, для ее решения существует ряд методов: STUN (TURN), UDP hole punching, IPv6 Teredo и т.д.[2].

Большинство данных, которые передаются по ВС, являются избыточными, при этом степень избыточности во многом зависит от типа данных. Например, для видеоданных степень избыточности в несколько раз больше чем для графических данных, чья степень избыточности, в свою очередь, больше чем степень избыточности текстовых данных. Другим фактором, влияющим на степень избыточности, является принятая система кодирования. Примером систем кодирования могут быть обычные языки общения, которые являются ни чем другим, как системами кодирования понятий и идей для высказывания мыслей. Установлено, что "кодирование" текстовых данных с помощью русского языка дает в среднем избыточность на 20-25% большую, чем "кодирование" аналогичных данных на английском языке[3].

Для человека избыточность данных часто связана с качеством информации, поскольку избыточность улучшает понятность и восприятие информации. Но, когда речь идет о хранении и передаче информации средствами компьютерной техники - избыточность играет отрицательную роль, поскольку она приводит к возрастанию стоимости хранения и передачи информации. Особенно актуальной эта проблема является в случае обработки значительных объемов информации (десятки и сотни гигабайт) при незначительных объемах носителей данных. В связи с этим, задача уменьшения избыточности или кодирования данных является актуальной. Для ее решения применяются алгоритмы сжатия, как классические, например алгоритм LZW (Лемпела-Зива-Велча), так и разработанные относительно недавно, например, сжатие на основе преобразования BWT (Барроуза – Уиллера)[4].

**Реализация.** Систему хранения и передачи биомедицинских данных целесообразно разрабатывать, исходя из следующих принципов:

- возможность свободного расширения вычислительной сети, как за счет замены оборудования на более производительное, так и при помощи добавления новых вычислительных узлов и переконфигурирования сети;
- эффективный, быстродействующий механизм управления обменом по сети;
- возможность добавления и расширения ПО используемого для обработки данных.

Для построения ВС, целесообразно использовать стандартное сетевое оборудование и компьютеры общего назначения, а предоставлять возможность обмена данными с персональных и мобильных устройств. Это позволит легко конфигурировать сеть, в соответствии с требованиями заказчика, заменять вышедшие из строя узлы, а также модернизировать сеть. При этом использование стандартных компонентов позволит сохранить стоимость аппаратной части системы и затраты на ее обслуживание на достаточно низком уровне. На рисунке 1 приведена структурная схема такой системы.

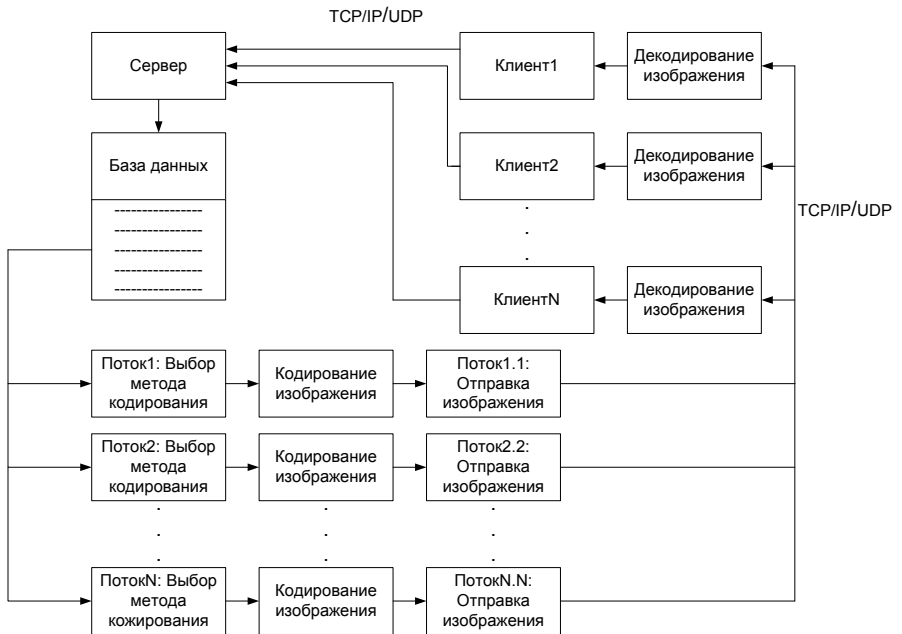


Рис. 1 – Структурная схема системы хранения данных

Основная сложность в построении системы заключается в разработке специализированного ПО для кодирования данных. Все методы кодирования данных основаны на простом принципе - если наиболее часто встречающиеся элементы закодировать более короткими кодами, а реже встречающиеся – более длинными, то для хранения всех данных потребуется меньше места,

чем, если бы все элементы представлялись кодами одинаковой длины. Точная взаимосвязь между частотами появления элементов, и оптимальными длинами кодов описана в так называемой теореме Шеннона об источнике шифрования, которая определяет предел максимального сжатия без потерь и энтропию [4].

Если вероятность появления элемента  $s_i$  равна  $p(s_i)$ , то наиболее выгодно представлять этот элемент кодом длиной в  $\log_2 p(s_i)$  бит. Если при кодировании добиться того, что длина всех элементов будет приведена к  $\log_2 p(s_i)$  битам, то и длина всей кодируемой последовательности будет минимальной из возможных. При этом, если распределение вероятностей всех элементов  $F = \{p(s_i)\}$  неизменно, и вероятности элементов взаимно независимы, то средняя длина кодов может быть рассчитана по формуле (1):

$$H = -\sum_i p(s_i) * \log_2 p(s_i) \quad (1)$$

Однако вероятность появления элемента не может быть независимой, в этом случае, для каждого нового кодируемого элемента  $s_i$  распределение вероятностей  $F$  примет некоторое значение  $F_k$ , то есть для каждого элемента  $F = F_k$  и  $H = H_k$ .

Можно сказать, что источник находится в состоянии  $k$ , которому соответствует некий набор вероятностей  $p_k(s_i)$  для всех элементов  $s_i$ .

Поэтому, учитывая эту поправку, можно выразить среднюю длину кодов по формуле (2)

$$H = -\sum_k P_k * H_k = -\sum_{k,i} P_k * p_k(s_i) \log_2 p_k(s_i) \quad (2)$$

где  $P_k$  — вероятность нахождения источника в состоянии  $k$  [4].

Рассмотрим применение LZW алгоритма для кодирования данных. Этот алгоритм использует расширяющийся словарь, в котором первоначально содержится 256 слов длиной в одну букву. В процессе работы словарь увеличивается до своего максимального объема  $|V_{\max}|$  слов. Обычно, объем словаря достигает нескольких десятков тысяч слов. Каждое слово имеет свою известную длину. Таким образом, количество слов в словаре точно равно его текущему объему. В процессе работы словарь заполняется по алгоритму приведенному на рисунке 2:

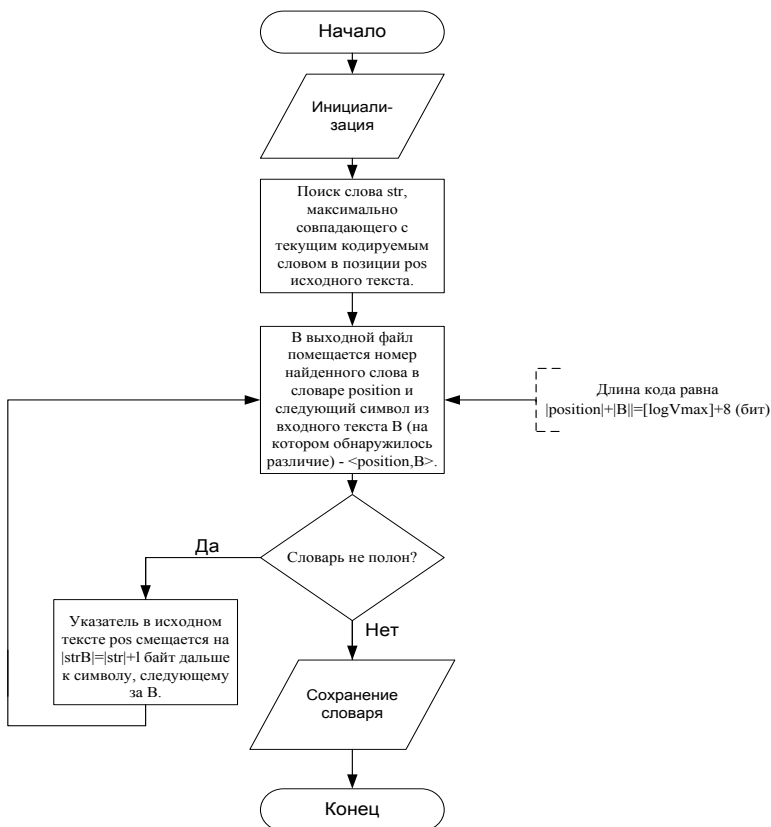


Рис. 2 – Алгоритм заполнения словаря LZW метода

Так как словарь увеличивается постепенно и одинаково для кодировщика и декодировщика, для кодирования позиции нет необходимости использовать  $\lceil \log V_{\max} \rceil$  бит, а можно брать лишь  $\lceil \log V \rceil$  бит, где  $V$ -текущий объем словаря [4].

Альтернативой алгоритму LZW являются – алгоритмы на основе BWT преобразования. Основная задача этого преобразования заключается в том, чтобы переставить символы в последовательности таким образом, чтобы увеличить коэффициент сжатия для методов кодирования.

Данные состоят из набора стабильных сочетаний символов (контекстов). Контекст позволяет предсказать некоторый неизвестный символ, а преобразование собирает вместе символы, соответствующие сходным контекстам. таким образом, чем больше стабильных контекстов в блоке данных, тем лучше будет сжиматься полученный в результате преобразования блок. На практике, в результате преобразования текстов более половины из всех символов следует за такими же [4].

Рассмотрим характерные особенности алгоритмов сжатия (компрессии) данных. Скорость компрессии для BWT - на уровне LZW, но при этом у LZW скорость декомпрессии, как правило, в 3-4 раза быстрее. Степень компрессии сильно зависит от типа данных. BWT оперирует сразу целым блоком данных. Это делает затруднительным использование алгоритма в тех областях применения, где требуется быстрое сжатие данных, символ за символом. В этом отношении BWT более требователен, чем методы семейства LZW, использующие для сжатия скользящее окно.

**Выводы.** Представление и хранение биомедицинских данных требует больших объёмов данных и предъявляет высокие требования к сетевому оборудованию при их передаче по каналам связи и к ёмкости внешних носителей при хранении. Выбор наиболее оптимального метода кодирования и нахождение путей максимально эффективной передачи данных по каналам связи является важной задачей.

**Список литературы: 1.** Персональные компьютеры в сетях TCP/IP – Киев, издательская группа BHV, 1997. **2** Rosenberg, J., Weinberger, J., Huitema, C., Mahy, R., “STUN - Simple Traversal of User Datagram Protocol (UDP) Through Network Address Translators (NATs)”, IETF, RFC3489, March 2003. **3.** Сэлмон Д. Сжатие данных, изображений и звука. Data Compression Methods. Серия: Мир программирования. Издательство: Техносфера, 2004. - 368 с. **4.** Ватолин Д., Ратушняк А., Смирнов М., Юкин В. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео. - М.: ДИАЛОГ-МИФИ, 2002. - 384 с.

**Bibliography (transliterated): 1.** Personal'nye komp'jutery v setjah TCP/IP – Kiev, izdatel'skaja gruppa BHV, 1997. **2** Rosenberg, J., Weinberger, J., Huitema, C., Mahy, R., “STUN - Simple Traversal of User Datagram Protocol (UDP) Through Network Address Translators (NATs)”, IETF, RFC3489, March 2003. **3.** Sjelomon D. Szhatie dannyh, izobrazhenij i zvuka. Data Compression Methods. Serija: Mir programmirovanija. Izdatel'stvo: Tehnosfera, 2004. - 368 s. **4.** Vatolin D., Ratushnjak A., Smirnov M., Jukin V. Metody szhattija dannyh. Ustrojstvo arhivatorov, szhatie izobrazhenij i video. - M.: DIALOG-MIFI, 2002. - 384 s.

*Поступила (received) 13.12.2013*

**А. И. ГАПОН**, канд. техн. наук, проф. НТУ “ХПИ”  
**О. Н. ЕВСЕЕНКО**, магистр НТУ “ХПИ”;

## **ПОСТРОЕНИЕ МОДЕЛИ ТЕПЛООВОГО ОБЪЕКТА НА РС-СЕТКАХ**

Производится выбор метода моделирования. Осуществляется построение математической модели теплового объекта на основе метода РС-сеток. Снимаются переходные характеристики объекта. Исследуется реакция объекта на серию ШИМ-импульсов.

**Ключевые слова:** модель, тепловой объект, метод РС-сеток, масштабные коэффициенты, переходная характеристика, ШИМ.

**Введение.** Моделирование — важный процесс при изучении объекта, который позволяет упрощенно описать сложную систему, отразив её существенные свойства. Это позволяет использовать модель в качестве заместителя изучаемой системы.

Математическое моделирование основано на идентичности дифференциальных уравнений, описывающих явление в оригинале, и модели, отличающихся по своей природе.

Основные преимущества математического моделирования в сравнении с физическим: при математическом моделировании более наглядно осуществляется индикация и регистрация результатов исследований, можно в широких пределах варьировать исходные данные задачи для выбора оптимальных (по заданному критерию) параметров исследуемой системы, время решения задачи может быть изменено в широких пределах [1].

Реализация управления объектом должна основываться на дискретных системах — в них проще реализовать сложные алгоритмы управления. При использовании ЦВМ алгоритм задается в виде программы, сложность которой практически не влияет на конструкцию системы. Смена программы управления производится без больших затрат времени. В непрерывных САУ повышение сложности алгоритма управления требует включения в состав системы новых элементов, а замена алгоритма связана с существенным усложнением конструкции [2].

**Цель работы.** Построение математической модели реального теплового объекта и исследование его поведения на РС-сетках на импульсные воздействия, параметры которых вычисляются по методу, предложенному в работе [3].

**Анализ литературы.** Из многообразия методов математического моделирования [4] был сделан выбор в пользу метода РС-сеток. Суть метода моделирования на РС-сетках заключается в разделении моделируемой