

# Программные методы контроля достоверности информации в структуре пакетов передачи данных систем электронного документооборота

А.Р. Ахатов

Поставлена задача повышения достоверности передачи информации в системах электронного документооборота на основе программных методов. Разработаны принципы, правила, алгоритмы контроля достоверности информации на основе линейного посимвольного и поразрядного суммирования, плоскостного контроля и контроля по произвольному модулю. Исследованы эффективность и предельные возможности программных методов по критериям вероятности пропуска ошибок, трудоёмкости и стоимости контроля. Получены верхние границы ошибок для методов контроля информации по модулю  $q$ .

*Ключевые слова:* электронный документооборот, контроль достоверности.

## 1. Введение

В системах электронного документооборота (СЭД) используются технологии, позволяющие извлекать нужную информацию, проводить интеллектуальный анализ данных, создавать средства поиска закономерностей, извлекать знания, анализировать базы данных. СЭД производит обработку данных в любых форматах, представленных в виде офисных и отсканированных бумажных документов, web-страниц, графических изображений, чертежей, видеофайлов и т.д. Критериями качества работы СЭД являются обеспечение целостности, сохранности и достоверности информационных ресурсов. Предъявление таких требований, особенно достоверности, связано с тем, что СЭД функционируют в локальных, региональных и глобальных инфокоммуникационных сетях, а в передаваемой информации из-за сбоев и отказов в работе технических средств, различного рода помех в каналах связи, ошибок операторов информация может стать недостаточно достоверной. Источниками ошибок также являются процессы записи в память ЭВМ, сканирование, перезапись информации с одного вида носителей информации на другой. Искажения в информации не только снижают точность решения задач, но и влекут за собой ошибочный результат, а следовательно, неверное решение, причём этап ввода информации считается самой ненадёжной частью информационного процесса. Специалисты подтверждают, что наибольшее значение вероятности искажений принадлежит средствам сканирования и распознавания ( $\approx 10^{-2}$ ) и человеку-оператору ( $\approx 10^{-3}$ ) [1]. В связи с этим, разработка методов контроля, направленных на обнаружение и исправление ошибок, в первую очередь, на этапах ввода и подготовки информации при меньших материальных и временных затратах, остаётся до сих пор весьма актуальной и нерешённой задачей.

В современных телекоммуникационных сетях для передачи информации используется множество различных протоколов. К 1979 году окончательно сформировался стек базовых

протоколов и начался экстенсивный рост сети Интернет. По мере выявления недостатков протоколов и новых потребностей после 1989 года началась активная разработка новых направлений и приложений в Интернет. Это, прежде всего, мультимедиа, базой для внедрения которой стали протоколы MIME и HTTP, системы сетевой и информационной безопасности, мобильные услуги, IPv6, MPLS/GMPLS и различные виды сервисов.

Персональные компьютеры (ПК), которые служат основными средствами подготовки, ввода и передачи данных СЭД, могут подключаться к региональным, государственным или глобальным телекоммуникациям. Так, например, подключаясь к сети Интернет, ПК используют стек протоколов TCP/IP.

Любой пакет, используемый для передачи информации в глобальных сетях, можно представить в следующем виде:

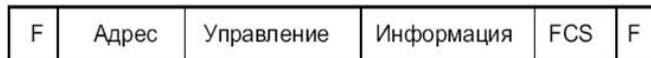


Рис. 1. Структура пакета

Здесь флаг F задаёт границы кадра, FCS — контрольная сумма. Поле «управление» предназначено для организации и осуществления процесса передачи. Оно содержит поля номера кадров, номера текущего кадра, номер следующего кадра, который отправитель текущего кадра ожидает получить, и другую служебную информацию. Все указанные поля выполняют функцию заголовка пакета.

Поле «информация» содержит непосредственно передаваемое сообщение. Данные в этом поле могут отображать информацию о различных объектах СЭД и в этом случае являются предметом исследования следующего рода: дискретно-цифровая информация, т.е. данные тематических, картографических, географических баз данных, фактографические данные о технико-экономических, финансовых показателях; алфавитно-цифровая информация; текстовая информация; графическое представление документов; непрерывные по природе данные, т.е. результаты метрологических измерений, данные физических, химических, биологических экспериментов и др.

Следует отметить, что в существующих протоколах основное усилие по достоверной передаче информации направлено на построение технологий защиты от искажений в каналах связи и не предприняты меры обнаружения и устранения ошибок ввода, также искажающих смысл и содержание информации. Причём можно отметить следующий немаловажный момент, требующий внимания для исследования способов повышения достоверности передачи информации на основе пакетов. Например, в IP-пакетах значительный объём ресурса расходуется на передачу заголовков, вместе с тем большая часть информации в IP-заголовках остаётся неизменной при передаче информации от пакета к пакету на протяжении всего потока.

Отметим, что обусловленная самой технологией передачи пакетов, избыточная информация непосредственно влияет на скорость передачи информации и при этом эффективность канала немало зависит от применяемых методов контроля достоверности информации наряду с методами сжатия и кодирования.

В настоящее время существует множество вариантов контроля достоверности данных на основе избыточности, используемых в сжатых заголовках протоколов передачи информации [2, 3]. Среди них характерной является практика использования механизма обнаружения ошибок TWICE.

Суть такого метода заключается в том, что компрессор данных включает в заголовок счётчик числа пакетов, увеличивающийся при каждой передаче, а декомпрессор данных контролирует последовательность счётчика и при обнаружении скачка более чем на 1 принимает решение о потере пакета. Затем декомпрессор обновляет переданный контекст. Для восстановления контекста при этом используется обратный канал, по которому декомпрессор посылает запросы на обновление заголовков.

Надо отметить, что механизм TWICE повышает достоверность переданной информации, главным образом, за счёт выявления потери пакетов. Однако вопрос обнаружения самой ошибки остаётся открытым.

В качестве метода, обнаруживающего ошибку, можно показать схему ROCCO, основанную на использовании контрольной суммы. Вычисляемая для исходного несжатого заголовка контрольная сумма включается в сжатый заголовок и позволяет определить, во-первых, корректность контекста, а во-вторых, правильность его локального восстановления.

Из изложенного выше вытекают две проблемы, связанные с повышением достоверности при передаче информации в современных телекоммуникационных системах. Первая проблема – борьба с ошибками на Прикладном и Представительском уровнях модели OSI, т.е. борьба с ошибками человека-оператора, сканирующих и других устройств, предназначенных для ввода информации. Вторая проблема – рассмотрение решения задач контроля информации на Транспортном, Сетевом, Физическом и Канальном уровнях. Для решения каждой из отмеченных проблем требуется проведение специальных исследований и разработок, связанных с созданием программных средств контроля достоверности информации в структуре пакетов передачи данных.

В [4] авторами предложен новый подход использования естественной избыточности информации путём разработки алгоритмов и программных средств контроля информации, основанных на принципах статистического и словарного кодирования, таких как коды Хаффмана, Лемпеля – Зива – Велча, преобразования Барроуза – Вилера, арифметического кодирования, которые охватывают большую часть ошибок человека-оператора и сканирующих устройств, совершаемых на уровне пользователей услуг сети прикладного уровня (Прикладной, Представительский, Сеансовый). Разработанные алгоритмы повышения достоверности информации устраняют недостатки, имеющиеся в существующих системах контроля информации и обеспечивают приемлемую достоверность информации в пакетах за счёт программных средств, использующих статистическую избыточность, которая является естественной, содержится непосредственно в информации.

Целью настоящей работы являются исследование и разработка методов, обеспечивающих повышение достоверности информации, в которых также учитываются ошибки на уровне сетевых служб многоуровневой модели OSI (Сетевой, Канальный, Физический). При передаче пакетов поля заголовков с изменяющимся содержанием и поля с неизменным содержанием можно представить в качестве различных типов объектов. Для первого типа объектов, представляющих переменную часть пакета, например, контрольные суммы, порядковый номер кадра в пакете, тип управляющего кадра и др., возможна разработка методов программного контроля информации, основанного на использовании искусственной и естественной избыточности [5]. Для второго типа объектов, каковыми являются неизменные части пакетов, например, адреса отправителя и получателя, идентификаторы начала и конца кадра и др., целесообразна реализация программных способов контроля и коррекции ошибок на основе экспертных систем и баз данных, использующих вложенную структурно-технологическую избыточность [6].

Рассмотрим круг вопросов, связанных с исследованием и разработкой программных методов контроля достоверности информации переменной части пакетов, содержащую непосредственно передаваемое получателю сообщение. Причём следует отметить, что эффективный программный контроль достоверности информации улучшает эффективность сети, качество и скорость передачи информации за счёт автоматизации процессов обнаружения и исправления ошибок и уменьшения времени реакции при диалоге систем.

## 2. Программные методы контроля информации

Известны системные (организационные), кодовые, аппаратурные и программные методы повышения достоверности информации.

Системные методы, в основном, включают организационные мероприятия, направленные на повышение достоверности информации, такие как поддержание характеристик оборудования в заданных пределах; повышение культуры обработки информации; обучение и стимулирование персонала; создание оптимального числа копий или предысторий программ исходных и текущих данных; организация первичной обработки данных и создание специальных процедур доступа к массивам и др.

В теории передачи информации широко развиты кодовые и аппаратные методы, обеспечивающие помехоустойчивую передачу. В системах, защищённых только кодовыми и аппаратными методами контроля передачи данных, вероятность пропуска ошибок оказывается менее  $10^{-5}$ , однако в них не учитываются ошибки человека-оператора и устройств ввода информации, которые составляют большую долю – до 85 % в общем объёме искажений.

Следовательно, известные методы контроля информации в современных пакетах могут быть удачно дополнены программными методами, использующими избыточность для контроля достоверности информации и способными устранить существующие недостатки.

Вопросы обеспечения достоверности передачи цифровых данных, точности выходной информации, автоматизированного ведения информационной базы с управляемой искусственной избыточностью были исследованы в работах Куцыка Б.С., Вагина Н.И., Горшкова Л.Ф., Пополянско А.Н., Коршиковой Л.А. Мамиконова А.Г., Кульба В.В., Шелкова А.Б., Липаева В.В., Зильбершатца А., Стоунбрейкера М., Ульмана Д., Зингера Е.З. на практике автоматизированных систем управления предприятиями на базе локальных вычислительных сетей в 80-х годах XX века.

Программные методы повышения достоверности передачи цифровой информации на основе использования естественной информационной избыточности при контроле параметров технологического процесса, результатов физических и химических экспериментов исследованы и разработаны в работах Крамаренко В.В., Брагина В.Н., Жуманова И.И.

Как было отмечено, в современных информационных системах предприятий и организаций, наряду с цифровой информацией, обрабатывается большое количество служебных документов с текстовой информацией, обладающей структурно-технологической, статистической и семантической избыточностью. Причём использование вложенных в тексты логических и статистических связей символов и букв, семантики и свойств естественного языка как раз и создаёт благоприятные условия для обеспечения качества передачи, ввода и обработки текстовой информации, в том числе контроля и коррекции орфографических ошибок [7, 8].

Однако в работах, посвящённых использованию программных методов повышения достоверности информации, теоретические и практические стороны проблемы контроля ошибок в текстах изучены недостаточно. Отсутствуют приемлемые методы, алгоритмы и правила контроля ошибок человека-оператора, сканирования и распознавания и орфографических ошибок, особенно в текстах на узбекском языке.

Следовательно, актуально проведение исследований основных концепций повышения достоверности текстовой информации в структурах современных протоколов передачи данных и разработки алгоритмов, функциональных и прикладных модулей, программных средств для обнаружения и исправления ошибок человека-оператора на этапе подготовки информации на основе программных методов и теоретическое обоснования их эффективности по критерию минимальной вероятности необнаруженных ошибок.

Вместе с тем, уместно отметить, что увеличение возможностей компьютерных технологий и внедрение современных телекоммуникационных систем, в свою очередь, способствует внедрению программных методов как для территориальных, так и для глобальных и мобильных сетей передачи данных.

Основным достоинством программных методов является то, что они могут быть легко программно реализованы на современных ПК. Известны следующие методы, использующие

искусственную избыточность данных:

- счётный контроль с получением контрольных сумм;
- методы посимвольного и поразрядного суммирования по модулю 9;
- цифровой контроль;
- метод контрольных чисел;
- счётный контроль с получением контрольных итогов;
- методы сравнения различных источников данных или метод контроля шифров по справочнику;
- балансовые методы контроля;
- методы избыточных цифр;
- контроль по модулям 10 и 11.

Представляет интерес отражение особенностей каждого из названных методов контроля информации.

Счётный контроль с получением контрольных сумм. Вследствие простоты своей реализации – это самый распространённый метод контроля. При этом широко применяются методы посимвольного и поразрядного суммирования. Среди них эффективными являются метод суммы информационных знаков, поразрядное суммирование, суммирование по модулю 9, цифровой счётный контроль и метод контрольных чисел. Установлено, что в благоприятных условиях обнаруживающая способность этих методов на порядок выше посимвольного и поразрядного суммирования.

Счётные методы контроля с получением контрольных итогов по эффективности обнаружения ошибок не уступают методам контрольных сумм. В данном случае контролируемые массивы вводятся в память ЭВМ два раза с сопоставлением контрольных сумм, что препятствовало их широкому применению.

Среди методов, основанных на сравнении различных источников данных, следует отметить метод контроля по справочнику, которые были эффективны лишь при небольшом числе номеров в справочнике. В нередких случаях практики управления нашли применение балансовые методы контроля. Их обнаруживающая способность увеличивается за счёт использования статистических и логических связей, существующих между контролируруемыми показателями. В случае отсутствия этих факторов данный метод не применяется.

Методы контроля избыточных цифр считаются наиболее эффективными. Их эффективность на один – два порядка выше других методов. Однако при эксплуатации они используют значительные ресурсы для вычисления контрольных сумм на основе вложенной в память логики и, кроме того, процесс контроля связан с увеличением затрат машинного времени.

Более распространёнными являются методы контроля по модулям 10 и 11. Известны работы с предложениями применить модифицированные варианты модульных методов, которые улучшают показатель достоверности данных и снижают трудоёмкость их эксплуатации.

Однако отмеченным программным методам повышения достоверности информации присущи недостатки, такие как:

- пропускаются ошибки, обусловленные перестановкой символов и цифр, а для устранения подобного рода ошибок нужно вычислить не просто арифметическую сумму чисел или цифр по принципу отмеченного алгоритма, а сумму, преобразованную определёнными способами;
- контролируемые массивы данных вводятся в память ЭВМ многократно для обнаружения и исправления ошибок, что приводит к значительным расходам полезного машинного времени.

Для устранения имеющихся пробелов рассматриваемой теории требуется проведение дополнительных исследований с целью оценки эффективности каждого программного метода контроля, что является трудоёмкой, но нужной работой.

В [9] авторами исследована и оценена эффективность следующих программных методов контроля:

- использующих посимвольное и поразрядное суммирование (соответственно алгоритмы П1 и П2);
- суммы информационных знаков (алгоритм П3);
- поразрядное суммирование по модулю 9 (алгоритм П4);
- цифровой контроль (алгоритм П5);
- контрольные числа (алгоритм П6);
- двойной просчёт (алгоритм П7);
- по справочнику (алгоритм П8);
- по модулям 10 и 11 (алгоритмы П9, П10).

В п.5 настоящей работы даны некоторые из результатов аналитических и экспериментальных исследований.

Следует отметить, что основным принципом контроля информации на основе изложенных программных методов является то, что в них для контроля информации рассчитываются контрольные суммы по соответствующим правилам, что характерно практике передачи информации в телекоммуникационных системах.

Наряду с этим, обобщение результатов исследований основных подходов, принципов, правил программного контроля информации приводит к необходимости организации способов контроля при передаче пакетов данных и позволяет предложить следующие варианты.

В первом варианте ПК-отправитель рассчитывает контрольную сумму по соответствующему алгоритму, и считается, что ПК-получатель обладает данным правилом суммирования для контроля информации. Контрольная сумма при этом передаётся в виде самой информации. В данном случае принимается допущение, что вероятность ошибок, происходящих в современных каналах связи, считается значительно меньшей (на несколько порядков) по сравнению с вероятностью ошибок технических устройств ввода и оператора в звене отправителя. Далее переданные ПК-отправителем и рассчитанные ПК-получателем в декодированной информации контрольные суммы сравниваются. В случае их совпадения система принимает решение о достоверности переданной информации, а в случае несовпадения – о наличии ошибок в поле «информация». Для исправления ошибок система переспрашивает информацию вместе с контрольной суммой. Контрольная сумма рассчитывается и проверяется заново. Итерационный процесс продолжается до принятия решения о достоверной передаче информации.

Рассматриваемый способ передачи и контроля информации относится к уровню пользователей услуг сети прикладного уровня сетевой модели OSI. В данном случае, аналогично модулю UDP, средства контроля встраиваются в IP-пакет и синхронизируются с алгоритмами их передачи.

Во втором варианте передачи и контроля информации ПК-отправитель передаёт сперва рассчитанную по соответствующему алгоритму контрольную сумму ПК-получателю в поле «информация» пакета, и ПК-получатель избавляется от повторного расчёта контрольных сумм.

ПК-получатель удостоверяется о достоверности принятой контрольной суммы по алгоритму, вложенному в протоколах. В дальнейшем решение о достоверности информации принимается на основе сравнения уже полученной контрольной суммы с контрольной суммой, переданной вместе с информацией.

Ниже изложены результаты проведённых исследований для практического применения разработанных нами программных методов, использующих искусственную избыточность.

### 3. Принцип использования искусственной избыточности

Рассмотрим разработанный нами метод поразрядного суммирования. Будем считать, что сообщения формируются из  $m$  информационных символов в десятичном коде и для контроля достоверности используется  $n$  контрольных сумм. Для этого последовательность  $a_1, a_2, \dots, a_i, \dots, a_l, \dots$  делится на равные группы, включающие  $m$  символов каждая. Количество информационных символов определяется по формуле  $m = C_n^1 + C_n^2 + \dots + C_n^{n-1}$ , где  $n$  – число контрольных сумм.

Требуется, чтобы алгоритм контроля информации путём суммирования мог установить не только наличие ошибки в принятой последовательности, но и указать номер позиции, в которой произошли ошибки, для автоматического их исправления.

Для простоты представления результатов исследований изложим работу метода при  $n = 4$ . Обозначим через  $a_i$  символ  $i$ -го разряда в проверяемой последовательности ( $i = 1, m$ ), а через  $S_j$  –  $j$ -ю контрольную сумму ( $j = 1, 4$ ). Будем считать количество контролируемых символов в передаваемой последовательности  $m = 14$ .

Контрольные суммы передаваемой последовательности данных вычисляются по схеме [10]

$$\left. \begin{aligned} S_1^0 &= a_1 + a_5 + a_6 + a_7 + a_{11} + a_{12} + a_{13} \\ S_2^0 &= a_2 + a_3 + a_8 + a_9 + a_{11} + a_{12} + a_{14} \\ S_3^0 &= a_3 + a_6 + a_8 + a_{10} + a_{11} + a_{13} + a_{14} \\ S_4^0 &= a_4 + a_7 + a_9 + a_{10} + a_{12} + a_{13} + a_{14} \end{aligned} \right\}, \quad (1)$$

в связи с чем образуются новые последовательности символов с учётом контрольных сумм

$$a_1, a_2, \dots, a_i, \dots, a_{14}, S_1^0, S_2^0, S_3^0, S_4^0. \quad (2)$$

Теперь по схеме (1) вычисляются контрольные суммы ( $S_j^j$ ) принятой последовательности и сравниваются, т.е.:

$$\Delta S_j = S_j^0 - S_j^j, \quad (j = 1 \div 4).$$

Полученные разности контрольных сумм используются для контроля, обнаружения и исправления ошибочных символов.

Излагаемый метод контроля достоверности информации построен на основе пяти правил:

- $\pi_1$  – общее правило контроля достоверности информации;
- $\pi_2$  – обнаружение и исправление однократных транспозиционных ошибок;
- $\pi_3$  – обнаружение и исправление двукратных смежных транспозиционных ошибок;
- $\pi_4$  – обнаружение ошибок типа пропуск или добавление дополнительного символа в строке;
- $\pi_5$  – обнаружение случайных символьных ошибок более высокой кратности.

Исследованы эффективность и предельные возможности метода контроля информации при введённых правилах. Установлено, что при построении алгоритмов контроля информации по правилам  $\pi_2$ ,  $\pi_3$  и  $\pi_5$  ошибки не обнаруживаются в следующих случаях:

- когда одновременно искажаются один, два, три и т.д. символов  $a_i$  и соответствующие им контрольные суммы  $S_j$ ;
- при искажении значение символа  $a_i$  увеличивается на некоторую величину и сумма  $S_j$  уменьшится на эту же величину.

В качестве критерия эффективности метода принят критерий вероятности необнаруженных ошибок. Тогда эффективность алгоритмов, построенных по правилу  $\pi_2$ , равна

$$P_H^1 = 67mnp^2(1-p)^{m+n-2};$$

по правилу  $\pi_3$  –

$$P_H^2 = 67[C_m^2 C_n^2 p^4 (1-p)^{m+n-4}];$$

по правилу  $\pi_5$  –

$$P_H^3 = 67[C_m^k C_n^3 p^{3k} (1-p)^{m+n-k-3}],$$

где  $k$  – кратность ошибок ( $k > 2$ ).

По правилу  $\pi_4$  форматные ошибки не обнаруживаются, когда происходят «пропуск» и «добавление» символов одновременно. Вероятность ошибок типа «пропуск» или «добавление» символов человека-оператора принята равной  $P = 2.4 \cdot 10^{-3}$  ошибок/знак. Учитывая это, получаем вероятность пропуска форматных ошибок:  $P_H^4 = 5.76 \cdot 10^{-6}$  ошибок/знак.

Суммарная вероятность необнаруженных ошибок рассматриваемых правил контроля информации определяется

$$P_H = P_H^1 + P_H^2 + P_H^3 + P_H^4 = \{67[mnp^2(1-p)^{m+n-2} + C_m^2 C_n^2 p^4 (1-p)^{m+n-4} + C_m^k C_n^3 p^{3k} (1-p)^{m+n-k-3}] + 5.76 \cdot 10^{-6}$$

Установлено, что в случае применения разработанного метода обнаруживаются и исправляются все ошибки типа однократных транспозиционных и двукратных смежных транспозиционных, при этом минимальная вероятность необнаруженных ошибок  $P_H \approx 10^{-5} - 10^{-6}$ .

В СЭД и в системах управления распределёнными базами данных более эффективным оказалось применение разработанных алгоритмов контроля информации по плоскостному суммированию.

#### 4. Алгоритмы контроля информации по плоскостному суммированию

Считается, что последовательность знаков и букв, составляющих исходный текст

$$a_1, a_2, \dots, a_i, \dots, a_n \quad (3)$$

с помощью специальной программы преобразуется в последовательность десятичных чисел

$$\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_i, \dots, \alpha_n \quad (4)$$

и соответственно на контроль информации вместо последовательности (3) будет передаваться цифровая последовательность (4).

Для контроля информация  $\alpha_i$  формализуется в виде двумерной матрицы. Затем вычисляются контрольные суммы (КС) путём линейного либо модульного суммирования элементов матрицы по горизонтальным ( $G_1^0, G_2^0, \dots, G_n^0$ ) и вертикальным ( $V_1^0, V_2^0, \dots, V_m^0$ ) сечениям, т.е. по строкам и столбцам массива  $A[i, j]$ :

$$A[i, j] = \begin{array}{cccc|c} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1m} & G_1^0 \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2m} & G_2^0 \\ \dots & \dots & \dots & \dots & \vdots \\ \alpha_{i1} & \alpha_{i2} & \dots & \alpha_{im} & G_i^0 \\ \dots & \dots & \dots & \dots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \dots & \alpha_{nm} & G_n^0 \\ \hline V_1^0 & V_2^0 & \dots & V_m^0 & \hline \end{array}$$

Далее вычисляются КС (т.е.  $G_1^{\exists}, G_2^{\exists}, \dots, G_n^{\exists}, V_1^0, V_2^0, \dots, V_m^0$ ) преобразованной информации и сравниваются КС исходной и преобразованной информации, т.е.

$$\Delta G_j = G_i^0 - G_j^{\exists}, \quad (j = 1 \div 4).$$

$$\Delta V_j = V_i^0 - V_j^{\exists}, \quad (j = 1 \div 4).$$

В зависимости от характера и типов ошибок в сообщениях принимаются решения о достоверности преобразованной информации.

В табл. 1 изложены возможные варианты решения и алгоритмы обработки информации на ЭВМ при обнаружении и исправлении ошибок.

Пропуск ошибки в данных может происходить в двух ситуациях:

– искажаются по две цифры в двух строках массива таким образом, что соответствующие горизонтальные и вертикальные суммы не изменятся, т.е. в каждой строке происходят ошибки типа перестановки цифр, или в результате искажений одна цифра увеличивается (или уменьшается) на некоторую величину, а вторая цифра уменьшается (или увеличивается) на ту же величину;

– искажения информации происходят в двух столбцах матрицы таким образом, что их соответствующие вертикальные и горизонтальные суммы не изменяются.

Таблица 1. Варианты решений и алгоритмы обработки информации при обнаружении и исправлении ошибок

Результаты сравнения КС	Решение на ЭВМ	Алгоритмы обработки информации
$G_i^0 = G_i^{\exists}, V_j^0 = V_j^{\exists}$	Принято верно	Запись $A[i, j]$ и КС
Только $G_k^0 \neq G_k^{\exists}$ ( $1 \leq l \leq m$ )	Ошибка в КС $G_k^0$ .	Исправлять КС $G_k^0$ : $G_k^0 = G_k^{\exists}$ и запись $A[i, j]$ и КС
Только $V_l^0 \neq V_l^{\exists}$ ( $1 \leq l \leq m$ )	Ошибка в КС $V_l^0$	Исправлять $V_l^0 = V_l^{\exists}$ и запись $A[i, j]$ и КС
$G_k^0 \neq G_k^{\exists}, V_l^0 \neq V_l^{\exists},$ ( $1 \leq k \leq n$ ); ( $1 \leq l \leq m$ )	Ошибка в $\alpha_{kl}$	Исправлять $\alpha_{kl}$ : $\alpha_{kl} = \alpha_{kl} + \Delta G_k$ или $\alpha_{kl} = \alpha_{kl} + \Delta V_l$ , где $\Delta G_k = G_k^0 - G_k^{\exists}$ и $\Delta V_l = V_l^0 - V_l^{\exists}$
$G_k^0 \neq G_k^{\exists}; G_p^0 \neq G_p^{\exists}$ и $V_l^0 \neq V_l^{\exists}$ ( $1 \leq k, p \leq n$ ) ( $1 \leq l \leq m$ ), $k \diamond p$	Ошибка типа перестановки в символах $\alpha_{kl}$ и $\alpha_{pl}$	Поменять местами символы $\alpha_{kl}$ и $\alpha_{pl}$ , затем запись всех КС и $A[i, j]$ в память
$G_k^0 \neq G_k^{\exists}; V_l^0 \neq V_l^{\exists}$ и $V_q^0 \neq V_q^{\exists}; 1 \leq k \leq n;$ $1 \leq l, q \leq m; l \diamond q$ .	Перестановка в символах $\alpha_{kl}$ и $\alpha_{pq}$	Поменять местами символы $\alpha_{kl}$ и $\alpha_{pq}$ , затем запись всех КС и $A[i, j]$ в память
$G_k^0 \neq G_k^{\exists}; G_p^0 \neq G_p^{\exists}$ и $V_l^0 \neq V_l^{\exists}$ и $V_q^0 \neq V_q^{\exists};$ $1 \leq k, p \leq n; 1 \leq l, q \leq m;$ $k \neq p$ и $l \neq q$	Ошибка в двух символах $\alpha_{kl}$ и $\alpha_{pq}$	Исправить ошибки $\alpha_{kl}$ и $\alpha_{pq}$ $\alpha_{kl} = \alpha_{kl} + \Delta G_k$ ; $\alpha_{pq} = \alpha_{pq} + \Delta G_p$ или $\alpha_{kl} = \alpha_{kl} + \Delta V_l$ ; $\alpha_{pq} = \alpha_{pq} + \Delta V_q$ , где $\Delta G_k = G_k^0 - G_k^{\exists}$ ; $\Delta G_p = G_p^0 - G_p^{\exists}$ ; $\Delta V_l = V_l^0 - V_l^{\exists}$ ; $\Delta V_q = V_q^0 - V_q^{\exists}$ и записать все данные в память
В остальных случаях	Ошибка высокого порядка	Вывод всех КС и $A[i, j]$

Будем считать, что число таких искажений равно 67, две строки и два столбца матрицы подбираются соответственно  $C_n^2$  и  $C_m^2$  способами, тогда вероятность необнаруженных ошибок алгоритма оценивается по формуле

$$P_H = C_m^2 [67 C_n^2 p^2 (1-p)^{n-2}]^2 + C_m^2 [67 C_m^2 p^2 (1-p)^{m-2}]^2.$$

Отметим, что полученные результаты дают возможность построения методик теоретических и экспериментальных исследований для обоснованного выбора практического применения того или иного метода контроля информации. В связи с этим исследованы границы эффективности и предельные возможности методов по различным критериям.

## 5. Оценка эффективности программных методов

В качестве критериев оценки эффективности контроля выбраны вероятность пропуска ошибок –  $P_H$ , трудоёмкость –  $T_M$  и стоимость реализации метода контроля –  $C$ . За критерий трудоёмкости метода контроля принято процессорное время, затрачиваемое на контроль информации, а за критерий стоимости реализации – стоимость контроля одного пакета информации. Результаты получены при следующих заданных условиях: объём передаваемой тестируемой информации равен 8 Мбайт; значение экспериментально установленной средней вероятности ошибок человека-оператора и сканирующих и распознающих устройств –  $P = 3,7 \cdot 10^{-3}$ ; методом контроля защищена последовательность кадров, принадлежащих одному пакету; ошибок обработки, обусловленных сбоями устройств, нет.

На рис. 2 и 3 приводятся графики эффективности исследованных алгоритмов, полученные на основе критериев трудоёмкости и стоимости контроля, где по осям ординат расположены значения соответствующих критериев, а по осям абсцисс – задаваемые экспериментальные объёмы перерабатываемой информации.

Трудоёмкость контроля вычисляется по разработанной нами методике, и конечное выражение критерия определяется как

$$T_M = \frac{T_{обн}}{W_{пер}},$$

где  $T_{обн}$  – время, затрачиваемое на выполнение арифметических и логических операций для контроля одного пакета, объём информации которого рассчитан в десятичных знаках;  $W_{пер}$  – время, затрачиваемое на передачу, обработку информации объёмом  $N$  десятичных знаков.

Стоимость контроля определяется из выражения

$$C = \frac{C_K}{Q},$$

где  $C_K$  – стоимость ввода, формирования и контроля информации одного пакета,  $Q$  – общий объём перерабатываемой информации в десятичных знаках.

По результатам можно судить о том, что выбор того или иного метода повышения достоверности данных тесно связан с материальными и временными затратами: чем эффективнее алгоритм контроля, тем сложнее его реализация. Время реализации метода контроля резко возрастает с повышением требований, предъявленных к достоверности данных.

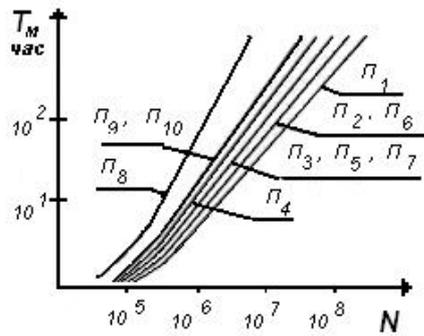


Рис.2. Зависимость трудоёмкости ( $T_M$ ) исследованных алгоритмов от объёма перерабатываемой информации ( $N$ )

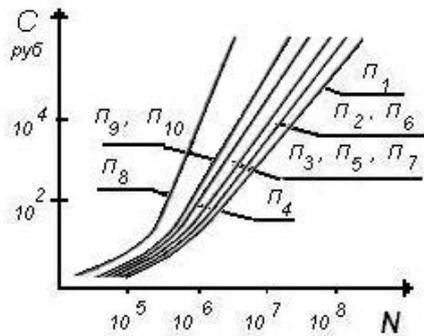


Рис.3. Зависимость стоимости реализации ( $C$ ) исследованных алгоритмов от объёма перерабатываемой информации ( $N$ )

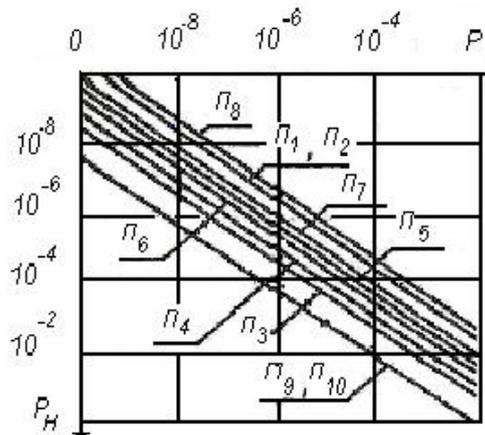


Рис.4. Зависимость вероятности пропуска ошибок ( $P_H$ ) от средней вероятности ошибок на десятичный символ ( $P$ ) для исследованных алгоритмов

На рис. 4 показан график зависимости вероятности пропуска ошибок алгоритмами контроля  $P_H$  от средней вероятности ошибок  $P$  на десятичный символ. Эти графики показывают, что при вероятности ошибок в пределах  $10^{-3} - 10^{-4}$  соотношение способностей обнаружения ошибок исследованных алгоритмов достигает двух порядков, трудоёмкость и стоимость контроля снижается в 2 – 3 раза за счёт автоматического исправления ошибок.

Наибольший эффект по достоверности информации достигается при применении алгоритмов:  $\Pi_2$  – поразрядного суммирования;  $\Pi_9$  – суммирования по модулю 10;  $\Pi_{10}$  – суммирования по модулю 11.

В связи с этим ниже изложены результаты исследований, связанных с построением методов контроля информации по произвольному модулю  $q$ .

## 6. Модель контроля информации

Избыточные коды с контролем по модулю  $q$ , равному основанию кода, широко используются для обнаружения ошибок человека-оператора, занятого вводом цифровой информации [11, 12]. Разработанный нами метод предназначен для обнаружения ошибок путем контроля по модулю  $q$ , равному основанию кода. По методу считается, что последовательность информационных символов  $\alpha_1\alpha_2\dots\alpha_n$  принадлежит коду, если выполняется следующее условие

$$\sum_{i=1}^n f_i(\alpha_i) \equiv 0 \pmod{q}, \quad (5)$$

где:  $q$  – основание кода;  $\alpha_i \in \{0, q-1\}$  для  $i = \overline{1, n}$ ;  $f_i(\alpha_i)$  – функция, закреплённая за  $i$ -й позицией;  $f_i(\alpha_i) \in \{0, q-1\}$  для  $\alpha_i = \overline{0, q-1}$ ,  $i = \overline{1, n}$ . В простейшем случае кодирования  $f_i(\alpha_i) = \alpha_i$  для некоторого  $i$ .

Для обоснования предлагаемого метода нами исследованы различные модели расчёта контрольных сумм по строкам и столбцам перекодированной последовательности и вероятности необнаружения ошибок при применении модульных способов суммирования, принадлежащих к классу избыточных кодов. В табл. 2 даны обозначения алгоритмов (графа 1), значения модулей суммирования, математические модели правил суммирования, вероятности необнаруженных ошибок (графа 2) и их экспериментальные значения (графа 3).

Таблица 2. Вероятности необнаружения ошибок для различных алгоритмов контроля

Обозначение алгоритма	Вероятность необнаруженной ошибки $P_H$	Расчётное значение $P_H$
Модуль $q = 9$		
Модель - $K = \sum_{i=1}^{N/2} (a_{2i-1}) \bmod q - \sum_{i=1}^{N/2} (a_{2i}) \bmod q$		
$B_1$	$C_{n/2}^2 \sum_{i=1}^t \frac{\alpha_i}{(A_i^2 + t)} p^2 (1-p)^{N-2}$	$7 \cdot 10^{-5}$
Модуль $q = 9$		
Модель - $K = \sum_{i=1}^N a_i - qF$		
$B_2$	$88P_{3N} \omega^2 p^2 \sum_{i=1}^9 (10 - \delta_i)$	$5 \cdot 10^{-5}$
Модуль $q = 10$		
Модель - $K = \sum_{i=1}^N a_i - qF$		
$B_3$	$80P_{3N} \omega^2 p^2 \sum_{i=1}^3 (10 - \delta_i)$	$1.5 \cdot 10^{-5}$
Модуль $q = 11$		
Модель - $(\sum_{i=1}^N a_i + K) \bmod q = 0$		
$B_4$	$8P_{3N} \omega^2 P^2 [\sum_{i=2}^3 (10 - \delta_i)(\delta_i - 1)]$	$1.2 \cdot 10^{-5}$

Здесь приняты следующие обозначения:

$q$  – модуль суммирования;

$K$  – контрольная сумма;

$n$  – количество разрядов в числе;

$N$  – количество разрядов в строке;

$t$  – основание системы счисления ( $t = 10$ );

$A_i^2$  – число размещений из  $t$  по 2;

$a_i$  – число сочетаний пар цифр;

$F$  – число десятков, образующееся при сложении;

$\delta_i$  и  $\delta_j$  – числовые значения ошибок в  $i$ -м и  $j$ -м разрядах;

$P_i$  – вероятность  $i$ -ых сочетаний знаков в искажённом коде;

$p$  – вероятность  $i$ -го искажения в разряде кода;

$\omega$  – частота появления цифры в разряде строки.

Определено, что рассмотренные алгоритмы эффективно проявляют себя при устранении: однократных ошибок замещения типа транскрипционной ошибки  $aabb \rightarrow aabc$ , например, в коде  $1238 \rightarrow 1237$ ; перестановки символов на соседних позициях – транспозиционные ошибки вида  $bc \rightarrow cb$ , например,  $1238 \rightarrow 1283$ ; идентичные ошибки замещения в двух соседних позициях – транскрипционные ошибки вида  $bb \rightarrow cc$ , например,  $1335 \rightarrow 1995$ .

Так как ошибки вида  $b \rightarrow c$  составляют большую часть всех ошибок, то задачу построения конкретного модуля проверки информации типа (1) естественно формулировать следующим образом: для заданного  $q$  найти функции  $f_i(x)$ , обеспечивающие обнаружение всех ошибок  $b \rightarrow c$  и возможно большего количества ошибок видов  $bc \rightarrow cb$  и  $bb \rightarrow cc$ . Следует уделить внимание повышению количества обнаруживаемых ошибок и других, менее вероятных видов, если это не помешает обнаружению ошибок вида  $b \rightarrow c$ ,  $bc \rightarrow cb$ ,  $bb \rightarrow cc$ .

Решению указанной задачи применительно к цифровой информации ( $q = 10$ ) посвящено сравнительно много исследований [11, 12]. Вместе с тем, представляет интерес исследование вопроса оценки границ обнаружения ошибок, достигаемых предложенным методом контроля информации по произвольному модулю  $q$ .

Нами для чётного  $q$  получена верхняя граница количества обнаруживаемых всех ошибок вида  $bc \rightarrow cb$ . Для  $q = 10$  и  $q = 11$  построены коды и определено, что коды при  $q = 11$  обнаруживают больше ошибок видов  $bc \rightarrow cb$  и  $bb \rightarrow cc$ . Теперь изложим методику определения границ количества обнаруженных ошибок предлагаемого избыточного кода.

## 7. Границы количества обнаруживаемых ошибок

Методика оценки границ количества обнаруживаемых ошибок более подробно изложена в работе автора [13]. Обозначим через  $N_1, N_2$  и  $N_3$  относительные количества обнаруживаемых ошибок соответственно видов  $b \rightarrow c$ ,  $bc \rightarrow cb$  и  $bb \rightarrow cc$ , понимая под относительным количеством  $N_u$  отношение количества обнаруживаемых ошибок некоторого вида к количеству всех возможных ошибок этого вида. Количество всех ошибок видов  $b \rightarrow c$ ,  $bc \rightarrow cb$  и  $bb \rightarrow cc$  будем считать соответственно равными  $nq(q-1)$ ,  $(n-1)q(q-1)$  и  $(n-1)q(q-1)$ .

Будем считать, что если обнаруживаются все ошибки вида  $b \rightarrow c$ , то  $N_1 = 1$  и  $q$  – чётное,  $N_2 \leq 1 - A$ ,  $N_3 \leq 1 - A$ , где

$$A = \frac{2}{q(q-1)}, \quad (6)$$

и метод достигает этой границы при модульном суммировании по  $q = 10$  и  $q = 11$ .

Доказано, что в коде типа (5) для  $N_2 < 1 - A$  и  $N_3 < 1 - A$  достаточно иметь один проверочный символ, например,  $\alpha_i$ . Для этого  $f_i(\alpha_i)$  представлена в виде сложной функции

$$f_i(\alpha_i) = F_i(y_i) = F_i[f_{i+1}(\alpha_i)],$$

где  $F_i(y_i) \in \overline{0, q-1}$  для  $y_i \in \overline{0, q-1}$ ,  $i = \overline{1, n-1}$ . В этом случае все ошибки вида  $b \rightarrow c$  обнаруживаются, если

$$F_i(y_1) \neq F_i(y_2). \quad (7)$$

где  $y_1 \neq y_2$ ;  $y_1, y_2 \in M$ ;  $i = \overline{1, n-1}$ ;  $M = \{\overline{0, q-1}\}$ . Здесь через  $M$  обозначено множество  $\{\overline{0, q-1}\}$ , из которого исключен элемент  $s \neq u$ .

Следует отметить, что функции  $F_i(y)$  могут быть одинаковыми для разных и даже для всех позиций  $i$ . Поэтому для достижения границы (6) достаточно найти хотя бы одну функцию  $F(y)$ .

Теперь рассмотрим случай, когда  $N_2 = N_3 = 1 - A$ . В этом случае граница (6) достигается, если выполняются следующие условия:

$$F_i(y_1) - y_1 \neq F_i(y_2) - y_2 \pmod{q}, \quad (8)$$

где  $y_1 \neq y_2$ ;  $y_1, y_2 \in M_u$ ;  $i = \overline{1, n-1}$ ;  $u \in \{\overline{0, q-1}\}$ .

$$F_i(y_1) + y_1 \neq F_i(y_2) + y_2 \pmod{q}, \quad (9)$$

где  $y_1 \neq y_2$ ;  $y_1, y_2 \in M_s$ ;  $s \in \{\overline{0, q-1}\}$ ;  $i = \overline{1, n-1}$ .

В условиях (8) и (9):

$$F_1(y) = F_1(y) + D, \quad D \in \overline{0, q-1}; \quad (10)$$

$$F_2(y) = q - F_2(y).$$

Для лучшего обзора практического значения вышеприведённых выкладок дадим примеры функций для различных значений модуля  $q$ .

$q=8$ :  $F_1(y) = 2y + 1$  для  $y = \overline{1, 2, 3, 4}$ ;

$F_2(y) = -2y$  для  $y = \overline{5, 6, 7, 0}$ .

$q=9$ :  $F(y) = ty$  для  $y = \overline{0, 4}$ ;

$F(y) = t(y+1)$  для  $y = \overline{5, 8}$ ;

$F(9) = 5$  где  $t = 3$  или  $7$ .

Исследования показали, что для  $q = 10$  путём перебора на ЭВМ можно найти около 140 функций  $F(y)$ , которые удовлетворяют условиям (7), (8), (9) и которые нельзя получить друг из друга преобразованиями (10).

Если  $q = 10$ , то при сравнении с рассмотренными модулями не только учитываются наиболее вероятные однократные и двукратные ошибки, но и некоторые трёхкратные ошибки типа перестановок, например,  $1238 \rightarrow 1832$  или идентичные ошибки замещения в трёх соседних разрядах, например,  $13332 \rightarrow 19992$ . В соответствии с вышеизложенным, относительные количества обнаруживаемых ошибок указанных видов обозначим соответственно  $N_4$  и  $N_5$ .

Определено, что метод суммирования по модулю  $q = 11$  при принятой нами функции  $f_i(x)$  достигает границы (2) и обеспечивает:

$$N_1 = 1, N_2 = N_3 = 1 - A \quad N_4 = N_5 = 1 - 2A.$$

Кроме того, нами получены результаты при других функциях  $f_i(x)$  и  $N_1 = 1$ :

$$f_i(\alpha_i) = \alpha_i + q: \quad N_2 = 1 - A, N_3 = 1 - 3A, \\ N_4 = 1 - 2A, N_5 = 1 - 4A;$$

$$f_i(\alpha_i) = \alpha_i + 2q: \quad N_2 = N_3 = N_4 = 1 - 2A, \\ N_5 = 1 - 7A;$$

$$f_i(\alpha_i) = \alpha_i + 3q: \quad N_2 = 1 - 3A, N_3 = 1 - A, \\ N_4 = 1 - 6A, N_5 = 1 - 4A.$$

Следует отметить, что правомерность полученных результатов совпадает с мнениями авторов в работах [11, 12], которые исследовали достоверность передачи цифровой информации.

## 8. Заключение

Проведённый анализ теории и практики повышения достоверности информации в современных протоколах передачи данных на основе механизма использования кодовой и аппаратной избыточности позволил инициировать актуальное исследование, связанное с контролем и коррекцией ошибок на Прикладном и Представительском уровнях модели OSI, т.е. борьбой с ошибками человека-оператора, сканирующих и других устройств, предназначенных для ввода информации. Наряду с этим, уделено внимание задаче контроля информационной части заголовков пакетов на Транспортном, Сетевом, Физическом и Канальном уровнях, в связи с чем выбрано направление исследований, связанное с разработкой программных методов контроля достоверности информации в структуре пакетов передачи данных.

Исследованы основные подходы использования существующих программных методов контроля достоверности передачи и обработки информации в автоматизированных системах управления предприятиями и технологическими процессами за счёт использования искусственной и естественной избыточности, которые, в основном, ориентированы на повышение достоверности цифровой информации, служащие основным инструментом при коррекции также текстовой информации.

Увеличение возможностей компьютерных технологий и современных телекоммуникационных систем позволяет вновь рассмотреть и использовать сложившиеся подходы для организации контроля информации, передаваемой как в территориальных, так и в глобальных и мобильных сетях передачи данных.

Установлено, что при использовании программных методов повышения достоверности информации, теоретические и практические стороны проблемы контроля ошибок разнородной, в частности, текстовой информации изучены пока недостаточно. Отсутствуют приемлемые методы, алгоритмы и правила контроля и коррекции ошибок человека-оператора, сканирования и распознавания и орфографических ошибок.

В связи с этим исследованы и получены оценки эффективности программных методов контроля, использующих посимвольное и поразрядное суммирование, суммы информационных знаков, поразрядное суммирование по модулю 9, цифровой контроль, контрольные числа, двойной подсчёт, по справочнику, по модулям 10 и 11.

Исследование эффективности и предельных возможностей программных методов по критериям вероятности пропуска ошибок, трудоёмкости и стоимости контроля, а также получение верхних границ ошибок при применении моделей контроля информации по модулям 10 и 11 доказало, что разработанные методы удачно дополняют реализованные в современных пакетах передачи технологии и обеспечивают требуемую достоверность информации

при незначительных временных и материальных затратах, причём предложенным алгоритмом контроля информации обнаруживаются ошибки в звеньях подготовки и ввода информации, которые не учитываются в кодах, используемых при проверке информации по контрольным суммам в современных протоколах.

## Литература

1. Трубочёв А.П. и др. Оценка безопасности информационных технологий. Под общ. ред. В.А. Галатенко. - М.: СИП РИА, 2001. - 356 с.
2. Degermark, M., Nordgren, B. and Pink S. "IP Header Compression", RFC 2507, February 1999.
3. Casner, S. and Jacobson V. "Compressing IP/UDP/RTP Headers for Low-Speed Serial Links", RFC 2508, February 1999.
4. A.R.Akhatov, I.I.Jumanov, O.I.Djumanov. An Effective Quality Control of Textual Information on the Basis of Statistical Redundancy in Distributed Mobile IT Systems and e-Applications//3rd International Conference in Central Asia on Internet, Tashkent, 2007, IEEE Catalog Number: 07EX1695C, ISBN: 1-4244-1007-X, Library of Congress: 2007920881.
5. А.Р.Ахатов. Методы и алгоритмы обеспечения достоверности текстовой информации на основе статистической избыточности// Вестник ТашГТУ, № 2, 2007, ТГТУ, Ташкент, с.41-44.
6. А.Р.Ахатов, И.И.Жуманов. Алгоритм контроля качества текстов в системах электронного документооборота // Вестник ТУИТ, № 2, 2007, Ташкент, с. 68-72.
7. Akhatov A.R., Jumanov I.I. Improvement of Text Information Processing Quality in Documents Processing Systems//2nd IEEE/IFIP International Conference in Central Asia on Internet ICI-2006, September 19-21, International Hotel Tashkent, Uzbekistan, <http://ieeexplore.ieee.org/xpl/RecentCon.jsp?punumber=4055161> (дата обращения: 05.05.2009).
8. Jumanov I.I., Akhatov A.R., Tursinxanov N.M. Methods and Algorithms of Input Information Protection in Electronic Document Processing Systems//2nd IEEE/IFIP International Conference in Central Asia on Internet ICI-2006, September 19-21, International Hotel Tashkent, Uzbekistan, <http://ieeexplore.ieee.org/xpl/RecentCon.jsp?punumber=4055161> (дата обращения: 05.05.2009).
9. Жуманов И.И., Ахатов А.Р., Мингбаев Н.С. Методы обнаружения и исправления ошибок при передаче текстов на узбекском языке // Сб. науч. тр. «Вопросы кибернетики». РИСО АН РУз. № 163. – Ташкент, 2002 – с.54-60.
10. Ахатов А.Р. Логический алгоритм обнаружения и исправления ошибок в машинописных текстах на узбекском (латиница, кириллица), русском, английском языках// Республиканский журнал «Проблемы информатики и энергетики», № 2, Ташкент, Узбекистан 2003 г., с.66-71.
11. Б.Б.Трофимов, В.А.Литвинов. Методы обнаружения ошибок в алфавитно-цифровых последовательностях на этапе подготовки и ввода данных в ЭЦВМ // Управляющие системы и машины, № 4, 1977, с. 20-27.
12. A.M. Andrew. A variant of modulus 11 checking // The computer Bulletin. Vol. 14, № 8, 1970, p.261-265.
13. А.Р. Ахатов. Оценка эффективности программных методов контроля достоверности информации на основе избыточных кодов// Журнал «Инфокоммуникации: Сети – Технологии – Решения», Узбекское агентство связи и информатизации, Центр научно-технических и маркетинговых исследований, Ташкент, 2007, № 2, с. 32-37.

*Статья поступила в редакцию 08.08.2008;  
исправленный вариант представлен 07.01.2009*

**Ахатов Акмал Рустамович**

к.т.н., доцент кафедры информационных технологий Самаркандского государственного университета, (703004, Узбекистан, Самарканд, Университетский бульвар, 15)  
тел. (8366) 220-6881, (8366) 233-6202, e-mail: akmalar@rambler.ru

**Program Methods for Information Fidelity Checking in Data Transmission Packets of Document Management Systems****A. R. Akhatov**

We consider the problem of increasing fidelity of information transmission in document management systems by using program techniques. Principles, rules, and algorithms of checking information fidelity have been developed based on linear letterwise and bitwise sums, two-dimensional sums, and modular sums. The effectiveness and performance bounds of the algorithms have been investigated with respect to error non-detection probabilities, complexity, and cost. Upper bounds of errors have been obtained for modular checking.

*Keywords:* document management systems, fidelity checking.