СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

**Научный журнал Российской академии наук
(издается под руководством Отделения нанотехнологий
и информационных технологий РАН)**

Издается с 1989 года
Журнал выходит ежеквартально

**Учредитель:
Федеральный исследовательский центр
«Информатика и управление» Российской академии наук**

РЕДАКЦИОННЫЙ СОВЕТ

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Редакция

Журнал включен в базу данных Russian Science Citation Index (RSCI),
интегрированную с Web of Science
Журнал входит в систему Российского индекса научного цитирования (РИНЦ)
Журнал включен в базу данных CrossRef (систему DOI — Digital Object Identifier),
в базу данных Ulrich's periodicals directory
и в информационную систему «Общероссийский математический портал Math-Net.Ru»

Журнал реферируется в «Реферативном журнале» ВИНИТИ
и в системе Google Scholar

Журнал включен в сформированный Минобрнауки России Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук

http://www.ipiran.ru/journal/collected

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

Том 31 № 4 Год 2021

СОДЕРЖАНИЕ

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

Том 31 № 4 Год 2021

СОДЕРЖАНИЕ

НЕКОТОРЫЕ ВОПРОСЫ ОЦЕНКИ КАЧЕСТВА ИНФОРМАЦИОННЫХ СИСТЕМ

А. А. Зацаринный[1], *Ю. С. Ионенков*[2]

Аннотация: Статья посвящена вопросу оценки качества информационных систем (ИС). Представлены понятия качества продукции в целом и качества ИС в частности. Рассмотрены отличия понятий «качество» и «эффективность». Приведены наиболее распространенные в настоящее время зарубежные и отечественные модели качества ИС. Отмечено, что большинство из них относится к оценке качества программных средств, в то время как нормативная база оценки качества ИС в целом развита недостаточно. Рассмотрены основные методы оценки качества ИС. Сформулированы предложения по совершенствованию нормативной базы в области оценки качества ИС.

Ключевые слова: качество; оценка качества; метод оценки качества; показатель; информационная система

DOI: 10.14357/08696527210401

1 Введение

Информационные системы приобретают все большую значимость, особенно в условиях принятого курса на цифровую трансформацию общества. При этом сложность ИС постоянно возрастает, в основном за счет расширения их функциональности, а также условий применения. Эти и другие факторы привели к повышению требований к качеству ИС и безопасности их применения, причем как со стороны заказчиков услуг (пользователей), так и разработчиков. Поскольку интересы пользователей могут существенным образом различаться, а представления специалистов о функционировании ИС не совпадать, оценка качества этих систем представляется весьма сложной задачей.

В современной литературе качество системы (услуги, продукции) понимается как степень соответствия их полезных свойств потребностям и предпочтениям потребителей.

Качество ИС является комплексной характеристикой совокупности технических, технологических, эксплуатационных, экономических и других показателей и требует постоянного совершенствования подходов к ее оценке. Необходимо определить перечень показателей качества, по которым будут оцениваться сис-

[1]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, AZatsarinny@ipiran.ru

[2]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, uionenkov@ipiran.ru

темы, а также провести обобщенную оценку по всем выбранным показателям. Оценка качества ИС проводится на всех стадиях их жизненного цикла. При этом оценка качества ИС в современных условиях требует учета сложившихся к настоящему времени методов его оценки, а также адаптации этих методов к специфике и конкретным типам ИС.

В статье рассматриваются вопросы, относящиеся к определению понятия «качество ИС», включая наиболее распространенные модели качества ИС, а также методы оценки их качества с учетом понятия «качество» в целом.

2 Понятие качества информационных систем

В настоящее время в нормативных документах используются несколько определений понятия качества продукции (объекта):

- совокупность свойств продукции, обусловливающих ее пригодность удовлетворять определенные потребности в соответствии с ее назначением (ГОСТ 15467-79) [1];
- совокупность характеристик объекта, относящихся к его способности удовлетворять установленные и предполагаемые потребности (ИСО 8402-94) [2];
- степень соответствия совокупности присущих характеристик объекта требованиям (ГОСТ Р ИСО 9000-2015) [3].

Из этих определений качества можно выделить следующие элементы, которые включает в себя понятие качества:

- свойства продукции;
- возможность выражения качества в количественных оценках;
- способность удовлетворять не только существующие потребности, но и будущие;
- соответствие своему назначению;
- соответствие стандартам.

Существующие стандарты исходят из того, что качество объекта определяется его потребительскими свойствами. Подобная постановка вопроса представляется естественной, так как для потребителя важно в первую очередь то, насколько объект удовлетворяет его потребности.

Исходя из приведенных выше определений качества продукции, а также особенностей ИС, качество ИС — это совокупность свойств системы, обусловливающих возможность ее использования для удовлетворения определенных в соответствии с ее назначением потребностей.

Отметим соотношение понятий «качество» и «эффективность»: если качество представляется как совокупность свойств системы, обусловливающих ее пригодность для использования по назначению, то эффективность характеризует

степень соответствия системы назначению и приспособленности к достижению целей, поставленных при ее создании.

Определение качества и требования к качеству — разные категории. Требования к качеству могут быть выражены структурированной системой характеристик (показателей) качества. Такая система показателей называется моделью качества.

3 Модели качества информационных систем и их отражение в существующих стандартах

Наиболее распространенными моделями качества ИС в настоящее время являются модели МакКола [4], Боэма [5], Гилба [6] и созданные на их основе модели, закрепленные нормативными документами — ГОСТ 28195-89 [7], ГОСТ ИСО/МЭК 9126-2001 [8], ГОСТ Р ИСО/МЭК 25010-2015 [9] и ГОСТ РВ 51987-2002 [10].

Отметим, что большинство данных моделей направлено на оценку качества программных средств. Приведенные стандарты содержат достаточно подробный перечень показателей, позволяющих оценивать качество программного обеспечения для ИС, и широко используются на практике. При этом перечень требований и показателей качества ИС в целом содержит лишь один стандарт — ГОСТ РВ 51987-2002 [10].

Модель МакКола [4] классифицирует все требования к программному обеспечению по 11 показателям качества программного обеспечения. Эти 11 факторов сгруппированы в три категории — эксплуатация продукта, пересмотр продукта и факторы перехода продукта.

К факторам эксплуатации продукта относятся: правильность; надежность; эффективность; целостность; удобство использования. Факторы пересмотра продукта включают: ремонтопригодность; гибкость; тестируемость. Факторы перехода продукта — переносимость, возможность повторного использования и совместимость.

Критерии качества — это числовые уровни факторов, поставленные в качестве целей при разработке. Объективно оценить или измерить факторы качества непосредственно довольно трудно, поэтому МакКол ввел метрики качества, которые, с его точки зрения, легче измерять и оценивать. Оценки метрик по его шкале принимают значения от 0 до 10.

К критериям качества отнесены:

- прослеживаемость;
- функциональная полнота;
- последовательность проектирования;
- правильность;
- устойчивость к ошибкам;
- эффективность выполнения;

– управление доступом;
– контроль за доступом;
– удобство работы;
– удобство обучения;
– способность к взаимодействию;
– простота работы;
– краткость;
– полнота протоколирования;
– информативность;
– расширяемость;
– широта использования;
– модульность;
– независимость от программной платформы;
– независимость от аппаратной платформы;
– унификация интерфейсов;
– унификация данных.

Боэм [5] предложил свою модель, представляющую, по существу, расширение модели МакКола. В ней атрибуты качества подразделяются по способу использования программной системы. Определены 19 промежуточных атрибутов, в которые входят все 11 факторов качества модели МакКола. Эти промежуточные атрибуты представляют собой группы так называемых примитивных атрибутов, которые оцениваются с помощью определенных для них метрик. По сравнению с моделью МакКола в модели Боэма дополнительно измеряются:

– способность к восстановлению функций;
– адекватность;
– ясность;
– удобство внесения изменений;
– понятность;
– универсальность;
– документирование;
– экономическая эффективность.

Модель Гилба [6] соответствует общей концепции предыдущих моделей, но имеет несколько отличий. Модель качества встраивается в проектную специфи-

кацию, а каждый атрибут должен быть измеримым и детализироваться в процессе жизненного цикла программной системы. Помимо атрибутов качества в модель входят атрибуты ресурсов. Модель основана на четырех качественных (применимость, полезность, приспособляемость, удобство использования) и четырех ресурсных (время, бюджет, исполнители, средства разработки) атрибутах, которые можно расширять. Гилб разработал 7 принципов использования модели, главный из которых — принцип измеримости: все атрибуты могут и должны быть на практике измеримыми.

Необходимо отметить, что приведенные выше модели носят концептуальный характер; реальных систем, которые полностью соответствовали бы этим моделям, нет. Вместе с тем существующие стандарты, разработанные с учетом имеющихся моделей качества, учитывают особенности реальных ИС.

Рассмотрим некоторые особенности оценки качества применительно к существующим стандартам.

ГОСТ 28195-89 [7] включает в себя общие положения по оценке качества программных средств, описывает процессы планирования уровня качества, а также процессы контроля значений показателей качества в процессе разработки и испытаний. Показатели качества разбиты на 6 групп и 19 комплексных показателей. Группы определяют пользовательские свойства программных средств, комплексные показатели — программные свойства, от значений которых зависит значение пользовательских свойств. В зависимости от типа программного средства выбирается номенклатура показателей качества, которая должна быть зафиксирована в техническом задании на разработку программного средства.

В ГОСТ ИСО/МЭК 9126-2001 [8] оценка качества программных систем основана на трехуровневом рассмотрении. Уровень цели — то, что пользователь желает видеть в программном обеспечении. Уровень атрибутов — свойства, отражающие приближение к целям. Уровень метрик — количественные характеристики степени наличия атрибутов. В модели выделены 6 целей: функциональность; надежность; практичность, или удобство использования; эффективность; сопровождаемость; переносимость, или мобильность. Цели подразделяются на атрибуты качества.

Полный список атрибутов качества программных систем по стандарту ИСО/МЭК 9126-2001 приведен в табл. 1.

ГОСТ Р ИСО/МЭК 25010-2015 [9] уточняет ряд положений предыдущего стандарта и включает те же характеристики качества программного обеспечения с некоторыми поправками:

– вместо использования специфичных для программного обеспечения определений там, где это возможно, приняты универсальные определения для расширения области применения до компьютерных систем;

– в качестве комплексного показателя качества добавлено «покрытие контекста» с единичными показателями «полнота контекста» и «гибкость»;

Таблица 1 Список атрибутов качества программных систем по стандарту ИСО/МЭК 9126-2001

Комплексные показатели качества	Единичные показатели качества
Функциональные возможности	Пригодность Правильность Способность к взаимодействию Согласованность Защищенность
Надежность	Стабильность Устойчивость к ошибкам Восстанавливаемость
Практичность	Понятность Обучаемость Простота использования
Эффективность	Характер изменения во времени. Характер изменения ресурсов
Сопровождаемость	Анализируемость Изменяемость Устойчивость Тестируемость
Мобильность	Адаптируемость Простота внедрения Соответствие Взаимозаменяемость

- как комплексный показатель качества была добавлена «безопасность» с единичными показателями «конфиденциальность», «целостность», «безотказность», «отслеживаемость» и «подлинность»;
- добавлена как комплексный показатель качества «совместимость»;
- добавлены единичные показатели качества «функциональная полнота», «емкость», «защищенность от ошибки пользователя», «доступность», «готовность», «модульность» и «возможность многократного использования».

Перечень требований и показателей качества ИС в целом содержит лишь один стандарт — ГОСТ РВ 51987-2002 [10]. Положения этого стандарта могут применяться при формировании требований технического задания, при сравнительном анализе, оценке и обосновании технических решений, при проведении испытаний (в том числе сертификационных), при контроле качества функционирования создаваемых, модернизируемых и эксплуатируемых ИС, т. е. практически на всех этапах жизненного цикла современных систем. Стандарт содержит 10 характеристик качества функционирования ИС (надежность, своевременность, полнота, актуальность, безошибочность, конфиденциальность

и др.) и 16 основных показателей качества функционирования ИС, для которых задаются конкретные значения. Вопросы применения данного стандарта достаточно подробно рассмотрены в [11, 12]. Вместе с тем по большинству обобщенных показателей качества данный стандарт предлагает всего один–два частных показателя качества, что недостаточно для объективной и всесторонней оценки системы. Кроме того, данный документ разработан около двух десятков лет назад и не в полной мере учитывает особенности современных ИС. Целесообразна разработка аналога данного стандарта, содержащего новый расширенный перечень обобщенных и частных показателей качества для современных ИС.

Таким образом, подавляющее большинство моделей качества и стандартов посвящены оценке качества только программных средств. В то же время оценка качества ИС должна охватывать все ее элементы: технические средства; программное обеспечение; другие виды обеспечения, включая организационное; подсистему эксплуатации и др.

4 Предложения по совершенствованию существующей нормативной базы

Представленный выше анализ указывает на необходимость совершенствования существующей нормативной базы оценки качества ИС. Так, в некоторых работах ФИЦ ИУ РАН сделаны попытки доработать единственный стандарт, относящийся к ИС в целом (ГОСТ РВ 51987-2002), в направлении учета особенностей современных ИС, включая вопросы информационной безопасности, организационного обеспечения и др. [13, 14]. В частности, предложен перечень показателей качества для типовой ИС на основе доработанных и расширенных требований этого стандарта, в котором увеличено число обобщенных и частных показателей качества, а также введены показатели, учитывающие особенности современных ИС (табл. 2).

Приведенный перечень показателей качества ИС может служить основой для разработки стандарта в данной области. Вместе с тем следует отметить, что разработка стандартов связана с длительными сроками и сложностью их разработки. Для организации разработки стандарта необходимо провести соответствующие исследования и обосновать заявку на выполнение работы, что требует участия подготовленных специалистов в соответствующей предметной области. В этом плане у отечественных предприятий и научно-исследовательских организаций существуют определенные трудности. Кроме того, разработка стандартов включает ряд стадий и занимает длительное время.

Таким образом, существует необходимость развития и совершенствования существующей нормативной базы в области требований и показателей качества функционирования ИС, а также методического аппарата, позволяющего производить оценку их качества на всех стадиях жизненного цикла.

Таблица 2 Показатели качества для типовой информационно-телекоммуникационной системы

Обобщенные показатели	Частные показатели
Надежность	Коэффициент готовности Среднее время наработки на отказ Среднее время восстановления Вероятность безотказной работы Вероятность представления/доведения информации
Своевременность	Среднее время доведения информации Среднее время доступа к данным Среднее время обработки информации Вероятность обработки информации за $T_{\text{зад}}$ Среднее время выполнения технологических операций
Полнота	Доля информации, представленной в требуемом объеме Доля реализованных решений Полнота контроля Вероятность оперативного отражения объектов и явлений
Достоверность	Вероятность безошибочной обработки Средняя наработка на ошибку Среднее время коррекции информации Коэффициент информационного технического использования
Конфиденциальность	Вероятность сохранения конфиденциальности информации в течение заданного периода Время вскрытия информации Вероятность навязывания ложного сообщения
Защищенность от несанкционированного доступа (НСД)	Вероятность сохранения защищенности от НСД Время успешной попытки НСД Вероятность преодоления механизмов защиты
Организационное обеспечение	Наличие, достаточность и укомплектованность эксплуатирующих подразделений Профессиональная подготовка персонала Морально-психологический фактор Эргономика

5 Методы оценки качества информационных систем

Согласно ГОСТ 15467-79, оценка уровня качества — это совокупность операций, включающая выбор номенклатуры показателей качества оцениваемой продукции, определение этих показателей и сопоставление их с базовыми.

Выделяются три основных метода оценки качества систем:

(1) дифференциальный;

(2) комплексный;

(3) смешанный.

Дифференциальный метод основан на использовании единичных показателей качества оцениваемого и базового образцов. При этом определяют, достигнут ли уровень базового образца в целом, по каким показателям он достигнут, какие показатели существенно отличаются от базовых.

Выбор номенклатуры единичных показателей для оценки качества системы проводится с учетом требований заказчика, условий разработки, производства и эксплуатации и т. д.

Расчет относительных показателей качества проводится по формуле:

$$Q_i = \frac{P_i}{P_{i\text{б}}},$$

где P_i — значение i-го показателя качества системы; $P_{i\text{б}}$ — значение i-го базового показателя; $i = 1, \ldots, n$ — число оцениваемых показателей качества.

Дифференциальный метод оценки уровня потребительских показателей качества может применяться на стадиях исследования и обоснования требований к ИС, разработки, производства и эксплуатации систем. Его достоинство в том, что исключается необходимость определения коэффициента весомости оцениваемого показателя качества, а недостатки — сравнительная форма фиксации значения оценки («лучше»–«хуже») и возможность суждения о качестве системы в целом лишь в тех случаях, когда значения всех единичных показателей качества оцениваемого товара выше или ниже соответствующих базовых значений показателей.

Комплексный метод оценки основан на применении обобщенного показателя качества, который представляет собой функцию от единичных показателей. При этом проводится сопоставление обобщенных показателей качества оцениваемого и базового образцов. Обобщенный показатель может быть средневзвешенным или интегральным.

Средневзвешенный показатель применяют, если нельзя установить функциональную зависимость обобщенного показателя от единичных показателей качества, но возможно с приемлемой точностью установить веса единичных показателей. Весовые коэффициенты определяются экспертным путем с использованием математических методов (метод ранжирования, метод приписывания баллов, метод парного сравнения и т. п.) [15].

Синтез полученных коэффициентов важности и определение показателя качества соответствующей системы осуществляется по формуле:

$$V_j = \sum_{i=1}^{N} w_i V_{ij},$$

где V_j — показатель качества j-й системы; w_i — весовой коэффициент i-го критерия; V_{ij} — коэффициент важности j-й системы по i-му критерию.

Интегральный показатель применяется тогда, когда можно установить суммарный полезный эффект от использования системы и суммарные затраты на ее создание и эксплуатацию. Интегральный показатель может быть рассчитан по следующей формуле:

$$\text{И} = \frac{\text{Э}}{\sum_{t=0}^{T} (\text{З}_{ct} + \text{З}_{\text{э}t})\, \text{К}_t},$$

где Э — суммарный полезный эффект от использования системы; З_{ct} — затраты на создание системы; $\text{З}_{\text{э}t}$ — затраты на эксплуатацию системы; К_t — коэффициент приведения разновременных затрат к одному году; T — нормативный срок службы.

Смешанный метод основан на одновременном использовании единичных и комплексных показателей качества. Он применяется, когда совокупность единичных показателей слишком велика и анализ каждого из них дифференциальным методом не позволяет сделать аргументированных выводов или когда обобщенный показатель при комплексном методе недостаточно полно учитывает все характеристики системы.

Следует отметить, что при любом из приведенных выше методов оценки качества основной и наиболее трудной задачей остается выбор перечня показателей качества. В [16–19] предложены подходы к выбору показателей эффективности и методам оценки эффективности ИС. Представляется, что эти подходы и методы могут быть использованы и для оценки качества ИС.

Опыт работ ФИЦ ИУ РАН по созданию и эксплуатации сложных крупномасштабных территориально распределенных ИС показывает, что при оценке их качества наиболее целесообразно использовать комплексный метод на основе средневзвешенного показателя качества. Для данного метода характерны простота формализации и возможность работы с большой размерностью данных, к тому же он достаточно апробирован. Кроме того, к достоинствам данного метода можно отнести ясный физический смысл, а также учет индивидуальных представлений лица, принимающего решение.

6 Заключение

В настоящее время сформировалось концептуальное видение качества как одной из фундаментальных категорий, определяющих образ жизни, социальную и экономическую основу для развития человека и общества.

Для оценки качества современная наука и практика выработали систему количественной оценки свойств систем (объектов), так называемых моделей качества, которая основывается на использовании показателей качества. При этом под уровнем качества систем (объектов) понимается относительная характеристика качества, представляющая собой результат сравнения совокупности значений показателей качества систем (объектов) с соответствующей совокупностью ба-

зовых показателей. Определение уровня качества производится с помощью специальных методов оценки.

Следует отметить, что существующие модели оценки в основном ориентированы на оценку качества программных средств, в то время как нормативная база оценки качества ИС в целом развита недостаточно.

В данной статье рассмотрены общие вопросы, касающиеся качества ИС, и наиболее распространенные модели качества ИС, а также методы оценки их качества. Кроме того, представлены предложения по совершенствованию нормативной базы в данной области.

Литература

1. ГОСТ 15467-79. Управление качеством продукции. Основные понятия. Термины и определения. — М.: Стандартинформ, 2009. 21 с.
2. ИСО 8402-94. Управление качеством и обеспечение качества: Словарь. http://stroysvoimirukami.ru/iso-8402-94.
3. ГОСТ Р ИСО 9000-2015. Системы менеджмента качества. Основные положения и словарь. — М.: Стандартинформ, 2019. 53 с.
4. *McCall J., Richards P., Walters G.* Factors in software quality. — Rome, NY, USA: Rome Air Development Center, 1977. Report NTIS AD-A049-014, 015, 055. 3 vols.
5. *Боэм Б., Браун Дж., Каспар Х. и др.* Характеристики качества программного обеспечения / Пер. с англ. — М.: Мир, 1981. 208 с. (*Boehm B. W., Brown J. R., Kaspar H., Lipow M., Macleod G. J., Merrit M. J.* Characteristics of software quality. — Amsterdam: North-Holland, 1978. 216 p.)
6. *Gilb T.* Principles of software engineering management. — Reading, MA, USA: Addison Wesley, 1988. 464 p.
7. ГОСТ 28195-89. Оценка качества программных средств. Общие положения. — М.: Стандартинформ, 2001. 31 с.
8. ГОСТ ИСО/МЭК 9126-2001. Информационная технология. Оценка программной продукции. Характеристики качества и руководства по их применению. — М.: Стандартинформ, 2013. 13 с.
9. ГОСТ Р ИСО/МЭК 25010-2015. Информационная технология. Системная и программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Модели качества систем и программных продуктов. — М.: Стандартинформ, 2015. 36 с.
10. ГОСТ РВ 51987-2002. Информационная технология. Комплекс стандартов на автоматизированные системы. Требования и показатели качества функционирования информационных систем. — М.: Стандартинформ, 2003. 52 с.
11. Методическое руководство по оценке качества функционирования информационных систем (в контексте стандарта ГОСТ РВ 51987 «Информационная технология. Комплекс стандартов на автоматизированные системы. Требования и показатели качества функционирования информационных систем. Общие положения»). — М.: 3 ЦНИИ МО РФ, 2003. 352 с.
12. *Костогрызов А. И., Степанов П. В.* Инновационное управление качеством и рисками в жизненном цикле систем. — М.: ВПК, 2008. 404 с.

13. *Ионенков Ю. С.* Научно-практические аспекты оценки эффективности информационно-телекоммуникационных систем // Радиолокация, навигация, связь: Сб. трудов XXIV Междунар. научн.-технич. конф. — Воронеж: Вэлберн, 2018. Т. 1. С. 140–149.
14. *Зацаринный А. А., Ионенков Ю. С.* Оценка эффективности информационно-телекоммуникационных систем / Под ред. А. А. Зацаринного. — М.: НИПКЦ Восход-А, 2020. 120 с.
15. *Ларичев О. И.* Теория и методы принятия решений. — М.: Логос, 2007. 392 с.
16. *Зацаринный А. А., Ионенков Ю. С.* К вопросу оценки эффективности автоматизированных систем с использованием метода анализа иерархий // Системы и средства информатики, 2015. Т. 25. № 3. С. 162–179.
17. *Зацаринный А. А., Ионенков Ю. С., Сучков А. П.* Некоторые аспекты оценки эффективности облачных технологий // Системы и средства информатики, 2018. Т. 28. № 3. С. 104–117.
18. *Зацаринный А. А., Ионенков Ю. С.* Некоторые методические аспекты выбора показателей эффективности информационных систем // Системы высокой доступности, 2019. № 4. С. 19–26.
19. *Зацаринный А. А., Волович К. И., Денисов С. А., Ионенков Ю. С., Кондрашев В. А.* Вопросы выбора показателей эффективности функционирования высокопроизводительного вычислительного комплекса на примере ЦКП «Информатика» ФИЦ ИУ РАН // Известия высших учебных заведений. Материалы электронной техники, 2020. Т. 23. Вып. 3. С. 241–247.

Поступила в редакцию 26.04.21

SOME ISSUES OF INFORMATION SYSTEM QUALITY ASSESSMENT

A. A. Zatsarinny and Yu. S. Ionenkov

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The article is devoted to the issue of assessing the quality of information systems (IS). The concepts of product quality in general and IS quality in particular are presented. The differences between the concepts of “quality” and “efficiency” are considered. The most common currently foreign and domestic models of IS quality are presented. It is noted that most of them relate to the assessment of the quality of software, while the regulatory framework for assessing the quality of IS as a whole is not sufficiently developed. The main methods of assessing the quality of IS are considered. The proposals for improving the regulatory framework in the field of IS quality assessment are formulated.

Keywords: quality; quality assessment; quality assessment method; indicator; information system

DOI: 10.14357/08696527210401

References

1. GOST 15467-79. 2009. *Upravlenie kachestvom produktsii. Osnovnye ponyatiya. Terminy i opredeleniya* [Product quality control. Basic concepts. Terms and definitions]. Moscow: Standardinform Publs. 21 p.
2. GOST ISO 8402-94. 1994. Upravlenie kachestvom i obespechenie kachestva: Slovar' [Quality management and quality assurance: Glossary]. Available at: http://stroysvoimirukami.ru/iso-8402-94/ (assessed October 2, 2021).
3. GOST R ISO 9000-2015. 2019. Sistemy menedzhmenta kachestva. Osnovnye polozheniya i slovar' [Quality management systems. Fundamentals and vocabulary]. Moscow: Standardinform Publs. 53 p.
4. McCall, J., P. Richards, and G. Walters. 1977. Factors in software quality. Rome, NY, USA: Rome Air Development Center. Report NTIS AD-A049-014, 015, 055. 3 vols.
5. Boehm, B. W., J. R. Brown, H. Kaspar, M. Lipow, G. J. Macleod, and M. J. Merrit. 1978. *Characteristics of software quality*. Amsterdam: North-Holland. 216 p.
6. Gilb, T. 1988. *Principles of software engineering management*. Reading, MA: Addison Wesley. 464 p.
7. GOST 28195-89. 2001. Otsenka kachestva programmnykh sredstv. Obshchie polozheniya [Assessment of the quality of software. General provisions]. Moscow: Standardinform Publs. 31 p.
8. GOST ISO/IEC 9126-2001. 2013. Informatsionnaya tekhnologiya. Otsenka programmnoy produktsii. Kharakteristiki kachestva i rukovodstva po ikh primeneniyu [Information technology. Software product evaluation. Quality characteristics and guidelines for their use]. Moscow: Standardinform Publs. 13 p.
9. GOST R ISO/IEC 25010-2015. 2015. Informatsionnaya tekhnologiya. Sistemnaya i programmnaya inzheneriya. Trebovaniya i otsenka kachestva sistem i programmnogo obespecheniya (SQuaRE). Modeli kachestva sistem i programmnykh produktov [Information technology. Systems and software engineering. Systems and software quality requirements and evaluation (SQuaRE). System and software quality models]. Moscow: Standardinform Publs. 36 p.
10. GOST RV 51987-2002. 2002. Informatsionnaya tekhnologiya. Kompleks standartov na avtomatizirovannye sistemy. Trebovaniya i pokazateli kachestva funktsionirovaniya informatsionnykh sistem [Information technology. Set of standards for automated systems. Requirements and quality indicators of information systems functioning]. Moscow: Standardinform Publs. 52 p.
11. Metodicheskoe rukovodstvo po otsenke kachestva funktsionirovaniya informatsionnykh system (v kontekste standarta GOST RV 51987 "Informatsionnaya tekhnologiya. Kompleks standartov na avtomatizirovannye sistemy. Trebovaniya i pokazateli kachestva funktsionirovaniya informatsionnykh sistem. Obshchie polozheniya") [Methodological guidelines for assessing the quality of functioning of information systems (in the context of GOST RV 51987 "Information technology. A set of standards for automated systems. Requirements and quality indicators for information systems functioning. General provisions")]. Moscow: 3 TsNII MO RF. 352 p.
12. Kostogryzov, A. I., and P. V. Stepanov. 2008. *Innovatsionnoe upravlenie kachestvom i riskami v zhiznennom tsikle sistem* [Innovative quality and risk management in the system lifecycle]. Moscow: VPK. 404 p.

13. Ionenkov, Yu. S. 2018. Nauchno-prakticheskie aspekty otsenki effektivnosti informatsionno-telekommunikatsionnykh sistem [Scientific and practical aspects of evaluating the effectiveness of information and telecommunication systems]. *Sb. trudov XXIV Mezhdunar. nauchn.-tekhnich. konf. "Radiolokatsiya, navigatsiya, svyaz"* [XXIV Scientific and Technical Conference (International) "Radar, Navigation, Communication" Proceedings]. Voronezh: Velbern. 1:140–149.
14. Zatsarinnyy, A. A., and Yu. S. Ionenkov. 2020. *Otsenka effektivnosti informatsionno-telekommunikatsionnykh sistem* [Evaluation of the effectiveness of information and telecommunication systems]. Moscow: NIPKTs Voskhod-A. 120 p.
15. Larichev, O. I. 2007. *Teoriya i metody prinyatiya resheniy* [Decision theory and methods]. Moscow: Logos. 392 p.
16. Zatsarinnyy, A. A., and Yu. S. Ionenkov. 2015. K voprosu otsenki effektivnosti avtomatizirovannykh sistem s ispol'zovaniem metoda analiza ierarkhiy [On the issue of assessing the effectiveness of automated systems using the method of analysis of hierarchies]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(3):162–179.
17. Zatsarinny, A. A., Yu. S. Ionenkov, and A. P. Suchkov. 2018. Nekotorye aspekty otsenki effektivnosti oblachnykh tekhnologiy [Some aspects of cloud computing efficiency estimation]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(3):104–117.
18. Zatsarinny, A. A., and Y. S. Ionenkov. 2019. Nekotorye metodicheskie aspekty vybora pokazateley effektivnosti informatsionnykh sistem [Some methodological aspects of the choice of performance indicators of information systems]. *Sistemy vysokoy dostupnosti* [Highly Available Systems] 15(4):19–26.
19. Zatsarinny, A. A., K. I. Volovich, S. A. Denisov, Y. S. Ionenkov, and V. A. Kondrashev. 2020. Voprosy vybora pokazateley effektivnosti funktsionirovaniya vysokoproizvoditel'nogo vychislitel'nogo kompleksa na primere TsKP "Informatika" FITs IU RAN [Choice of HPC cluster performance indicators for the example of the "Informatika" Center for Collective Use of the FRC CSC RAS]. *Proceedings of Higher Schools. Materials of Electronics Engineering* 23(3):241–247.

Received April 26, 2021

Contributors

Zatsarinny Alexander A. (b. 1951) — Doctor of Science in technology, professor, principal scientist, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; AZatsarinny@ipiran.ru

Ionenkov Yuriy S. (b. 1956) — Candidate of Science (PhD) in technology, senior scientist, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; uionenkov@ipiran.ru

ПРОГРАММА ПОСТРОЕНИЯ ВПОЛНЕ ИНТЕРПРЕТИРУЕМЫХ И RTF-АДЕКВАТНЫХ ЛИНЕЙНЫХ РЕГРЕССИОННЫХ МОДЕЛЕЙ

М. П. Базилевский[1]

Аннотация: Статья посвящена проблеме отбора «информативных» регрессоров (ОИР) в регрессионных моделях, оцениваемых с помощью метода наименьших квадратов (МНК). Построенные в результате такого отбора модели часто оказываются неадекватными и плохо интерпретируемыми. В работе впервые сформулированы определения «вполне интерпретируемой» и «RTF-адекватной» регрессионной модели. Рассмотрен ранее предложенный эффективный алгоритм решения задачи ОИР. На его основе разработан алгоритм построения вполне интерпретируемых и RTF-адекватных линейных регрессионных моделей. В нем для каждой регрессии последовательно осуществляется проверка: «информативности» переменных, мультиколлинеарности, соответствия знаков коэффициентов физическому смыслу факторов, адекватности модели по коэффициенту детерминации и значимости в целом по F-критерию Фишера, а также значимости коэффициентов по t-критерию Стьюдента. Предложенный алгоритм реализован в виде программы для эконометрического пакета Gretl. Разработанная программа универсальна и может быть использована для решения широкого круга задач анализа данных.

Ключевые слова: отбор «информативных» регрессоров; метод наименьших квадратов; вполне интерпретируемая и RTF-адекватная регрессия; критерий «информативности» переменных; мультиколлинеарность; F-критерий Фишера; t-критерий Стьюдента

DOI: 10.14357/08696527210402

1 Введение

Один из этапов регрессионного анализа [1] — спецификация модели, т. е. выбор состава переменных и математической формы связи между ними. Определение спецификации регрессии на практике зачастую сводится к решению задачи ОИР, более известной в зарубежной литературе как «subset selection», «variable selection» или «feature selection» in regression [2, 3]. В отечественной литературе описание методов ОИР можно найти в работе [4]. Решение задачи ОИР может осуществляться не только по какому-то одному критерию адекватности, а одновременно по нескольким критериям сразу. Такая технология многокритериального выбора получила название «конкурс» моделей [5, 6].

[1]Иркутский государственный университет путей сообщения, кафедра математики, mik2178@yandex.ru

В работе [7] автором был проведен анализ и выявлены следующие недостатки программного комплекса автоматизации процесса построения регрессионных моделей (ПК АППРМ), предназначенного для проведения «конкурса» моделей.

1. Полученные в результате «конкурса» модели часто представляют собой сложные нелинейные уравнения, коэффициенты которых весьма затруднительно интерпретировать. При этом игнорируется проблема мультиколлинеарности, которая может лишать смысла интерпретацию коэффициентов модели.
2. Полученные в результате «конкурса» модели часто оказываются неадекватными по коэффициенту детерминации R^2, незначимыми по F-критерию Фишера, а их оценки — незначимыми по t-критерию Стьюдента.

На основе выявленных недостатков в работе [7] было введено понятие «хорошо интерпретируемая качественная регрессия», удовлетворяющая пяти условиям, и разработан фундаментальный блок алгоритмов построения таких моделей.

Цель данной работы — создание конкретного алгоритма построения хорошо интерпретируемых и качественных линейных регрессий, а также его реализация в виде программы.

2 Вполне интерпретируемые и RTF-адекватные регрессионные модели

Размышления над предложенным в [7] понятием «хорошо интерпретируемая качественная модель» привели к тому, что целесообразнее заменить этот термин следующими двумя определениями.

Определение 1. Регрессионная модель считается вполне интерпретируемой, если она удовлетворяет трем условиям:

(1) ее спецификация изначально выбрана так, что после оценивания модели можно объяснить любой ее коэффициент или некоторый его аналог, за исключением, быть может, свободного члена;

(2) знаки коэффициентов оцененной модели соответствуют физическому смыслу входящих в уравнение факторов;

(3) эффект мультиколлинеарности незначителен.

Определение 2. Регрессионная модель, оцененная с помощью МНК, считается RTF-адекватной, если она удовлетворяет трем условиям:

(1) модель адекватна по коэффициенту детерминации R^2;

(2) коэффициенты модели значимы по t-критериям Стьюдента;

(3) модель значима в целом по F-критерию Фишера.

3 Эффективный алгоритм решения задачи отбора «информативных» регрессоров

Рассмотрим модель множественной линейной регрессии

$$y_i = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \cdots + \alpha_l x_{il} + \varepsilon_i\,, \quad i = \overline{1,n}\,, \tag{1}$$

где y_i, $i = \overline{1,n}$, — значения зависимой (объясняемой) переменной y; $x_{i1}, x_{i2}, \ldots, x_{il}$, $i = \overline{1,n}$, — значения l независимых (объясняющих) переменных (регрессоров) $x_1, x_2, \ldots, x_l$; ε_i, $i = \overline{1,n}$, — ошибки аппроксимации; $\alpha_0, \alpha_1, \ldots, \alpha_l$ — неизвестные параметры; n — объем выборки.

Задача отбора «информативных» регрессоров. Необходимо выделить из l возможных регрессоров m переменных, минимизируя сумму квадратов ошибок для регрессии (1).

В работах [7, 8] рассмотрен эффективный алгоритм «Selection B» решения поставленной задачи ОИР. В его основе лежит следующая техника оценивания регрессионной модели с помощью МНК. Предварительно проводится нормирование (стандартизация) всех переменных по формулам:

$$v_i = \frac{y_i - \overline{y}}{\sigma_y}\,; \quad z_{i1} = \frac{x_{i1} - \overline{x}_1}{\sigma_{x_1}}\,; \ldots; z_{il} = \frac{x_{il} - \overline{x}_l}{\sigma_{x_l}}\,,$$

где $\overline{y}, \overline{x}_1, \ldots, \overline{x}_l$ — средние значения переменных; $\sigma_y, \sigma_{x_1}, \ldots, \sigma_{x_l}$ — среднеквадратичные отклонения переменных; $v, z_1, \ldots, z_l$ — стандартизованные переменные, для которых среднее значение равно 0, а среднеквадратичное отклонение равно 1.

Тогда регрессии (1) ставится в соответствие ее стандартизованная модель:

$$v_i = \beta_1 z_{i1} + \beta_2 z_{i2} + \cdots + \beta_l z_{il} + u_i\,, \quad i = \overline{1,n}\,, \tag{2}$$

где $\beta_1, \ldots, \beta_l$ — неизвестные параметры (бета-коэффициенты); u_i, $i = \overline{1,n}$, — ошибки аппроксимации.

Оценки бета-коэффициентов находятся по формуле:

$$\tilde{\beta}_{\text{МНК}} = r_{xx}^{-1} r_{yx}\,,$$

где r_{xx} — матрица коэффициентов парной корреляции между объясняющими переменными; r_{xy} — вектор коэффициентов парной корреляции между объясняемой переменной y и объясняющими переменными $x_1, x_2, \ldots, x_l$.

Коэффициент детерминации для регрессий (1) и (2) находится по формуле:

$$R^2 = r_{yx}^{\text{T}} \tilde{\beta}_{\text{МНК}}\,. \tag{3}$$

Алгоритм «Selection B» предполагает предварительное задание пользователем следующих параметров:

1. Вектор знаков $\mathrm{Sign}_{1\times l}$. Компоненты этого вектора назначаются по следующему правилу:

$$\mathrm{Sign}_{1,j} = \begin{cases} 1\,, & \text{если } j\text{-я переменная оказывает положительное влияние на } y; \\ -1\,, & \text{если } j\text{-я переменная оказывает отрицательное влияние на } y; \\ 0\,, & \text{если затруднительно оценить влияние } j\text{-й переменной на } y. \end{cases}$$

   Для формирования вектора Sign желательно привлекать экспертов.

2. Число отбираемых регрессоров m.

3. Граница мультиколлинеарности d.

В алгоритме «Selection B» сначала определяется общее число альтернативных вариантов моделей $r = C_l^m$, после чего запускается основной цикл алгоритма. На каждом повторении формируются необходимые матрицы r_{xx} и r_{yx}, а затем последовательно проверяются условия:

$$\left|r_{xx}^J\right| < d\,; \tag{4}$$

$$\tilde{\beta}_j \mathrm{Sign}_{1,j} < 0\,, \quad j \in J\,, \tag{5}$$

где J — множество номеров объясняющих переменных, входящих в регрессию; r_{xx}^J — матрица парных коэффициентов корреляции для объясняющих переменных с номерами из множества J.

В неравенстве (4) определитель матрицы парных коэффициентов корреляции $|r_{xx}^J|$ представляет собой критерий обнаружения эффекта мультиколлинеарности. Если $|r_{xx}^J| = 0$, то имеет место совершенная мультиколлинеарность, а если $|r_{xx}^J| = 1$, то мультиколлинеарности нет. Если условие (4) выполняется, то такая регрессия сразу исключается из рассмотрения.

Если выполнено хотя бы одно из условий (5), то это означает, что знак бета-коэффициента $\tilde{\beta}_j$ стандартизованной регрессии не согласуется с соответствующим знаком вектора Sign. Следовательно, такая модель исключается из дальнейшего «конкурса».

Если очередная альтернативная регрессия прошла «испытание» условиями (4) и (5), то для нее по формуле (3) рассчитывается величина коэффициента детерминации. После окончания работы основного цикла определяется наилучшая модель по величине R^2.

Проведенный в работе [8] эксперимент показал, что решение задачи ОИР с использование пакета Gretl по алгоритму «Selection B» осуществляется более чем в 6 раз быстрее, нежели в ПК АППРМ.

4 Алгоритм построения вполне интерпретируемых и RTF-адекватных линейных регрессий

Алгоритм построения вполне интерпретируемых и RTF-адекватных линейных регрессий (см. рисунок) был разработан на основе фундаментального блока, предложенного в работе [7]. Согласно этому блоку, для того чтобы очередная альтернативная регрессионная модель была допущена к участию в «конкурсе», она должна пройти жесткую процедуру отсева, состоящую из 5 стадий:

(1) проверка «информативности» переменных;

(2) проверка присутствия эффекта мультиколлинеарности;

(3) проверка соответствия знаков коэффициентов физическому смыслу факторов;

(4) проверка адекватности по коэффициенту детерминации и значимости в целом по F-критерию Фишера;

(5) проверка значимости коэффициентов по t-критерию Стьюдента.

Первая стадия связана с вычислением критерия «информативности» переменных [6]:

$$\Gamma\left(x_1, x_2, \ldots, x_l\right) = \sum_{j \in J} \mathrm{Inf}\left(x_j\right),$$

где $\mathrm{Inf}\left(x_j\right)$ — некоторый вес (балл) регрессора x_j, например в пятибалльной шкале. Таким образом, этот критерий указывает на суммарную «важность» вхождения всех независимых переменных в зависимость.

На четвертой стадии для обеспечения значимости регрессии в целом по F-критерию Фишера значение коэффициента детерминации R^2 не должно быть меньше величины

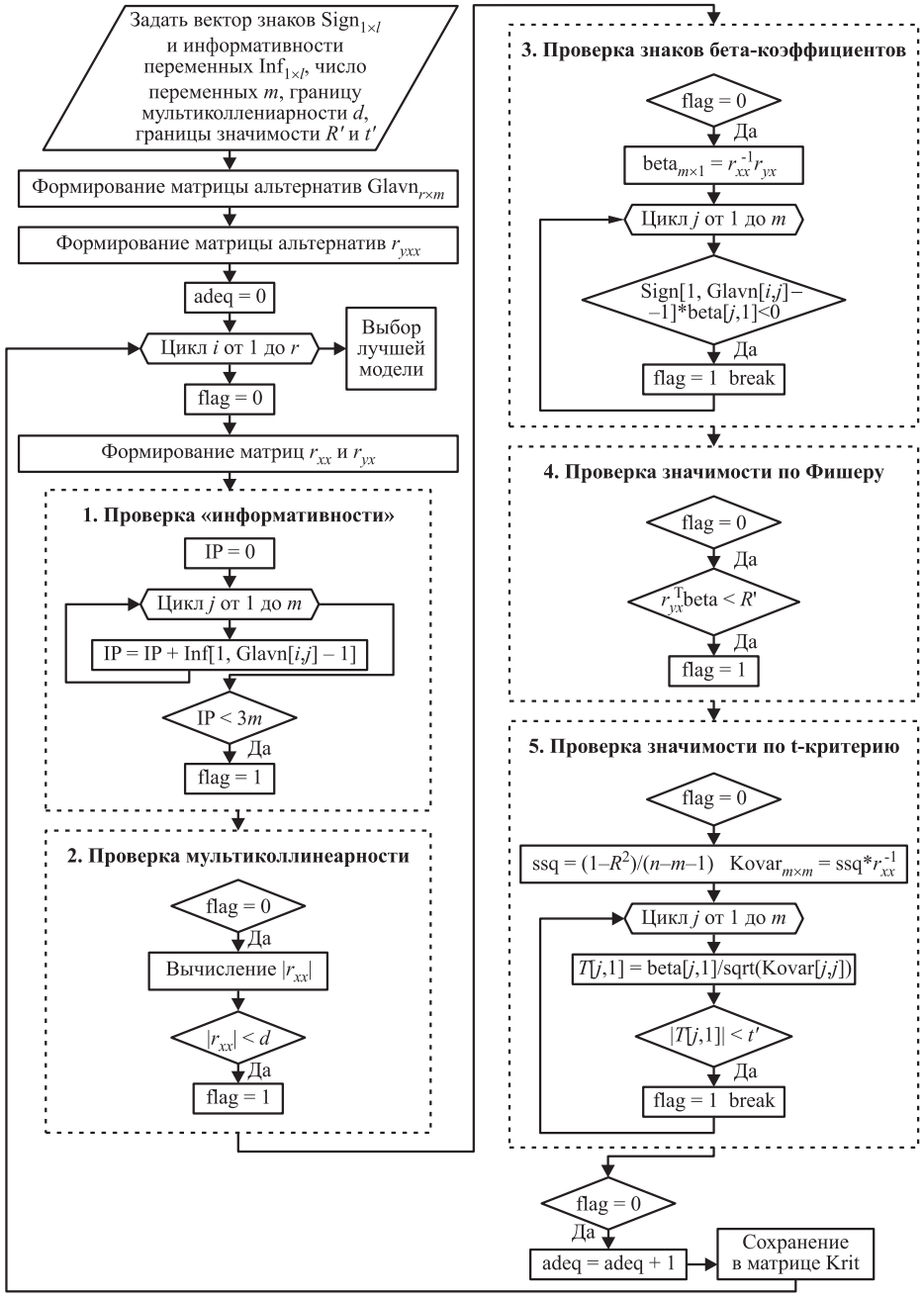
$$R' = \frac{F_{\text{крит}}(\alpha, m, n-m-1)}{(n-m-1)/m + F_{\text{крит}}(\alpha, m, n-m-1)},$$

где α — заданный уровень значимости; $F_{\text{крит}}(\alpha, m, n-m-1)$ — критическая точка распределения Фишера.

На пятой стадии вычисление t-критериев Стюдента для регрессий (1) и (2) реализуется по формулам:

$$t_j = \frac{\tilde{\beta}_j}{\sqrt{s^2 (r_{xx}^J)_{jj}^{-1}}}, \quad j = \overline{1, m},$$

где $s^2 = (1 - R^2)/(n - m - 1)$ — величина остаточной дисперсии; $(r_{xx}^J)_{jj}^{-1}$ — j-й диагональный элемент матрицы $(r_{xx}^J)^{-1}$.



Алгоритм построения линейных регрессий

Пользователю предварительно необходимо задать следующие параметры:

1. Вектор знаков $\mathrm{Sign}_{1\times l}$.
2. Вектор «информативности» переменных $\mathrm{Inf}_{1\times l}$. Компоненты этого вектора представляют собой целые числа из интервала $[1, 5]$. Если $\mathrm{Inf}_{1\times j} = 5$, то «информативность» j-й переменной максимальна и составляет 5 баллов, а если $\mathrm{Inf}_{1\times j} = 1$, то «информативность» минимальна и составляет 1 балл. Для формирования этого вектора также желательно привлекать экспертов.
3. Число отбираемых регрессоров m.
4. Границу мультиколлинеарности d.
5. Границу значимости для F-критерия Фишера R'.
6. Границу значимости для t-критериев Стьюдента t'.

На рисунке $\mathrm{Glavn}_{r\times m}$ — матрица альтернатив, в которой объясняемая переменная имеет номер 1, а объясняющие переменные — $2, 3, \ldots, l+1$; adeq — счетчик вполне интерпретируемых и RTF-адекватных моделей; flag — бинарная переменная, которая отвечает за выполнение всех пяти требований, предъявляемым к моделям; Krit — критериальная матрица, в первом столбце которой содержатся значения коэффициента детерминации R^2, во втором — определителя $|r_{xx}|$, в третьем — критерия «информативности» переменных, в остальных — номера объясняющих переменных.

Предложенный алгоритм был реализован в виде программы с использованием эконометрического пакета Gretl.

5 Заключение

Разработанная программа построения вполне интерпретируемых и RTF-адекватных линейных регрессионных моделей позволяет находить среди многочисленных альтернатив по-настоящему качественные регрессии, значимые в целом по F-критерию Фишера, в которых все знаки коэффициентов соответствуют физическому смыслу факторов, незначителен эффект мультиколлинеарности и все коэффициенты значимы по t-критерию Стьюдента. Кроме того, выбор состава переменных приобретает определенную сознательность, поскольку в процесс построения модели привлекаются эксперты в исследуемой предметной области.

Литература

1. *Pardoe I.* Applied regression modeling. — Hoboken, NJ, USA: Wiley, 2020. 336 p.
2. *Miller A. J.* Subset selection in regression. — London, U.K.: Chapman & Hall/CRC, 2002. 256 p.
3. *Venkatesh B., Anuradha J.* A review of feature selection and its methods // Cybernetics Information Technologies, 2019. Vol. 19. P. 3–26.

4. *Стрижов В. В., Крымова Е. А.* Методы выбора регрессионных моделей. — М.: ВЦ РАН, 2010. 60 с.
5. *Носков С. И., Потороченко Н. А.* Диалоговая система реализации «конкурса» регрессионных зависимостей // Управляющие системы и машины, 1992. № 3-4. С. 111–116.
6. *Носков С. И.* Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных. — Иркутск: Облинформпечать, 1996. 321 с.
7. *Базилевский М. П.* Фундаментальный блок алгоритмов построения хорошо интерпретируемых качественных регрессионных моделей // Информационные технологии и математическое моделирование в управлении сложными системами, 2020. № 3(8). С. 1–10.
8. *Базилевский М. П.* Повышение эффективности алгоритма отбора по критерию детерминации информативных регрессоров в регрессионных моделях // Прикладная математика и информатика: современные исследования в области естественных и технических наук. — Тольятти, 2018. С. 196–202.

Поступила в редакцию 13.01.21

A PROGRAM FOR CONSTRUCTING OF QUITE INTERPRETABLE AND RTF-ADEQUATE LINEAR REGRESSION MODELS

M. P. Bazilevskiy

Department of Mathematics, Irkutsk State Transport University, 15 Chernyshevskogo Str., Irkutsk 664074, Russian Federation

Abstract: The article is devoted to the problem of feature selection in regression models estimated using the ordinary least squares method. Models constructed as a result of such selection are often inadequate and poorly interpreted. For the first time, the definitions of "quite interpretable" and "RTF-adequate" regression models are formulated. The previously proposed effective algorithm for solving the problem of feature selection is considered. On its basis, an algorithm has been developed for constructing quite interpretable and RTF-adequate linear regression models. In it, for each regression, the following tests are sequentially carried out: "informativeness" of variables, multicollinearity, correspondence of coefficients signs to the physical meaning of factors, adequacy of model in terms of coefficient of determination and significance in general according to Fisher's F-test, and significance of the coefficients according to the Student's t-test. The proposed algorithm is implemented as a program for the Gretl econometric package. The developed program is universal and can be used to solve a wide range of data analysis tasks.

Keywords: feature selection; ordinary least squares; quite interpretable and RTF-adequate regression; variable "informativeness" criterion; multicollinearity; Fisher's F-test; Student's t-test

DOI: 10.14357/08696527210402

References

1. Pardoe, I. 2020. *Applied regression modeling*. Hoboken, NJ, USA: Wiley. 336 p.
2. Miller, A. J. 2002. *Subset selection in regression*. London, U.K.: Chapman & Hall/CRC. 256 p.
3. Venkatesh, B., and J. Anuradha. 2019. A review of feature selection and its methods. *Cybernetics Information Technologies* 19:3–26.
4. Strizhov, V. V., and E. A. Krymova. 2010. *Metody vybora regressionnykh modeley* [Regression model selection methods]. Moscow: CC RAS. 60 p.
5. Noskov, S. I., and N. A. Potorochenko. 1992. Dialogovaya sistema realizatsii konkursa regressionnykh zavisimostey [Dialogue system for the implementation of the competition of regression dependencies]. *Upravlyayushchie sistemy i mashiny* [Control Systems and Computers] 3-4:111–116.
6. Noskov, S. I. 1996. *Tekhnologiya modelirovaniya ob"ektov s nestabil'nym funktsionirovaniem i neopredelennost'yu v dannykh* [Technology for modeling objects with unstable functioning and uncertainty in data]. Irkutsk: Oblinformpechat'. 321 p.
7. Bazilevskiy, M. P. 2020. Fundamental'nyy blok algoritmov postroeniya khorosho interpretiruemykh kachestvennykh regressionnykh modeley [The fundamental block of algorithms for constructing well-interpreted qualitative regression models]. *Informatsionnye tekhnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami* [Information Technology and Mathematical Modeling in the Management of Complex Systems] 3(8):1–10.
8. Bazilevskiy, M. P. 2018. Povyshenie effektivnosti algoritma otbora po kriteriyu determinatsii informativnykh regressorov v regressionnykh modelyakh [Improving the efficiency of the algorithm for selecting informative regressors by the criterion of determination in regression models]. *Prikladnaya matematika i informatika: sovremennye issledovaniya v oblasti estestvennykh i tekhnicheskikh nauk* [Applied Mathematics and Informatics: Contemporary Research in Natural and Technical Sciences]. Tol'yatti. 196–202.

Received January 13, 2021

Contributor

Bazilevskiy Mikhail P. (b. 1987) — Candidate of Science (PhD) in technology, associate professor, Irkutsk State Transport University, 15 Chernyshevkogo Str., Irkutsk 664074, Russian Federation; mik2178@yandex.ru

РАСПРЕДЕЛЕНИЯ СТАТИСТИК ОТНОШЕНИЯ ПРАВДОПОДОБИЯ ДЛЯ ВЫЯВЛЕНИЯ МОНОТОННОГО ТРЕНДА

М. П. Кривенко[1]

Аннотация: Рассматривается задача анализа монотонного тренда. За основу принимается модель изменения среднего нормального распределения, для которой распределения статистик отношения правдоподобия суть смеси χ^2- или β-распределений с весами, определяемыми через числа Стирлинга первого рода. Исследуются вопросы аппроксимации этих распределений. Предложен эффективный метод расчета критериев значимости путем отбрасывания незначимых элементов смеси, который позволяет на несколько порядков снизить сложность вычислений. Обсуждаются вопросы расширения критериев анализа тренда среднего нормального распределения до процедур выявления стохастической упорядоченности. Для этих целей применим критерий монотонного тренда для усредненных рангов, статистика которого при отсутствии изменений распределена как смесь χ^2-распределений. Кардинальное значение имеет использование ступенчатой структуры оценки: удалось не только получить оригинальные результаты, но и обеспечить жизнеспособность предлагаемых решений.

Ключевые слова: монотонный тренд; критерий отношения правдоподобия; непараметрическое обнаружение тренда; асимптотическое распределение; аппроксимация распределения вероятностей

DOI: 10.14357/08696527210403

1 Введение

Рассмотрим задачу выявления изменений характеристик исследуемого процесса. Наиболее привлекательным в такой ситуации оказывается формирование модели данных с использованием по возможности меньшего объема априорных предположений. Причина этого состоит в том, что подчас нет достаточно полной информации для построения простых адекватных практике теоретико-вероятностных схем, а любое сужение класса моделей увеличивает сомнения в устойчивости получаемых процедур. Подобная постановка задачи востребована, например, при лонгитюдном анализе данных в медицине [1], для мониторинга процесса формирования и оценки команд в социологии [2], в случае отслеживания экстремальных климатических явлений [3], при управлении доходами в эконометрике [4] и др.

[1]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, mkrivenko@ipiran.ru

Основу модели данных составляет предположение о том, что изменение их характеристик есть монотонная функция неизвестного вида от номера наблюдения (далее — монотонный тренд).

В работе рассматривается случай нормального распределения данных и неубывающий тренд единственного параметра — среднего. Если $f(x_j, \mu_j)$ — плотность нормального распределения для j-го наблюдения x_j и соответствующего среднего μ_j, $\lambda(\boldsymbol{\mu}, \mathbf{x}) = \prod_{j=1}^{N} f(x_j, \mu_j)$ — функция правдоподобия, то центральной становится задача нахождения такой оценки максимального правдоподобия (о. м. п.) монотонного тренда $\boldsymbol{\mu}^* = (\mu_1^*, \ldots, \mu_N^*)$, что

$$\lambda\left(\boldsymbol{\mu}^*, \mathbf{x}\right) = \sup_{\mu_1 \le \mu_2 \le \cdots \le \mu_N} \lambda(\boldsymbol{\mu}, \mathbf{x})\,.$$

Переход к случаю непараметрического тренда целесообразно осуществить так, чтобы расширить область применения уже построенных решений. Здесь наиболее практичным оказывается использование свободных от распределения процедур, заключающееся в переходе от наблюдений к их рангам.

Задача оценивания монотонного тренда впервые была поставлена в работах середины прошлого столетия, где содержалось доказательство существования и единственности о. м. п. монотонного тренда одномерного параметра. Предположения, положенные в основу нахождения оценки тренда и исследования ее выборочных свойств в последующих работах (см., например, [5]), в основном соответствовали тогдашним запросам практики. Позднее их удалось обобщить и добиться более значимых результатов при описании выборочных свойств о. м. п. монотонного тренда и построении критериев значимости [6, 7].

Для получения оценки используется PAV-процедура (PAV — Pool-Adjacent-Violators) устранения неверных неравенств. Обозначим через $\mu^*(s,t)$ о. м. п. параметра распределения при отсутствии изменений, построенную по наблюдениям $x_s, \ldots, x_t$, где $1 \le s \le t \le N$ (в случае нормально распределенных наблюдений $\mu^*(s,t)$ — обычное выборочное среднее для $x_s, \ldots, x_t$). Основной шаг PAV-процедуры заключается в объединении соседних групп наблюдений, которые дают оценки $\mu^*(s,j)$ и $\mu^*(j+1,t)$, не удовлетворяющие требуемым условиям.

Оценку монотонного тренда целесообразно представлять в виде

$$\boldsymbol{\mu}^* = \left(\underbrace{h_1, \ldots, h_1}_{m_1}, \ldots, \underbrace{h_l, \ldots, h_l}_{m_l} \right). \tag{1}$$

Здесь $h_j < h_{j+1}$, $j = 1, \ldots l-1$ и $m_1 + \cdots + m_l = N$. Назовем группу одинаковых значений элементов оценки ступенькой. Тогда можно говорить о числе ступенек l, а для j-й ступеньки — о ее длине m_j и высоте (абсолютной) h_j.

Представительный подкласс параметрических процедур проверки гипотезы об однородности выборки против конкурирующей гипотезы о монотонной

упорядоченности составляют критерии отношения правдоподобия (к. о. п.). Их статистики в зависимости от априорных предположений принимают тот или иной вид.

Если значения математического ожидания μ_0 и дисперсии σ^2 при нулевой гипотезе известны, то, используя в качестве промежуточного представление (1), получаем:

$$-2\sigma^2 \ln \frac{\max_{\mu_1=\cdots=\mu_N} \lambda(\boldsymbol{\mu}, \mathbf{x})}{\max_{\mu_1\le\cdots\le\mu_N} \lambda(\boldsymbol{\mu}, \mathbf{x})} = \sum_{j=1}^{N} \left(\mu_j^* - \mu_0\right)^2 .$$

Тогда в качестве статистики к. о. п. можно принять

$$\overline{\chi}_0^2 = \frac{1}{\sigma^2} \sum_{j=1}^{N} \left(\mu_j^* - \mu_0\right)^2 .$$

Если же математическое ожидание не задано, то аналогичным образом приходим к

$$\overline{\chi}^2 = \frac{1}{\sigma^2} \sum_{j=1}^{N} \left(\mu_j^* - \overline{x}\right)^2 ,$$

где $\overline{x}$ — выборочное среднее всех наблюдений.

Если ни математическое ожидание, ни дисперсия априори не известны, то в качестве статистики к. о. п. полагаем

$$\overline{e}^2 = \frac{\sum_{j=1}^{N} \left(\mu_j^* - \overline{x}\right)}{\sum_{j=1}^{N} \left(x_j - \overline{x}\right)^2} .$$

Наиболее важным для практики оказывается критерий $\overline{e}^2$. Использование критерия $\overline{\chi}_0^2$ оправдано в тех случаях, когда объектом статистического анализа становится результат некоторого преобразования исходной последовательности, для которого применяется модель нормального распределения (например, суммирование элементов последовательности испытаний Бернулли). Статистика $\overline{\chi}^2$ играет важную роль при статистическом анализе непараметрического тренда.

Непараметрические критерии используются для анализа как стохастической упорядоченности, так и монотонного тренда параметра распределения, когда прямым путем построить критерий не удается. В первую очередь здесь речь идет о критерии, основанном на числе ступенек: распределение случайной величины l при нулевой гипотезе в широких пределах не зависит от распределения x_j, что говорит в пользу устойчивости статистического вывода.

Чтобы построить ранговый аналог критерия отношения правдоподобия, придется несколько видоизменить постановку задачи. Будем считать, что последовательность наблюдений распадается на n групп по k элементов в каждой, причем j-й группе отвечает функция распределения $F_j(u)$ для $j = 1, \ldots, n$. Сопоставим каждому наблюдению x_s его ранг r_s среди всех наблюденных значений и перейдем к анализу последовательности средних рангов, состоящей из элементов

$$\overline{r}_j = \sum_{s=(j-1)k+1}^{jk} \frac{r_s}{k}\,, \quad j = 1, \ldots, n\,.$$

Ранговый аналог критерия отношения правдоподобия был введен в [8] и частично исследовался в [5, разд. 4.4] в следующей форме:

$$\overline{\chi}^2(\overline{\mathbf{r}}) = \frac{12}{N(N+1)} \sum_{j=1}^{n} k \left(\overline{r}_j^{\,*} - \frac{N+1}{2} \right)^2 , \tag{2}$$

где $\overline{r}_j^{\,*}$ — результат применения PAV-процедуры к последовательности средних рангов.

2 Распределения статистик

Основной результат относительно поведения перечисленных выше статистик при нулевой гипотезе заключается в том, что их условные по числу ступенек распределения суть соответственно χ^2-распределения и β-распределение. В результате центральной становится задача нахождения $\tau_N(j)$ — вероятности того, что о. м. п. монотонного тренда, полученная по N наблюдениям, содержит j ступенек. В [9] было доказано, что

$$\tau_N(j) = \frac{|s(N,j)|}{N!}\,, \quad j = 1, \ldots, N\,, \tag{3}$$

где $s(N,j)$ — числа Стирлинга 1-го рода. Позже было обосновано, что в качестве μ^* не обязательно рассматривать выборочное среднее. Таким образом, получаем:

$$\Pr\left\{\overline{\chi}_0^2 \le u\right\} = \sum_{j=1}^{N} \tau_N(j) \Pr\left\{\chi_j^2 \le u\right\}\,, \;\; u \ge 0\,; \tag{4}$$

$$\Pr\left\{\overline{\chi}^2 \le u\right\} = \frac{1}{N} + \sum_{j=2}^{N} \tau_N(j) \Pr\left\{\chi_{j-1}^2 \le u\right\}\,, \;\; u > 0\,; \tag{5}$$

$$\left.\begin{aligned} &\Pr\left\{\overline{e}^2 \le u\right\} = \frac{1}{N} + \sum_{j=2}^{N-1} \tau_N(j) \Pr\left\{\mathsf{B}_{(j-1)/2,(N-j)/2} \le u\right\}, \quad 0 < u < 1; \\ &\Pr\left\{\overline{e}^2 = 1\right\} = \frac{1}{N!}, \end{aligned}\right\} \quad (6)$$

где χ^2_j — случайная величина, имеющая χ^2-распределение с j степенями свободы; $\mathsf{B}_{s,t}$ — случайная величина, имеющая β-распределение с параметрами (s,t).

Теперь, в частности, можно найти первые семиинварианты распределений статистик к. о. п., что необходимо для решения задачи аппроксимации этих распределений. Для этого сначала строится характеристическая функция $\varphi(t)$ распределения статистики, а затем находятся первые семиинварианты. По аналогии с дзета-функцией Римана для $j = 1, 2, \ldots$ введем обозначения

$$\zeta_N(j) = \sum_{s=1}^{N} \frac{1}{s^j}.$$

Если положить $z = (1 - 2it)^{-1/2}$, то получаем:

$$\begin{aligned} \ln \varphi_{\overline{\chi}^2_0}(t) &= \sum_{j=1}^{N} \ln(z + j - 1) - \ln N!; \\ \ln \varphi_{\overline{\chi}^2}(t) &= \ln \varphi_{\overline{\chi}^2_0}(t) - \ln z, \end{aligned}$$

откуда следует:

$$\begin{aligned} \kappa_1\left\{\overline{\chi}^2_0\right\} &= \kappa_1\left\{\overline{\chi}^2\right\} + 1 = \zeta_N(1); \\ \kappa_2\left\{\overline{\chi}^2_0\right\} &= \kappa_2\left\{\overline{\chi}^2\right\} + 2 = 3\zeta_N(1) - \zeta_N(2); \\ \kappa_3\left\{\overline{\chi}^2_0\right\} &= \kappa_3\left\{\overline{\chi}^2\right\} + 8 = 15\zeta_N(1) - 9\zeta_N(2) + 2\zeta_N(3); \\ &\ldots \end{aligned}$$

В случае распределения статистики $\overline{e}^2$ ситуация несколько сложнее: сначала находятся факториальные моменты $\mu_{[k]}\{l\}$, далее обычные моменты выражаются через факториальные, что позволяет получить обычные моменты для $\overline{e}^2$. Отсюда стандартным образом находятся семиинварианты $\kappa_j\{\overline{e}^2\}$, которые в итоге выражаются через $\zeta_N(j)$.

Теперь можно получить асимптотические при $N \to \infty$ представления семиинвариантов и сделать вывод, что распределения всех трех статистик, а также и числа ступенек при нулевой гипотезе стремятся к нормальному. При этом логарифмический характер сходимости указывает на проблематичность использования в практических целях приближений, основанных на нормальном распределении.

Таблица 1 Значения $k(N)$, обеспечивающие применимость асимптотических результатов

N	$k(N)$
10–11	5
12–29	4
30–199	3
200–999	2
≥ 1000	1

Использование рангового аналога критерия отношения правдоподобия на практике опирается на асимптотический результат о том, что при $N \to \infty$ и $k \to \infty$ так, что $n = const$, статистика (2) имеет то же распределение, что и $\overline{\chi}^2$. Дело в том, что нормированный надлежащим образов вектор частных средних рангов стремится по распределению к вектору частных средних нормально распределенных величин с одним и тем же математическим ожиданием и некоторой фиксированной дисперсией, смещенных общим выборочным средним.

Составить представление о реальных возможностях процедуры, основанной на рангах, удается благодаря моделированию поведения статистики критерия. В частности, при различных значениях n находился такой минимальный объем выборки N, при котором значения критического уровня значимости при сравнении гипотетических частот и соответствующих эмпирических частот «практически» совпадали. Это позволило найти оценки $k(N)$, обеспечивающие применимость асимптотических результатов (табл. 1).

Громоздкость полученных представлений (4)–(6) вынуждает в первую очередь обратить внимание на условные варианты критериев: решение принимается при пороге, найденном в зависимости от наблюденного значения l. Полностью оправданным такой подход окажется только в том случае, если число ступенек не имеет никакого отношения к проблеме решения. Прибегая к методу моделирования, нетрудно выяснить, что это не так, и поэтому использовать условный критерий по возможности не следует.

3 Практика применения

Применение критерия значимости на практике связано с необходимостью эффективно решать задачи вычисления значений функций распределения статистик и нахождения соответствующих квантилей. В случае анализа монотонного тренда были исследованы два пути: прямой, когда непосредственно использовались (3)–(6), и путь аппроксимации исходных распределений известными.

Из выражения для производящей функции распределения вероятностей числа ступенек получаем:

$$N\tau_N(j) = (N-1)\tau_{N-1}(j) + \frac{1}{N-1}\left[(N-1)\tau_{N-1}(j-1)\right], \quad j = 2, \ldots, N\,.$$

Следовательно, в основу нахождения $\tau_N(j)$ могут быть положены соотношения:

$$\tilde{\tau}_j(0) = 0\,, \quad \tilde{\tau}_j(j+1) = 0 \text{ для } j = 1, \ldots, N-1\,;$$

$$\tilde{\tau}_j(1) = 1\,;$$

$$\tilde{\tau}_j(s) = \tilde{\tau}_{j-1}(s) + \frac{1}{j-1}\,\tilde{\tau}_{j-1}(s-1) \text{ для } j = 2,\ldots,N \text{ и } s = 1,\ldots,j\,;$$

$$\tau_N(j) = \frac{1}{N}\tilde{\tau}_N(j) \text{ для } j = 1,\ldots,N\,.$$

Для того чтобы снять вопрос о возникающих при использовании рекуррентных соотношений вычислительных погрешностях, пришлось обратиться к алгоритмам длинной целочисленной арифметики, позволяющим практически для любого большого значения N находить числа Стирлинга 1-го рода. После этого стало возможным получать точные значения $\tilde{\tau}_N(j)$ в виде числителя и знаменателя (3), а затем проводить оценивание точности вычисления $\tau_N(j)$, найденных с помощью обычных машинных операций для 8-байтовых чисел с плавающей запятой. Таким образом, прямыми вычислениями было показано, что

$$\max_{N\le 10^5}\max_{j=1,\ldots,N}|\tau_N(j) - \tilde{\tau}_N(j)| < 2{,}5\cdot 10^{-16}\,,$$

а это полностью удовлетворяет требованиям к точности при вычислениях значений функции распределения. Таким образом, нахождение весов в смесях распределений типа (4)–(6) возможно по рекуррентным формулам с помощью обычных машинных операций.

Снижения временны́х затрат при создании соответствующих стандартных процедур можно добиться с помощью двух приемов:

(1) вычисляя и сохраняя опорные наборы $\{\tau_N(1), \tau_N(2), \ldots\}$ при различных значениях N, кратных какому-либо значению (например, 10^3), а затем пересчитывая вероятности на их основе до ближайшего нужного N по прямым или обратным рекуррентным соотношениям;

(2) сокращая количество вычисляемых $\tau_N(j)$, ориентируясь на задаваемые ошибки вычислений.

Последний способ востребован еще и потому, что непосредственное использование (4)–(6) встречает значительные трудности даже при умеренных значениях N (вычисление значений χ^2- или β-распределений само по себе достаточно трудоемко). На помощь приходит свойство распределения l: оно сосредоточено в достаточно узкой области, смещающейся в бесконечность с ростом N со скоростью $\ln N$. Таким образом, начиная с некоторого слагаемого, часть суммы общего вида $\sum_{j=1}^{N}\tau_N(j)F_j(u)$, где $F_j(u) \le 1$, окажется пренебрежимо мала. Для этого достаточно найти такое минимальное $m(N,\varepsilon)$, что

$$\sum_{j=m(N,\varepsilon)+1}^{N}\tau_N(j) \le \varepsilon\,. \tag{7}$$

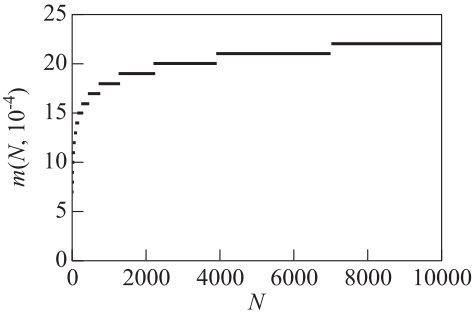


Рис. 1 Зависимость от N числа слагаемых $m(N, 10^{-4})$, гарантирующего при вычислении функции смеси распределений абсолютную погрешность 10^{-4}

Таблица 2 Отдельные значения $m(N, \varepsilon)$

ε	N			
	10	10^2	10^3	10^4
10^{-4}	8	14	18	22
10^{-8}	10	19	25	30
10^{-12}	10	23	30	32

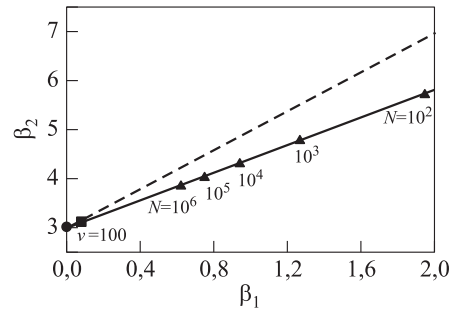


Рис. 2 Представление распределений в плоскости коэффициентов асимметрии β_1 и эксцесса β_2

Последнее неравенство решается прямым перебором. Если задать погрешность в виде $\varepsilon = 10^{-p}$, то (7) можно заменить неравенством

$$\sum_{j=1}^{m(N,\varepsilon)} s(N,j) \geq N! - \left[\frac{N!}{10^p}\right],$$

которое уже реализуемо с помощью длинной целочисленной арифметики. Но обычных операций с плавающей точкой при нахождения $\tau_N(j)$ оказывается вполне достаточно.

Характер получающихся результатов виден из рис. 1: необходимы лишь десятки элементов смеси, чтобы корректно использовать к. о. п. на практике. Кроме того, согласно табл. 2, значения $m(N, \varepsilon)$ остаются «практичными» при росте требований к точности вычислений ε: например, для предельных значений параметров N, ε сложность вычисления значений функции распределения по-прежнему снижается более чем на два порядка.

При аппроксимации распределений статистик к. о. п. отправной точкой стал выбор наиболее подходящих для этих целей распределений. Для этого в плоскости коэффициентов асимметрии β_1 и эксцесса β_2 были построены изображения известных распределений, для которых доступны представления семиинвариантов. Если принять определения $\beta_1 = \kappa_3^2/\kappa_2^3$ и $\beta_2 = \kappa_4/\kappa_2^2 + 3$, то в качестве иллюстрации на рис. 2 отражены случаи нормального распределения (черный кружок), однопараметрического гамма-распределения γ_p (штриховая полупрямая вида $\beta_2 = (3/2)\beta_1 + 3$), $\overline{\chi}^2$-распределения в виде сплошной линии из точек, которая лишь асимптотически дает прямую, отличающуюся от штриховой полупрямой. Именно вариант гамма-распределения оказался наиболее подходящим при аппроксимации распределений

всех трех статистик к. о. п. обнаружения тренда. В качестве приближения использовалась линейная комбинация вида $a\gamma_p + b$, неизвестные величины которой p, a и b определялись обычным образом через семиинварианты распределений (4)–(6).

Графики рис. 2 демонстрируют также непрактичность нормальной аппроксимации: на линии $\overline{\chi}^2$-распределения черными треугольниками помечены точки, отвечающие отдельным значениям параметра N. Видно, что даже для больших объемов выборки до точки $(0, 3)$, отвечающей семейству нормальных распределений, крайне «далеко». Для сравнения указан случай обычного χ^2-распределения с числом степеней свободы $\nu = 100$ (черный квадрат), для которого общепринято применение нормального приближения.

4 Заключение

Подход, основанный на модели монотонных изменений, нечасто используется на практике и слабо освещен в литературе; его применение позволяет не только строить процедуры статистического контроля источника данных, но и с единой точки зрения взглянуть на еще не известные законы изменений и вести их планомерное исследование. Пока еще недостаточно разработанную область он открывает и в теоретическом плане. Здесь исследователь встречается с задачами, наиболее сложными в иерархии проблем статистики: оценивание при ограничениях и в условиях совпадения размерности параметрического и выборочного пространств, а также проверка множества сложных гипотез.

Особую роль играет прикладной аспект проблемы статистического анализа. Сюда входит широкий спектр задач от обоснования жизнеспособности разрабатываемого подхода до обеспечения его использования в реальных условиях. Количество соответствующих публикаций исчисляется единицами, и причина такой «холодности» исследователей к практическим вопросам заключается в сложности предлагаемого аппарата и большой трудоемкости процедур обработки данных.

Идея ступенчатого представления о. м. п. монотонного тренда позволила максимально прояснить и упростить вывод распределений статистик критерия отношения правдоподобия при условии справедливости нулевой гипотезы и оказалась весьма продуктивной при получении новых результатов о выборочных свойствах элементов оценки.

Литература

1. *Fitzmaurice G. M., Laird N. M., Ware J. H.* Applied longitudinal analysis. — 2nd ed. — Hoboken, NJ, USA: Wiley, 2011. 701 p.
2. *Fernandez N. B., Aiken J., Smith J. T.* Use of maximal spanning trees and the gamma test of monotone trend in the development and assessment of teams // Procd. Soc. Behv., 2011. Vol. 26. P. 147–158.
3. *Roth M., Jongbloed G., Buishand A.* Monotone trends in the distribution of climate extremes // Theor. Appl. Climatol., 2019. Vol. 136. Iss. 3-4. P. 1175–1184.

4. *Kliestik T., Valaskova K., Nica E., Kovacova M., Lazaroiu G.* Advanced methods of earnings management: Monotonic trends and change-points under spotlight in the Visegrad countries // Oeconomia Copernicana, 2020. Vol. 11. Iss. 2. P. 371–400.
5. *Barlow R. E., Bartholomew D. J., Bremmer J. M., Brunk H. D.* Statistical inference under order restrictions: The theory and application of isotonic regression. — London, New York, Sydney, Toronto: Wiley, 1972. 388 p.
6. *Кривенко М. П., Мацкевич И. В.* Свойства элементов оценки монотонного тренда // Теория вероятностей и ее применения, 1988. Т. 33. Вып. 2. С. 336–348.
7. *Кривенко М. П., Стрельцов К. В.* О распределении числа ступенек в оценке монотонного тренда // Теория вероятностей и ее применения, 1996. Т. 41. Вып. 1. С. 53–64.
8. *Chacko V. J.* Testing homogeneity against ordered alternatives // Ann. Math. Stat., 1963. Vol. 34. Iss. 3. P. 945–956.
9. *Miles R. E.* The complete amalgamation into blocks, by weighted means, of a finite set of real numbers // Biometrika, 1959. Vol. 46. Iss. 3. P. 317–327.

Поступила в редакцию 26.07.21

DISTRIBUTIONS OF LIKELIHOOD RATIO STATISTICS FOR MONOTONE TREND DETECTION

M. P. Krivenko

Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The problem of analyzing a monotonic trend is considered. The model is based on the change in the mean of the normal distribution, for which the distributions of the likelihood ratio statistics are mixtures of chi-square or beta distributions with weights determined through the Stirling numbers of the first kind. The questions of approximation of these distributions are investigated. An effective method for calculating significance criteria by discarding insignificant elements of a mixture is proposed which makes it possible to reduce the complexity of calculations by several orders of magnitude. The issues of expanding the criteria for analyzing the trend of the mean normal distribution to procedures for identifying stochastic ordering are discussed. For these purposes, the authors apply the monotonic trend criterion for the averaged ranks, the statistics of which, in the absence of changes, is distributed as a mixture of chi-square distributions. The use of a graded assessment structure was of fundamental importance: it was possible not only to obtain original results, but also to ensure the viability of the proposed solutions.

Keywords: monotone trend; likelihood ratio test; nonparametric trend detection; asymptotic distribution; probability distribution approximation

DOI: 10.14357/08696527210403

References

1. Fitzmaurice, G. M., N. M. Laird, and J. H. Ware. 2011. *Applied longitudinal analysis*. 2nd ed. Hoboken, NJ: Wiley. 701 p.
2. Fernandez, N. B., J. Aiken, and J. T. Smith. 2011. Use of maximal spanning trees and the gamma test of monotone trend in the development and assessment of teams. *Procd. Soc. Behv.* 26:147–158.
3. Roth, M., G. Jongbloed, and A. Buishand. 2019. Monotone trends in the distribution of climate extremes. *Theor. Appl. Climatol.* 136(3-4):1175–1184.
4. Kliestik, T., K. Valaskova, E. Nica, M. Kovacova, and G. Lazaroiu. 2020. Advanced methods of earnings management: Monotonic trends and change-points under spotlight in the Visegrad countries. *Oeconomia Copernicana* 11(2):371–400.
5. Barlow, R. E., D. J. Bartholomew, J. M. Bremmer, and H. D. Brunk. 1972. *Statistical inference under order restrictions: The theory and application of isotonic regression*. London, New York, Sydney, Toronto: Wiley. 388 p.
6. Krivenko, M. P., and I. V. Mackevich. 1988. Properties of elements of an estimate of monotonous trend. *Theor. Probab. Appl.* 33(2):316–329.
7. Krivenko, M. P., and K. V. Streltsov. 1996. On the distribution of the number of steps in an estimate of monotone trend. *Theor. Probab. Appl*. 41(1):25–34.
8. Chacko, V. J. 1963. Testing homogeneity against ordered alternatives. *Ann. Math. Stat.* 34(3):945–956.
9. Miles, R. E. 1959. The complete amalgamation into blocks, by weighted means, of a finite set of real numbers. *Biometrika* 46(3):317–327.

Received July 26, 2021

Contributor

Krivenko Michail P. (b. 1946) — Doctor of Science in technology, professor, leading scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; mkrivenko@ipiran.ru

ПОСТКВАНТОВАЯ СХЕМА ЦИФРОВОЙ ПОДПИСИ НА АЛГЕБРЕ МАТРИЦ

Д. Н. Молдовян[1], *А. А. Молдовян*[2], *Н. А. Молдовян*[3]

Аннотация: Рассматривается вопрос использования конечной мультипликативной группы обратимых матриц размерности 2×2, заданных над полем $GF(p)$, как алгебраического носителя схем цифровой подписи, основанных на вычислительной трудности скрытой задачи дискретного логарифмирования (СЗДЛ) и удовлетворяющих общему критерию постквантовой стойкости. Показано существование достаточно большого числа коммутативных подгрупп, обладающих двухмерной цикличностью, что использовано при построении конкретной схемы подписи, представляющей интерес как постквантовая криптосхема. В разработанной схеме подписи применена новая форма задания СЗДЛ, которая характеризуется использованием коммутативной группы с двухмерной цикличностью в качестве скрытой группы и маскирующих операций двух разных типов: (1) обладающих свойством взаимной коммутативности с операцией экспоненцирования и (2) свободных от этого свойства. Для обеспечения корректности работы криптосхемы в процедуре проверки подлинности подписи применяется проверочное уравнение специального вида, а при генерации подписи один из элементов последней вычисляется как корень квадратного уравнения.

Ключевые слова: конечная группа матриц; вычислительно трудная задача; дискретный логарифм; цифровая подпись; постквантовая криптография

DOI: 10.14357/08696527210404

1 Введение

В настоящее время для защиты информации в информационно-телекоммуникационных системах широко применяются двухключевые криптографические алгоритмы, которые основаны на вычислительной трудности задачи факторизации и задачи дискретного логарифмирования (ЗДЛ). Последняя может быть решена на квантовом компьютере за полиномиальное время [1, 2], поэтому ожидаемое в ближайшем будущем появление практически доступных квантовых вычислителей и длительность процесса принятии новых криптографических

[1]Санкт-Петербургский институт информатики и автоматизации РАН Санкт-Петербургского Федерального исследовательского центра РАН, mdn.spectr@mail.ru

[2]Санкт-Петербургский институт информатики и автоматизации РАН Санкт-Петербургского Федерального исследовательского центра РАН, maa1305@yandex.ru

[3]Санкт-Петербургский институт информатики и автоматизации РАН Санкт-Петербургского Федерального исследовательского центра РАН, nmold@mail.ru

стандартов обусловили высокую степень актуальности разработки постквантовых двухключевых криптосхем и стандартов на их основе [3, 4].

Для разработки практичных постквантовых схем электронной цифровой подписи (ЭЦП) представляет интерес использование вычислительной трудности СЗДЛ, задаваемой в конечных некоммутативных ассоциативных алгебрах (КНАА) [5], и общего критерия постквантовой стойкости [6].

С целью повышения производительности схем ЭЦП, основанных на СЗДЛ и удовлетворяющих указанному критерию, в настоящей работе рассматривается вопрос применения конечных групп обратимых матриц, заданных над простым конечным полем $\mathrm{GF}(p)$, в качестве алгебраического носителя.

2 Общий критерий постквантовой стойкости

Пусть задан генератор G циклической группы, имеющей простое значение порядка q, и некоторый элемент Y', вычисленный по формуле $Y' = G^x$, где x — неизвестное натуральное число ($x < q$). Задача вычисления x по известным G и Y' называется ЗДЛ. Известный алгоритм решения ЗДЛ на квантовом компьютере связан с построением периодической функции $f(i,j) = Y'^i G^j$, где i и j — пара целых чисел, принимающей значения в конечной циклической группе. Гипотетический квантовый вычислитель позволяет за полиномиальное время найти длины периодов функции $f(i,j)$, в том числе и значение $(-1, x)$, которое является длиной одного из периодов:

$$Y'^i G^j = Y'^{(i-1)} G^{j+x} \Rightarrow f(i,j) = f(i-1, j+x).$$

В схемах ЭЦП, основанных на вычислительной сложности СЗДЛ, в качестве базовой операции, вносящей основной вклад в стойкость криптосхемы, используется операция возведения в степень x, выполняемая в скрытой группе, генерируемой некоторым элементом G, представляющим собой секретное значение. Открытый ключ формируется путем выполнения двух разных маскирующих операций над элементами G и G^x по формулам $Y = \psi_1(G^x)$ и $Z = \psi_2(G)$, где ψ_1 и ψ_2 — согласованные между собой маскирующие операции, обладающие свойством взаимной коммутативности с операцией экспоненцирования. Функция $f(i,j) = Y^i Z^j$ содержит период длины $(-1, x)$, но принимает значения, лежащие в достаточно большом числе различных циклических групп, что обеспечивает стойкость к атакам, использующим известные квантовые алгоритмы нахождения длины периода.

Однако можно предположить, что в будущем будут разработаны новые квантовые алгоритмы, которые позволят находить периоды для функций, принимающих значения, не ограниченные какой-либо фиксированной циклической группой. Для обеспечения стойкости к более широкому классу квантовых атак в работе [6] предложен общий критерий обеспечения постквантовой стойкости, формулируемый следующим образом: *построение периодической функции*

с использованием открытых параметров криптосхемы, которая содержит период с длиной, определяемой значением дискретного логарифма x, должно быть практически (вычислительно) невыполнимой задачей.

В качестве способа построения схем ЭЦП, удовлетворяющих этому критерию, в [6] предложено использовать коммутативную группу с двухмерной цикличностью в качестве скрытой группы и вычисление одного из элементов открытого ключа в зависимости от двух независимых элементов G и Q скрытой группы (элементов, образующих базис группы), порядок каждого из которых равен простому числу q. Группы, порождаемые базисом, в котором все элементы имеют одинаковое значение порядка, называются группами с многомерной цикличностью [7].

Разработанная в [6] конкретная постквантовая схема ЭЦП на основе общего критерия постквантовой стойкости использует подпись в виде трех элементов — двух чисел и одного вектора — и удвоенное проверочное соотношение. Первое приводит к существенному увеличению длины подписи, а второе — удвоению размера открытого ключа и уменьшению производительности процедуры проверки подлинности ЭЦП.

Построение более практичной схемы ЭЦП, реализующей общий критерий постквантовой стойкости, связано с поиском проверочных уравнений и маскирующих операций, отличных от тех, что были применены в [6]. При этом в качестве алгебраического носителя криптосхемы представляется интересным использовать конечную алгебру матриц размерности 2×2 над полем $\mathrm{GF}(p)$, которая фактически является частным случаем четырехмерных КНАА, заданным по прореженной таблице умножения базисных векторов. При этом прореженность таблицы обеспечивает существенное снижение вычислительной сложности операции умножения, что дает дополнительное повышение производительности схемы ЭЦП.

3 Используемый алгебраический носитель

Рассмотрим множество обратимых матриц размерности 2×2 над полем $\mathrm{GF}(p)$ с характеристикой $p = 2q + 1$, где q — 256-битное простое число, и матрицу G порядка q, которая не входит в множество скалярных матриц вида

$$S_n = \begin{pmatrix} n & 0 \\ 0 & n \end{pmatrix},$$

где $n = 1, 2, \ldots, p-1$. Значение λ, представляющее собой примитивный элемент поля $\mathrm{GF}(p)$, определяет матрицу S_λ, всевозможные целочисленные степени которой пробегают значения всех скалярных матриц, множество которых является циклической группой порядка $p-1$. Скалярная матрица $S = S_\lambda^2$ имеет порядок, равный q. Матрицы S и G перестановочны и образуют базис $\langle S, G \rangle$ коммутативной (все матрицы вида $S^i G^j$ попарно перестановочны) группы $\Gamma_{\langle S,G \rangle}$,

содержащейся в рассматриваемой алгебре матриц и обладающей порядком q^2 и двухмерной цикличностью. Любая другая матрица G' порядка q, которая неперестановочна с G, также задает коммутативную группу $\Gamma_{\langle S,G'\rangle}$, обладающую порядком q^2 и двухмерной цикличностью. Произвольные две группы из множества групп $\Gamma_{\langle S,G'\rangle}$ пересекаются в циклической группе скалярных матриц. Существование большого числа попарно неперестановочных матриц G' определяет существование большого числа различных групп с двухмерной цикличностью, содержащихся в алгебре матриц. Следовательно, на такой алгебре потенциально можно построить схему ЭЦП, основанную на СЗДЛ с использованием конечной коммутативной группы с двухмерной цикличностью в качестве скрытой группы.

4 Используемые маскирующие операции

Пусть $Q = SG^r$, где $r < q$ — случайное натуральное число. Очевидно, что $\langle Q, G\rangle$ — базис группы $\Gamma_{\langle Q,G\rangle}$ порядка q^2, обладающей двухмерной цикличностью. При использовании $\Gamma_{\langle Q,G\rangle}$ в качестве скрытой группы зададим формирование открытого ключа в виде трех матриц U, Y и Z следующим образом.

1. Сгенерировать две невырожденные матрицы A и B порядка p^2-1, которые неперестановочны с G и между собой. Эти матрицы задают три маскирующих операции отображения

$$\varphi_1(X) = AXB^{-1};\ \ \varphi_2(X) = BXA^{-1};\ \ \psi(X) = BXB^{-1},$$

   где X — отображаемая матрица, из которых только операция ψ взаимно коммутативна с операцией возведения в степень.

2. Сгенерировать два случайных натуральных числа $x < q$ и $u < q$. Затем вычислить первый элемент открытого ключа в виде матрицы

$$U = \varphi_1\left(G^x Q^u\right) = AG^x Q^u B^{-1}.$$

3. Вычислить второй элемент открытого ключа в виде матрицы

$$Y = \psi(G) = BGB^{-1}.$$

4. Вычислить третий элемент открытого ключа в виде матрицы

$$Z = \varphi_2(Q) = BQA^{-1}.$$

Заметим, что связанные между собой операции φ_1 и φ_2 позволяют задать следующие композиционные операции отображения, обладающие свойством

взаимной коммутативности с операцией экспоненцирования, выполняемой над парами и тройками матриц:

$$\psi_1(X_1X_2) = \varphi_1(X_1)\,\varphi_2(X_2) = AX_1X_2A^{-1} \Rightarrow$$
$$\Rightarrow \psi_1\left((X_1X_2)^k\right) = (\psi_1(X_1X_2))^k\ ;$$
$$\psi_2(X_1X_2X_3) = \varphi_1(X_1)\,\psi(X_2)\,\varphi_2(X_3) = AX_1X_2X_3A^{-1} \Rightarrow$$
$$\Rightarrow \psi_2\left((X_1X_2X_3)^k\right) = (\psi_2(X_1X_2X_3))^k\ .$$

Пара чисел (x, u) и матрицы G, Q, A и B являются секретными и используются при вычислении подписи к заданному электронному документу M.

5 Формирование цифровой подписи

Процедура генерации ЭЦП выполняется с использованием личного секретного ключа подписанта в виде набора значений (x, u, G, Q, A) и некоторой стойкой специфицированной 256-битной хеш-функции h следующим образом.

1. Сгенерировать случайное натуральное число $k < q$ и вычислить матрицу
$$K = G^k.$$
2. Сгенерировать случайное натуральное число $t < q$ и вычислить матрицу
$$R = AKQ^tA^{-1}.$$
3. Вычислить первый элемент ЭЦП в виде значения хеш-функции от документа с присоединенной к нему матрицей R:
$$e = h(M, R).$$
4. Вычислить второй элемент ЭЦП в виде числа s:
$$s = \sqrt{\frac{1}{e}\left(k - \frac{tx}{u+1}\right)} \bmod q\,.$$
5. Если значение под корнем в правой части последней формулы будет равно квадратичному невычету по модулю q, то перейти к шагу 2.
6. Вычислить третий элемент ЭЦП в виде числа d:
$$d = \left(\frac{t}{s(u+1)} - 1\right) \bmod q\,.$$

На выходе получаем 768-битную подпись в виде трех 256-битных значений (e, s, d). Последние две формулы вытекают из уравнения, используемого на первом шаге процедуры проверки подлинности ЭЦП. Основной вклад в вычислительную сложность этого алгоритма вносят операции экспоненцирования. Легко видеть, что в среднем для генерации одной подписи требуется выполнить три операции возведения в 256-битную степень.

6 Верификация цифровой подписи

Для проверки соответствия подписи (e, s, d) документу M и открытому ключу (U, Y, Z) выполняется следующая процедура.

1. Вычислить матрицу
$$\tilde{R} = \left(UY^{es}Z(UZ)^d\right)^s .$$
2. Вычислить значение хеш-функции от документа с присоединенной к нему матрицей $\tilde{R}$:
$$\tilde{e} = h\left(M, \tilde{R}\right).$$
3. Если выполняется равенство $\tilde{e} = e$, то подпись подлинная, иначе ЭЦП отклоняется.

Основной вклад в вычислительную сложность этого алгоритма вносят три операции возведения в степень. Корректность работы описанной схемы ЭЦП заключается в том, что подпись, вычисленная в полном соответствии с процедурой формирования ЭЦП, проходит проверочную процедуру как подлинная подпись.

7 Доказательство корректности схемы электронной цифровой подписи

Пусть (e, s, d) — корректно сформированная подпись к документу M. Тогда при ее подстановке на вход проверочной процедуры имеем следующее:

$$\begin{aligned}
\tilde{R} &= \left(UY^{es}Z(UZ)^d\right)^s = \\
&= \left(AG^xQ^uB^{-1}BG^{es}B^{-1}BQA^{-1}\left(AG^xQ^uB^{-1}BQA^{-1}\right)^d\right)^s = \\
&= \left(AG^{x+es}Q^{u+1}A^{-1}AG^{xd}Q^{d(u+1)}A^{-1}\right)^s = \left(AG^{x+es+xd}Q^{u+1+d(u+1)}A^{-1}\right)^s = \\
&= AG^{xs(d+1)+es^2}Q^{sd(u+1)+s(u+1)}A^{-1} = \\
&= AG^{xst/(s(u+1))+e(1/e)(k-tx/(u+1))}Q^{(t/(u+1)-s)(u+1)+s(u+1)}A^{-1} = \\
&= AG^kQ^tA^{-1} = R \Rightarrow h\left(M, \tilde{R}\right) = h(M, R) \Rightarrow \tilde{e} = e .
\end{aligned}$$

Последнее равенство означает подлинность проверяемой ЭЦП.

8 Обсуждение

Периодическая функция, заданная на основе открытых параметров (элементов открытого ключа), принимает значения, которые зависят от матриц Q, G, Q^u и G^x. Это маскирует периодичности, связанные со степенями x и u. При этом без знания секретных значений x, u, G, Q, A и B представляется вычислительно невозможным задать периодическую функцию, содержащую периоды, длины которых зависят от степеней x и u.

В частности, рассмотрим периодическую функцию

$$F_1(i,j) = (U \circ Y \circ Z)^i (U \circ Z)^j = A \circ G^{xi+i+xj} \circ Q^{ui+i+uj+j} \circ A^{-1}.$$

Пусть длина периода этой функции равна (δ_i, δ_j). Тогда в силу независимости векторов G и Q имеем:

$$\begin{cases} x\delta_i + \delta_i + x\delta_j \equiv 0 \bmod q; \\ u\delta_i + \delta_i + u\delta_j + \delta_j \equiv 0 \bmod q \end{cases} \Rightarrow \begin{cases} (x+1)\delta_i + x\delta_j \equiv 0 \bmod q; \\ (u+1)\delta_i + (u+1)\delta_j \equiv 0 \bmod q. \end{cases}$$

Последняя система с неизвестными δ_i и δ_j при $u \neq -1 \bmod q$ имеет единственное решение $(\delta_i, \delta_j) = (0,0)$, поскольку ее главный определитель отличен от нуля: $\Delta = u+1$. Таким образом, функция $F_1(i,j)$ может иметь только периоды длиной (aq, bq) при некоторых целочисленных значениях a и b.

Для длины (δ_i, δ_j) периода функции

$$F_2(i,j) = (Z \circ U)^i Y^j = B \circ Q^{i+ui} \circ G^{xi+j} \circ B^{-1}$$

можно записать:

$$\begin{cases} \delta_i + u\delta_i \equiv 0 \bmod q; \\ x\delta_i + \delta_j \equiv 0 \bmod q \end{cases} \Rightarrow \begin{cases} \delta_i \equiv 0 \bmod q; \\ \delta_j \equiv 0 \bmod q, \end{cases}$$

т. е. функция $F_2(i,j)$ не может содержать периоды, длина которых отлична от (aq, bq). Таким образом, описанная схема ЭЦП удовлетворяет общему критерию постквантовой стойкости.

Представляет практический интерес сравнение описанной схемы ЭЦП с аналогами, предложенными в ходе конкурса НИСТ (Национального института стандартов и технологий США) [3], и схемой подписи [6] по производительности и размерам открытого ключа и подписи. С учетом компромисса между размерами открытого ключа и ЭЦП и производительностью алгоритмов формирования и проверки подлинности подписи предпочтительными представляются следующие кандидаты на постквантовый стандарт ЭЦП: Falcon

Сравнение предложенной схемы ЭЦП с известными

Схема ЭЦП	Длина подписи, байт	Длина открытого ключа, байт	Скорость формирования ЭЦП, о. е.	Скорость верификации ЭЦП, о. е.
Falcon-512	657	897	5	3
qTESLA-p-I	2 592	15 000	3	6
Dilithium	2 044	1 184	2	1
[6]	192	768	6	4
Данная работа	96	384	12	12

[https://falcon-sign.info/], qTESLA [https://qtesla.org/] и Dilithium [https://pq-crystals.org/dilithium/index.shtml]. В таблице дается сопоставление версий сравниваемых схем ЭЦП, соответствующих 128-битной стойкости. Сравнение показывает, что схемы ЭЦП, основанные на СЗДЛ, обладают существенно меньшим суммарным размером подписи и открытого ключа и более высокой производительностью. При этом предложенная в данной работе схема ЭЦП по сравнению со схемой [6] обладает меньшей длиной подписи и более высокой производительностью.

9 Заключение

Предложен новый способ построения схем ЭЦП, основанных на вычислительной трудности СЗДЛ, которые удовлетворяют общему критерию постквантовой стойкости. Способ обеспечивает существенное уменьшение размеров открытого ключа и подписи. Разработана схема ЭЦП, представляющая интерес для разработки на ее основе стандарта постквантовой ЭЦП.

Литература

1. *Shor P. W.* Polynomial-time algorithms for prime factorization and discrete logarithms on quantum computer // SIAM J. Comput., 1997. Vol. 26. P. 1484–1509.
2. *Jozsa R.* Quantum algorithms and the Fourier transform // Proc. R. Soc. Lon. Ser. A, 1998. Vol. 454. P. 323–337.
3. Federal Register. Announcing Request for Nominations for Public-Key Post-Quantum Cryptographic Algorithms. https://www.gpo.gov/fdsys/pkg/FR-2016-12-20/pdf/2016-30615.pdf.
4. Post-quantum cryptography / Eds. J. Ding, R. Steinwandt. — Lecture notes in computer science ser. — Springer, 2019. Vol. 11505. 420 p.
5. *Молдовян А. А., Молдовян Н. А.* Новые формы скрытой задачи дискретного логарифмирования // Труды СПИИРАН, 2019. Т. 18. № 2. С. 504–530.
6. *Moldovyan D. N., Moldovyan A. A., Moldovyan N. A.* Digital signature scheme with doubled verification equation // Computer Science J. Moldova, 2020. Vol. 28. P. 80–103.

7. *Moldovyan N. A.* Fast signatures based on non-cyclic finite groups // Quasigroups Related Systems, 2010. Vol. 18. P. 83–94.

Поступила в редакцию 08.09.20

POST-QUANTUM SIGNATURE SCHEME ON MATRIX ALGEBRA

D. N. Moldovyan, A. A. Moldovyan, and N. A. Moldovyan

St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg Federal Research Center of the Russian Academy of Sciences, 39, 14th Line V.O., St. Petersburg 199178, Russian Federation

Abstract: The paper considers the use of a finite multiplicative group of invertible matrices of dimension 2×2 set over the field $\mathrm{GF}(p)$ as algebraic carrier of the digital signature schemes based on the computational difficulty of the hidden discrete logarithm problem and satisfying the general criterion of post-quantum resistance. The existence of a sufficiently large number of commutative subgroups with two-dimensional cyclicity is shown. This fact is used in the construction of a specific signature scheme which is of interest as a post-quantum cryptosystem. In the introduced digital signature scheme, a new form of the hidden discrete logarithm problem is applied. The said form is characterized by the use of a commutative group with two-dimensional cyclicity as a hidden group and masking operations of two different types: (i) having the property of mutual commutativity with the exponentiation operation and (ii) free from this property. To ensure the correct operation of the cryptographic scheme, a special type of verification equation is used in the signature authentication procedure, and when generating a signature, one of the elements of the latter is calculated as a root of quadratic equation.

Keywords: finite group of matrices; computationally difficult problem; discrete logarithm; digital signature; post-quantum cryptography

DOI: 10.14357/08696527210404

References

1. Shor, P. W. 1997. Polynomial-time algorithms for prime factorization and discrete logarithms on quantum computer. *SIAM J. Comput.* 26:1484–1509.
2. Jozsa, R. 1998. Quantum algorithms and the Fourier transform. *Proc. R. Soc. Lon. Ser. A* 454:323–337.
3. Federal Register. Announcing Request for Nominations for Public-Key Post-Quantum Cryptographic Algorithms. Available at: https://www.govinfo.gov/content/pkg/FR-2016-12-20/pdf/2016-30615.pdf (accessed September 20, 2021).
4. Ding, J., and R. Steinwandt, eds. 2019. *Post-quantum cryptography*. Lecture notes in computer science ser. Springer. Vol. 11505. 420 p.

5. Moldovyan, A. A., and N. A. Moldovyan. 2019. Novye formy skrytoy zadachi diskretnogo logarifmirovaniya [New forms of defining the hidden discrete logarithm problem]. *Trudy SPIIRAN* [SPIIRAS Proceedings] 18(2):504–530.
6. Moldovyan, D. N., A. A. Moldovyan, and N. A. Moldovyan. 2020. Digital signature scheme with doubled verification equation. *Computer Science J. Moldova* 28:80–103.
7. Moldovyan, N. A. 2010. Fast signatures based on non-cyclic finite groups. *Quasigroups Related Systems* 18:83–94.

Received September 8, 2020

Contributors

Moldovyan Dmitriy N. (b. 1986) — Candidate of Science (PhD) in technology, scientist, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg Federal Research Center of the Russian Academy of Sciences, 39, 14th Line V.O., St. Petersburg 199178, Russian Federation; mdn.spectr@mail.ru

Moldovyan Alexander A. (b. 1951) — Doctor of Science in technology, professor, principal scientist, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg Federal Research Center of the Russian Academy of Sciences, 39, 14th Line V.O., St. Petersburg 199178, Russian Federation; maa1305@yandex.ru

Moldovyan Nikolay A. (b. 1953) — Doctor of Science in technology, professor, principal scientist, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg Federal Research Center of the Russian Academy of Sciences, 39, 14th Line V.O., St. Petersburg 199178, Russian Federation; nmold@mail.ru

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ, 2021. Т. 31. № 4. С. 48–60

ОБ ОДНОМ СПОСОБЕ ОБНАРУЖЕНИЯ ЭКСПЛУАТАЦИИ УЯЗВИМОСТЕЙ И ЕГО ПАРАМЕТРАХ

Ю. В. Косолапов[1]

Аннотация: При успешной эксплуатации уязвимости, приводящей к запуску вредоносного кода, обычно запускаемый код вызывает некоторую системную функцию. В качестве способа обнаружения эксплуатации уязвимости программы P предлагается алгоритм на основе проверки на нетипичность расстояния между вызовом с номером i и вызовом с номером $i-j$, где $j \in \in \{1,\ldots,T\}$, $T \in \mathbb{N}$. Под расстоянием понимается разность виртуальных адресов вызова этих функций, а типичность определяется путем проверки принадлежности вычисленного расстояния профилю расстояний, построенному ранее для защищаемой программы P. Кроме параметра T алгоритм обнаружения использует параметр $W(\in \mathbb{N})$ — количество профилей, по которым проверяется типичность вызова. При этом для $j \in \{1,\ldots,W\}$ профиль с номером j строится по парам вызовов, разность индексов вызовов которых в легитимной последовательности равна j. Чем больше T и W, тем меньше вероятность ложного обнаружения нелегитимного исполнения кода, однако с ростом этих параметров растет и время проверки типичности. В ходе экспериментального исследования выяснено, что достаточные значения параметров (W,T) зависят от набора отслеживаемых функций. Поэтому для каждого набора отслеживаемых функций (и каждой защищаемой программы P) эти параметры алгоритма должны находиться отдельно. Установлено, что при $W > 1$ существенно снижается вероятность ложного обнаружения.

Ключевые слова: уязвимости программного обеспечения; расстояние между вызовами функций; защита программ

DOI: 10.14357/08696527210405

1 Введение и постановка задачи

Наличие уязвимостей в системном или прикладном программном обеспечении (ПО) снижает защищенность информационных систем (ИС), использующих это ПО. Автоматизированный поиск уязвимостей с использованием таких технологий, как символьное исполнение и фаззинг, не гарантирует отсутствия уязвимостей в коде [1]. Повысить защищенность ИС можно путем обновления ПО и за счет использования систем обнаружения вторжений (СОВ). При таком подходе уязвимость кода должна быть *известна* разработчикам ПО и разработчикам СОВ, чтобы первые могли разработать обновление, а вторые — правило

[1]Институт математики, механики и компьютерных наук им. И. И. Воровича, Южный федеральный университет, itaim@mail.ru

обнаружения. Для *неизвестных* уязвимостей обнаружение их эксплуатации считается более сложной задачей, так как в этом случае СОВ должна быть способна различать легитимное и нелегитимное выполнение кода. Источником данных для принятия решения о легитимности выполнения кода часто рассматривается последовательность вызовов функций (ПВФ) системного и/или прикладного ПО [2–4]. Это обосновывается тем, что целью разработчиков эксплоитов (программного кода, эксплуатирующего уязвимость) обычно ставится получение контроля над уязвимой системой путем запуска специального кода (шеллкода), из которого осуществляется вызов системной и/или прикладной функции [5]. Считалось, что появление вызовов функций из шеллкода в легитимной последовательности вызовов может быть обнаружено [2]. Однако в [6] показано, что шеллкод может быть эффективно построен так, чтобы его выполнение не отличалось от легитимного. Другим возможным способом защиты от эксплуатации уязвимостей является использование технологии рандомизации адресного пространства процесса. Использование мелкогранулярной рандомизации адресного пространства программы, как это предложено, например, в [7], хотя существенно и затрудняет эксплуатацию уязвимостей, но, как показано в [5], не делает такую эксплуатацию невозможной. К тому же для мелкогранулярной рандомизации может потребоваться доступ к исходному коду защищаемой программы, а также внесение изменений в статический компоновщик и динамический загрузчик [7].

Цель работы — построение способа (без модификации программы и компонентов сборки/запуска программ) для выявления нелегитимного выполнения кода, противодействующего атакам (см. [5, 6]), а также оценка значений параметров, обеспечивающих низкую вероятность ложного обнаружения. Идея способа приведена в работе автора [8].

2 Способ обнаружения эксплуатации уязвимости

Под отслеживаемыми функциями будем понимать набор функций, по последовательности вызовов которых может быть определено аномальное выполнение программы. Такой набор может быть определен на основе результатов исследования образцов деструктивного кода, как это сделано в [4]. Пусть $\mathcal{L}(P)$ — множество *отслеживаемых* имен функций, используемых программой P, $\mathrm{Path}_t(P(I))$ — ПВФ *в рамках одного потока* программы P, запущенной в момент времени t с входными данными I:

$$\mathrm{Path}_t(P(I)) = \left(f_1^{t,I}, \ldots, f_{N_{t,I}}^{t,I}\right). \tag{1}$$

Вызов $f_k^{t,I}$ в трассе (1) с порядковым номером k представи́м в виде тройки

$$f_k^{t,I} = \left(\mathsf{ba}_k^{t,I}, \mathsf{ca}_k^{t,I}, n_k^{t,I}\right),$$

где $\mathsf{ba}_k^{t,I} = \mathcal{B}(f_k^{t,I})$ — адрес в виртуальной памяти процесса, по которому загружен исполнимый модуль, содержащий по адресу $\mathsf{ca}_k^{t,I} = \mathcal{C}(f_k^{t,I})$ вызов функции с именем $n_k^{t,I} = \mathcal{N}(f_k^{t,I})$, $n_k^{t,I} \in \mathcal{L}(P)$. Двойной верхний индекс в обозначениях используется для указания на то, что эти значения в общем случае зависят от времени запуска программы и входных данных. Для $t_1 \neq \neq t_2$ адреса загрузки программы и модулей в адресное пространство процесса могут быть разными (из-за применения рандомизации адресного пространства), поэтому ПВФ $\mathrm{Path}_{t_1}(P(I))$ и $\mathrm{Path}_{t_2}(P(I))$ (при одних и тех же входных данных программы) в общем случае отличаются адресами ba и ca вызываемых функций. При этом порядок вызова функций не меняется: $N_{t_1,I} = N_{t_2,I}$, $n_i^{t_1,I} = n_i^{t_2,I}$, $i = 1, \ldots, N_{t_1,I}$. Имеет место равенство

$$\mathsf{ca}_k^{t_2,I} = \mathsf{ca}_k^{t_1,I} + \left(\mathsf{ba}_k^{t_2,I} - \mathsf{ba}_k^{t_1,I}\right) = \mathcal{T}_{t_1,t_2}\left(\mathsf{ca}_k^{t_1,I}\right), \tag{2}$$

позволяющее перейти от наблюдаемых адресов в момент t_2 к адресам в момент t_1.

Замечание 2.1. Адреса размещения модулей (или иначе — карта размещения) в любой момент времени известны. Поэтому при использовании обозначения $\mathcal{T}_{t_1,t_2}(A)$ подразумевается, что

(1) используя карту размещения в момент t_1, на основе A определяется начальный адрес b^{t_1} модуля, адресному пространству которого принадлежит A;

(2) в карте размещения для момента t_2 определяется начальный адрес b^{t_2} того же модуля;

(3) к A прибавляется $b^{t_2} - b^{t_1}$: $\mathcal{T}_{t_1,t_2}(A) = A + b^{t_2} - b^{t_1}$. Очевидно, что $\mathcal{T}^{-1}_{t_1,t_2}(A) = A + b^{t_1} - b^{t_2}$.

Для натурального s число

$$\rho\left(f_{k-s}^{t,I}, f_k^{t,I}\right) = \rho_{k-s,k}^{t,I} = \mathsf{ca}_k^{t,I} - \mathsf{ca}_{k-s}^{t,I}$$

назовем *s-расстоянием* между $f_{k-s}^{t,I}$ и $f_k^{t,I}$ для программы P. С учетом (2) получаем выражение перехода от расстояний в момент t_1 к расстояниям в момент t_2 и наоборот:

$$\rho_{k-s,k}^{t_2,I} = \mathcal{T}_{t_1,t_2}\left(\mathsf{ca}_k^{t_1,I}\right) - \mathcal{T}_{t_1,t_2}\left(\mathsf{ca}_{k-s}^{t_1,I}\right). \tag{3}$$

Очевидно, что $\rho_{k-s,k}^{t_2,I} = 0$ тогда и только тогда, когда $\rho_{k-s,k}^{t_1,I} = 0$.

Замечание 2.2. Будем предполагать, что если в P успешно эксплуатируется уязвимость на входных данных I', то запускаемый шеллкод содержит вызовы некоторых функций из списка $\mathcal{L}(P)$. Вызовы из шеллкода будем называть

нелегитимными, а остальные — легитимными. Последовательность вызовов функций (1) в этом случае может быть представлена в виде:

$$\mathrm{Path}_t(P(I')) = \left(\underbrace{f_1^{t,I'}, \ldots, f_i^{t,I'}}_{\text{легитимные}}, \underbrace{f_{i+1}^{t,I'}, \ldots, f_{i+k}^{t,I'}}_{\text{нелегитимные}}, \underbrace{f_{i+k+1}^{t,I'}, \ldots}_{\text{легитимные}} \right), \quad k \geqslant 1 .$$

Также будем предполагать, что шеллкод не может быть записан в те области памяти процесса, в которых уже размещен код программы. Поэтому $\rho(f_i^{t,I'}, f_{i+1}^{t,I'}) \neq 0$.

Замечание 2.3. Так как выражение (2) позволяет переходить между адресами в разные моменты времени и из контекста обычно известны входные данные программы P, то в обозначении вызовов иногда будем опускать верхний двойной индекс.

Для выявления нелегитимных вызовов необходимо иметь набор данных (профиль), характеризующий легитимное исполнение P. Пусть $\mathfrak{I}(P)$ — множество значений *легитимных* входных данных P. В силу (2) профиль может формироваться путем многократного запуска P на входных данных из $\mathfrak{I}(P)$. Отметим, что для всех возможных данных из $\mathfrak{I}(P)$ построение профиля представляет сложную задачу, так как $\mathfrak{I}(P)$ может быть потенциально бесконечным. При использовании конечного $\mathfrak{I}(P)$, с одной стороны, мощность этого множества должна быть приемлемой для обучения, а с другой стороны, набор $\mathfrak{I}(P)$ должен обладать свойством репрезентативности. Последнее означает, что входящие в набор данные должны быть «похожи» на те данные, для обнаружения аномалий на которых будет использоваться профиль. Например, если P — это браузер, то в $\mathfrak{I}(P)$ следует включать сайты тех типов, которые обычно посещает этот пользователь этого браузера.

Для $(n, m) \in \mathfrak{L}(P) \times \mathfrak{L}(P)$ будем писать

$$(n, m, \rho) \propto_s \mathrm{Path}_t(P(I)), \tag{4}$$

если найдется такое k, что $m = \mathfrak{N}(f_k)$, $n = \mathfrak{N}(f_{k-s})$ и $\rho = \mathfrak{C}(f_k) - \mathfrak{C}(f_{k-s})$. Запись (4) означает, что в $\mathrm{Path}_t(P(I))$ найдется пара вызовов, первым из которых будет вызов функции с именем n, а следующим, s-м от него, — вызов функции с именем m, при этом s-расстояние между ними равно ρ. Для каждой пары $(n, m) \in \mathfrak{L}(P) \times \mathfrak{L}(P)$ рассмотрим множество возможных s-расстояний между функциями n и m:

$$D_{(n,m),s} = \{\rho \in \mathbb{Z} | \exists I \in \mathfrak{I}(P) : (n, m, \rho) \propto_s \mathrm{Path}_t(P(I))\} .$$

В общем случае $D_{(n,m),s} \neq D_{(m,n),s}$. *Профилем s-расстояний* программы P назовем множество

$$\mathfrak{D}_s(P) = \left\{ D_{(n,m),s} : (n, m) \in \mathfrak{L}(P) \times \mathfrak{L}(P) \right\}, \tag{5}$$

построенное *по всем потокам* P. Будем говорить, что тройка (n, m, ρ) $(\in \mathfrak{L}(P) \times \mathfrak{L}(P) \times \mathbb{Z})$ *принадлежит* профилю $\mathcal{D}_s(P)$ и писать $(n, m, \rho) \sim \mathcal{D}_s(P)$, если $\rho \in D_{(n,m),s}$. (Иногда будем говорить, что число ρ для пары (n, m) принадлежит профилю $\mathcal{D}_s(P)$.) Если $\mathcal{D}_s(P)$ построен в момент t_1, а тройка (n, m, ρ) получена по последовательности вида (1) в момент $t = t_2$ на входных данных I', то будем говорить, что тройка (n, m, ρ) *соответствует* $\mathcal{D}_s(P)$, если в $\mathrm{Path}_{t_2}(P(I'))$ найдется такая пара вызовов f_{k-s} и f_k, что $n = \mathcal{N}(f_{k-s})$, $m = \mathcal{N}(f_k)$, $\mathcal{C}(f_k) - \mathcal{C}(f_{k-s}) = \rho$, при этом (см. (3))

$$\left(n, m, \mathcal{T}^{-1}_{t_1,t_2}\left(\mathcal{C}(f_k)\right) - \mathcal{T}^{-1}_{t_1,t_2}\left(\mathcal{C}\left(f_{k-s}\right)\right)\right) \sim \mathcal{D}_s(P). \tag{6}$$

Соответствие тройки (n, m, ρ) профилю $\mathcal{D}_s(P)$ обозначим $(n, m, \rho) \sim_{t_2} \mathcal{D}_s(P)$, тем самым подчеркивая, что тройка (n, m, ρ) получена в момент t_2, отличный от момента формирования профиля $\mathcal{D}_s(P)$. (Когда будет ясно, о какой паре функций (n, m) идет речь, будем просто говорить о соответствии числа ρ профилю $\mathcal{D}_s(P)$.)

Замечание 2.4. Так как с помощью (2) и (6) можно перейти от адресации в момент t_1 к адресации в момент t_2 и наоборот, то далее будем предполагать, что базовые адреса загрузки модулей не меняются от запуска к запуску.

Для натурального W рассмотрим набор

$$\mathcal{D}_1(P), \ldots, \mathcal{D}_W(P). \tag{7}$$

В простейшем варианте предлагаемый метод обнаружения аномалий заключается в нахождении в последовательности вида (1) таких пар *соседних* вызовов, 1-расстояние между которыми *не соответствует* ни одному профилю вида (5) из набора (7). Целесообразность использования $W > 1$ поясним на примере. Рассмотрим рис. 1, *а*, где представлены два *возможных* пути выполнения P в момент t в зависимости от входных данных: $\mathrm{Path}_t(P(I)) = (a, b, c, d)$ и $\mathrm{Path}_t(P(I')) = (a, d)$. Предположим, что при формировании $\mathcal{D}_1(P)$ последовательность (a, b, c, d) реализовалась на некоторых входных данных этапа

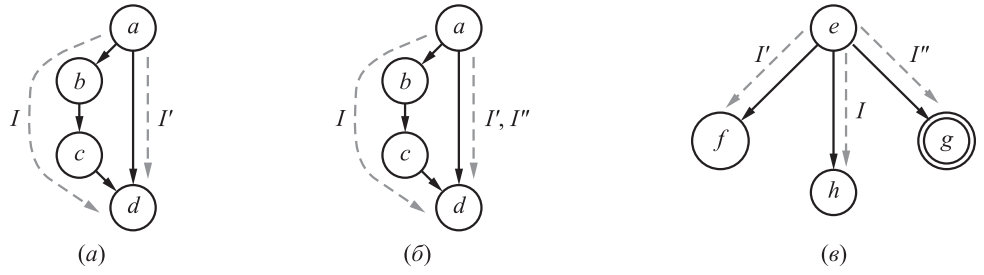


Рис. 1 Последовательности вызовов для разных входных данных

обучения, а последовательность (a, d) не реализовывалась для всех входных данных этого этапа, причем $(\mathcal{N}(a), \mathcal{N}(d), \rho = \rho(a, d)) \not\sim_t \mathcal{D}_1(P)$. Тогда появление цепочки (a, d) в пути выполнения P приведет к тому, что легитимное выполнение P при проверке только по профилю $\mathcal{D}_1(P)$ будет расценено как нелегитимное (ложное обнаружение). Однако $(\mathcal{N}(a), \mathcal{N}(d), \rho) \sim \mathcal{D}_3(P)$, так как, по предположению, цепочка (a, b, c, d) реализовалась на этапе формирования профиля. Таким образом, при конечном $\mathcal{I}(P)$ чем больше W, тем меньше вероятность ложного обнаружения нелегитимного вызова.

При конечном $\mathcal{I}(P)$ наличие в последовательности вида (1) такой пары (f_{i-1}, f_i), что тройка $(\mathcal{N}(f_{i-1}), \mathcal{N}(f_i), \mathcal{C}(f_i) - \mathcal{C}(f_{k-1}))$ не принадлежит ни одному профилю из набора (7), может соответствовать двум случаям. Во-первых, это может свидетельствовать о нелегитимности входных данных, и в этом случае аномалия будет правильно расценена как эксплуатация уязвимости в P. Во-вторых, входные данные могут быть легитимными, однако при их обработке в P задействуется ПВФ, которая не задействовалась при формировании набора (7). В этом случае вызов будет ложно расценен как нелегитимный. Для снижения вероятности ложного обнаружения предлагается подход на основе проверки легитимности расстояний до $T(\in \mathbb{N})$ предыдущих вызовов. Опишем этот способ.

Пусть $T, W, k \in \mathbb{N}$, $k \geqslant T + 1$. Для вызова f_k из трассы вида (1) рассмотрим вектор

$$\mathbf{v}_k = (v_1, \ldots, v_T), \tag{8}$$

где

$$v_i = \begin{cases} \min\limits_{1,\ldots,W} \{s : (\mathcal{N}(f_{k-i}), \mathcal{N}(f_k), \mathcal{C}(f_k) - \mathcal{C}(f_{k-i})) \sim \mathcal{D}_s(P)\} - 1; \\ W, \text{если } \forall s \in \{1, \ldots, W\} : (\mathcal{N}(f_{k-i}), \mathcal{N}(f_k), \mathcal{C}(f_k) - \mathcal{C}(f_{k-i})) \not\sim \mathcal{D}_s(P), \end{cases} \quad i = 1, \ldots, T.$$

Вес вектора (8) определим по формуле $\mathcal{W}(\mathbf{v}_k) = (2\sum_{i=1}^{T} i v_i)(T(T+1)W)^{-1}$, причем

$$\mathcal{W}(\mathbf{v}_k) = \begin{cases} 1, & \text{если } \rho(f_{k-j}, f_k) \text{ не содержится ни в одном профиле из (7);} \\ 0, & \text{если } \rho(f_{k-j}, f_k) \text{ содержится в профиле } \mathcal{D}_1(P) \end{cases}$$

для всех $j \in \{1, \ldots, T\}$. Через $\tau(T, W)$ обозначим такое пороговое значение из отрезка $[0, 1]_{\mathbb{R}}$, что при

$$\mathcal{W}(\mathbf{v}_k) < \tau(T, W) \tag{9}$$

вызов $f_k^{t,I}$ принимается за легитимный, а при $\mathcal{W}(\mathbf{v}_k) \geqslant \tau(T, W)$ — за нелегитимный. Алгоритм CheckTraceBack (см. алгоритм) реализует способ обнаружения

Алгоритм CheckTraceBack

Исходные параметры: 1) ПВФ $\mathrm{Path}_{t_2}(P(I))$ вида (1) длины n_I $(I \notin \mathcal{I}(P))$, 2) T, W, $\tau(T,W)$, 3) набор (7); 4) L — признак самообучения
Результат: Сообщение о нетипичной (not typical)/типичной (typical) ПВФ
result = typical
цикл $k = T+1, \ldots, n_I - 1$ **выполнять**
  Вычислить вектор $\mathbf{v}_k$ вида (8)
  если *не выполняется неравенство* (9) **тогда**
    result = nottypical
    Выйти из цикла
  конец условия
  иначе
    если $L = 1$ **тогда**
      Добавить в $\mathcal{D}_1(P)$ расстояние ρ для пары (n, m)
    конец условия
  конец условия
конец цикла
возвратить result

аномального выполнения по предыдущим T вызовам. Для предлагаемого способа задачей является подбор таких T, W, $\tau(T,W)$ и $\mathcal{I}(P)$, что вероятности ложного обнаружения и ложного пропуска не превышают заданных значений. В настоящей работе задача подбора $\tau(T,W)$ и $\mathcal{I}(P)$ не решается ($\tau(T,W) = 1$, а набор $\mathcal{I}(P)$ формируется с учетом результатов [4]).

3 Подбор параметров

Рассмотрим трассу вызовов (1). Заметим, что из условия

$$(\mathcal{N}(f_{k-j}), \mathcal{N}(f_k), \mathcal{C}(f_k) - \mathcal{C}(f_{k-j})) \sim_t \mathcal{D}_j(P) \tag{10}$$

не вытекает, что между вызовами функций с именами $\mathcal{N}(f_{k-j})$ и $\mathcal{N}(f_k)$, расположенными по адресам соответственно $\mathcal{C}(f_{k-j})$ и $\mathcal{C}(f_k)$, существует путь через $j-1$ отслеживаемых функций. Запись (10) означает лишь, что найдутся два таких адреса $\mathcal{C}_1$ и $\mathcal{C}_2$, что вызов функции с именем $\mathcal{N}(f_{k-j})$ расположен по адресу $\mathcal{C}_1$, а вызов функции с именем $\mathcal{N}(f_k)$ — по адресу $\mathcal{C}_2$, при этом между вызовами этих функций происходит $j-1$ вызовов отслеживаемых функций и $\mathcal{C}_2 - \mathcal{C}_1 = \mathcal{C}(f_k) - \mathcal{C}(f_{k-j})$. Рассмотрим пример, изображенный на рис. 1, *б* и 1, *в*. На этих рисунках представлены возможные пути выполнения двух разных участков кода одной программы. Назовем их условно *участок 1* (рис. 1, *б*) и *участок 2* (рис. 1, *в*). Пусть легитимные данные I используются в момент формирования

профиля, а легитимные данные I' и нелегитимные данные I'' — в момент тестирования, $\mathcal{N}(b) = \mathcal{N}(e)$, $\mathcal{N}(d) = \mathcal{N}(f) = \mathcal{N}(h)$, $\mathcal{C}(d) - \mathcal{C}(b) = \mathcal{C}(f) - \mathcal{C}(e)$. На нелегитимных данных I'' эксплуатируется уязвимость на *участке 2*, где из шеллкода осуществляется вызов g. (Успех в определении вызова g как нетипичного зависит от того, насколько нетипичны расстояния от T предыдущих вызовов до вызова g; нахождение значений параметров алгоритма для успешного выявления нетипичного выполнения требует отдельного исследования, в частности необходим набор известных эксплоитов для программы.) В момент формирования профиля на входных данных I в $\mathcal{D}_2(P)$ будет добавлено расстояние $\mathcal{C}(d) - \mathcal{C}(b)$ для пары $(\mathcal{N}(b), \mathcal{N}(d))$, а в $\mathcal{D}_3(P)$ будет добавлено расстояние $\mathcal{C}(d) - \mathcal{C}(a)$ для пары $(\mathcal{N}(a), \mathcal{N}(d))$. Поэтому на входных данных I' и I'' при $W \geqslant 3$ вызов d после вызова a будет расценен алгоритмом CheckTraceBack как типичный при $\tau(W, T) = 1$. Заметим, что алгоритмом CheckTraceBack при $W \geqslant 2$ вызов f также будет расценен как типичный, поскольку

$$(\mathcal{N}(e), \mathcal{N}(f), \mathcal{C}(f) - \mathcal{C}(e)) = (\mathcal{N}(b), \mathcal{N}(d), \mathcal{C}(d) - \mathcal{C}(b)) \sim_t \mathcal{D}_2(P),$$

где t — момент запуска P. Однако при этом от e до f, как следует из рис. 1, *в*, не существует цепочки из трех вызовов, начинающейся e и заканчивающейся f. Другими словами, решение о легитимности f принимается на основе расстояния, которое *не связано* с f. С одной стороны, это снижает вероятность ложного обнаружения аномального исполнения, а с другой — может привести к росту вероятности ложного пропуска.

Подбор T и W для алгоритма CheckTraceBack выполняется экспериментально. В работе в качестве P рассмотрен браузер FireFox, а в качестве $\mathcal{L}(P)$ — набор функций $\mathcal{L}_{MW}(P)$, типичный для вредоносного ПО [4]. На основе данных www.similarweb.com выбрано 15 различных тематик сайтов, и для каждой тематики выбраны первые десять наиболее посещаемых сайтов — сформировано 15 множеств сайтов, которые обозначим $S_1, \ldots, S_{15}$. Сайты выбирались так, чтобы $S_i \cap S_j = \varnothing$ для $i \neq j$. Используя эти множества, было составлено 15 наборов профилей вида (7), где $W = 10$: первый набор был получен по $\mathcal{S}_1 = S_1$, второй — по $\mathcal{S}_2 = S_1 \cup S_2$ и т. д. до пятнадцатого набора, который был построен по $\mathcal{S}_{15} = \cup_{i=1}^{15} S_i$. В каждом наборе (7) для формирования отдельного профиля использовалось от 300 до 600 тыс. вызовов отслеживаемых функций. Также сформирована выборка из 450 сайтов (по 30 сайтов для каждой из 15 тематик), из которой случайным образом выбрано четыре непересекающихся тестовых выборки по 25 сайтов каждая. Для каждой тестовой выборки построена трасса вида (1), состоящая из порядка 700 тыс. вызовов.

Обозначим через $A_{0,i}$ число таких соседних пар (f_{k-1}, f_k) вызовов в трассе, что $(\mathcal{N}(f_{k-1}), \mathcal{N}(f_k), \mathcal{C}(f_k) - \mathcal{C}(f_{k-1})) \not\sim \mathcal{D}_1(P)$, где $\mathcal{D}_1(P)$ принадлежит набору вида (7), построенному по $\mathcal{S}_i$. Через $A_{1,i}(W, T)$ обозначим число вызовов f_k, для которых $\mathcal{W}(\mathbf{v}_k) = 1$ (для $\tau(W, T) = 1$) при вычислении веса по набору, построенному по множеству $\mathcal{S}_i$. Для каждой пары (W, T), где $W \in \{1, \ldots, 10\}$,

$T \in \{5, 10, 15, \ldots, 100\}$, по трассе вызовов программы FireFox, полученной на основе тестовой выборки, вычислялись два числа: $A_0 = \sum_{i=1}^{15} A_{0,i}$ (не зависит от W и T) и $A_1(W,T) = \sum_{i=1}^{15} A_{1,i}(W,T)$. Величина $\Delta(W,T) = A_1(W,T)A_0^{-1}$ характеризует долю выявленных аномалий алгоритмом CheckTraceBack по отношению к числу аномалий, выявленных при проверке расстояний между соседними вызовами по профилю $\mathcal{D}_1(P)$. Чем меньше $\Delta(W,T)$, тем меньше в среднем алгоритмом CheckTraceBack обнаруживается ложных аномалий. Так как A_0 не зависит от W и T, то для $\tau(W,T) = 1$ из определения координат вектора (8) и определения веса этого вектора получаем, что последовательности $\Delta(1,T), \ldots, \Delta(i,T), \ldots$ и $\Delta(W,1), \ldots, \Delta(W,j), \ldots$ не возрастают.

Другими словами, при $\tau(W,T) = 1$ чем больше параметры W и T, тем ожидается меньше ложно положительных срабатываний алгоритма. Анализ значений $\Delta(W,T)$ может позволить выбрать T и W для алгоритма CheckTraceBack так, чтобы обеспечивались приемлемые вероятность ложного обнаружения и скорость обработки данных.

Для трех из четырех тестовых выборок (второй, третьей и четвертой) результаты близкие, поэтому на рис. 2, *а* и 2, *б* в случае $\mathfrak{L}(P) = \mathfrak{L}_{MW}(P)$ изображены результаты только для первой и четвертой выборок. На рис. 2, *а* графики сближаются при $T \geqslant 40$ для $W \geqslant 3$. При этом на рис. 2, *б* графики очень близки для $T \geqslant 20$ при $W \geqslant 4$. Выше отмечалось, что с ростом W и T увеличивается время вычисления веса (9). Поэтому в качестве рекомендуемых значений пары (W,T) в этом случае можно взять пару $(4, 30)$. В работе также проведены эксперименты по влиянию самообучения (параметр L в алгоритме CheckTraceBack). Результаты экспериментов на наборе, построенном по $\mathcal{S}_{15}$, показали, что самообучение практически не влияет на снижение числа ложных обнаружений.

С целью оценки зависимости значений (W,T) от набора $\mathfrak{L}(P)$ были проведены эксперименты на тех же наборах обучающих и тестовых выборок для набора отслеживаемых функций $\mathfrak{L}_{DA}(P)$, в который включены системные функции из модулей kernel32.dll и txfw32.dll, связанные с доступом к локальной файловой системе. Зависимость $\Delta(W,T)$ от значений (W,T) в этом случае показана на рис. 2, *в* и 2, *г*. По рис. 2, *г* можно заключить, что при $W \geqslant 5$ и $T \geqslant 35$ графики близки друг к другу, в то время как по рис. 2, *в* видно, что близость графиков наступает при $T \geqslant 50$ для $W \geqslant 5$. Снова, из соображений скорости обработки, в качестве пары (W,T) может быть выбрано значение $(5, 40)$.

Эксперименты также показали, что для $W \geqslant 5$ режим обучения практически не влияет на число обнаруживаемых ложных срабатываний, за исключением первой выборки, где число выявленных ложных срабатываний снижается не более чем на 2. Важно отметить, что на наборе функций $\mathfrak{L}_{DA}(P)$ число ложно обнаруживаемых алгоритмом CheckTraceBack нелегитимных вызовов в среднем меньше, чем на наборе функций $\mathfrak{L}_{MW}(P)$. Это подчеркивает актуальность задачи поиска оптимального набора отслеживаемых функций $\mathfrak{L}(P)$.

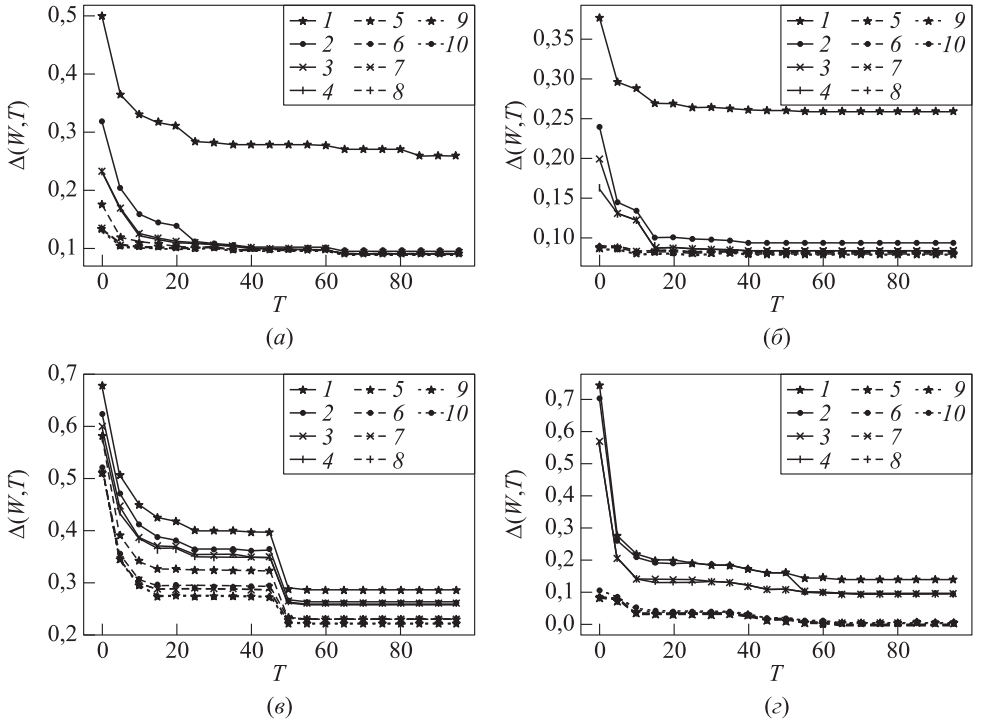


Рис. 2 Зависимость $\Delta(W,T)$ от (W,T) для $P = \mathsf{FireFox}$, $\mathfrak{L}(P) = \mathfrak{L}_{MW}(P)$ (*а*, *б*) и $\mathfrak{L}(P) = \mathfrak{L}_{DA}(P)$ (*в*, *г*): *1* — $W = 1$; *2* — 2; *3* — 3; *4* — 4; *5* — 5; *6* — 6; *7* — 7; *8* — 8; *9* — 9; *10* — $W = 10$

4 Заключение

Нахождение параметров алгоритма CheckTraceBack, при которых обеспечиваются приемлемые вероятности ложного обнаружения и ложного пропуска, является актуальной задачей. Эксперименты показывают, что, с одной стороны, параметры алгоритма зависят от набора отслеживаемых вызовов, а с другой стороны, режим самообучения практически не влияет на число обнаруживаемых ложных срабатываний. Отметим, что в работе не оценивается вероятность ложного пропуска алгоритмом эксплуатации уязвимости, так как для такой оценки необходимо иметь набор известных эксплоитов для фиксированной программы, нацеленных на запуск кода путем эксплуатации уязвимостей в этой программе. Для исследуемой программы FireFox при найденных значениях параметров (W,T) запуск эксплоита, рассмотренного в [8], обнаруживается всегда. Однако одного эксплоита, даже с разными вызываемыми из него функциями (из отслеживаемого набора), полагаем, недостаточно. Поэтому первым направлени-

ем дальнейшего исследования представляется проверка способности алгоритма CheckTraceBack к выявлению других известных эксплоитов. Вторым направлением исследования, которое тесно связано с первым, должно стать нахождение оптимального набора отслеживаемых функций $\mathfrak{L}(P)$. Наконец, третьим направлением является разработка способа защиты конфиденциальности профилей (5), так как эти профили хранятся на стороне пользователя защищаемой программы P и могут быть использованы им для построения эксплоита, обходящего защиту.

Литература

1. *Rohlf C., Ivnitskiy Y.* Attacking clientside JIT compilers // Black Hat USA, 2011. https://www.nccgroup.trust/globalassets/resources/us/presentations/documents/attacking_clientside_jit_compilers_paper.pdf.
2. *Forrest S., Hofmeyr S., Somayaji A.* The evolution of system-call monitoring // Annual Computer Security Applications Conference Proceedings. — Piscataway, NJ, USA: IEEE, 2008. P. 418–430. doi: 10.1109/ACSAC.2008.54.
3. *Singh A., Arora R., Pareek H.* Malware analysis using multiple API sequence mining CFG. arXiv.org, 2017. 12 p. arXiv:1707.02691 [cs.CR].
4. *Gupta S., Sharma H., Kaur S.* Malware characterization using Windows API call sequences // J. Cyber Security Mobility, 2018. Vol. 7. No. 4. P. 363–378. doi: 10.1007/978-3-319-49445-6_15.
5. *Snow K. Z., Monrose F., Davi L., Dmitrienko A., Liebchen C., Sadeghi A.-R.* Just-in-time code reuse: On the effectiveness of fine-grained address space layout randomization // Symposium on Security and Privacy Proceedings. — Piscataway, NJ, USA: IEEE, 2013. P. 574–588. doi: 10.1109/SP.2013.45.
6. *Wagner D., Soto P.* Mimicry attacks on host-based intrusion detection systems // 9th Conference on Computer and Communications Security Proceedings. — New York, NY, USA: ACM, 2002. P. 255–264. doi: 10.1145/586110.586145.
7. *Нурмухаметов А. Р., Жаботинский Е. А., Курмангалеев Ш. Ф., Гайсарян С. С., Вишняков А. В.* Мелкогранулярная рандомизация адресного пространства программы при запуске // Труды ИСП РАН, 2017. Т. 29. Вып. 6. С. 163–182. doi: 10.15514/ISPRAS-2017-29(6)-9.
8. *Косолапов Ю. В.* Об обнаружении эксплуатации уязвимостей, приводящей к запуску вредоносного кода // Моделирование и анализ информационных систем, 2020. Т. 27. Вып. 2. С. 138–151. doi: 10.18255/1818-1015-2020-2-138-151.

Поступила в редакцию 20.08.20

ON ONE METHOD FOR DETECTING EXPLOITATION OF VULNERABILITIES AND ITS PARAMETERS

Yu. V. Kosolapov

Institute for Mathematics, Mechanics, and Computer Science named after I. I. Vorovich, Southern Federal University, 8a Milchakova Str., Rostov-on-Don 344090, Russian Federation

Abstract: When a program vulnerability is successfully exploited, the exploit often calls some system function. Therefore, one of the possible ways to detect exploitation of a vulnerability of a specific program is to check for atypical distance between the call with the number i and the call with the number $i - j$ where $j \in \{1, \ldots, T\}$, $T \in \mathbb{N}$. Distance is understood as the difference between the addresses of the call of these functions and the typicality is determined by checking whether it belongs to the distance profile. In addition to the T parameter, the detection algorithm uses the parameter $W (\in \mathbb{N})$: it is the number of profiles against which the call is checked. In this case, for $j \in \{1, \ldots, W\}$, the profile with the number j is constructed from pairs of calls from a legitimate sequence, the difference of call indices in which is equal to j. The aim of this work is, on the one hand, to describe the detection algorithm and, on the other, to provide an experimental estimate of the sufficient values of the parameters W and T. As a result, in particular, it was found that the values of these parameters depend on the set of monitored functions; therefore, for each set of functions (and each protected program), these parameters must be found separately.

Keywords: software vulnerabilities; distance between function calls; program protection

DOI: 10.14357/08696527210405

References

1. Rohlf, C., and Y. Ivnitskiy. 2011. Attacking clientside JIT compilers. *Black Hat USA*. Available at: https://www.nccgroup.trust/globalassets/resources/us/presentations/documents/attacking_clientside_jit_compilers.pdf (accessed September 22, 2021).
2. Forrest, S., S. Hofmeyr, and A. Somayaji. 2008. The evolution of system-call monitoring. *Annual Computer Security Applications Conference Proceedings.* Piscataway, NJ: IEEE. 418–430. doi: 10.1109/ACSAC.2008.54.
3. Singh, A., R. Arora, and H. Pareek. 2017. Malware analysis using multiple API sequence mining CFG. arXiv.org. 12 p. Available at: https://arxiv.org/abs/1707.02691 (accessed September 18, 2021).
4. Gupta, S., H. Sharma, and S. Kaur. 2018. Malware characterization using Windows API call sequences. *J. Cyber Security Mobility* 7(4):363–378. doi: 10.1007/978-3-319-49445-6_15.

5. Snow, K. Z., F. Monrose, L. Davi, A. Dmitrienko, C. Liebchen, and A.-R. Sadeghi. 2013. Just-in-time code reuse: On the effectiveness of fine-grained address space layout randomization. *Symposium on Security and Privacy Proceedings*. Piscataway, NJ: IEEE. 574–588. doi: 10.1109/SP.2013.45.
6. Wagner, D., and P. Soto. 2002. Mimicry attacks on host-based intrusion detection systems. *9th Conference on Computer and Communications Security Proceedings*. New York, NY: ACM. 255–264. doi: 10.1145/586110.586145.
7. Nurmuhametov, A. R., E. A. Zhabotinskij, Sh. F. Kurmangaleev, S. S. Gajsarjan, and A. V. Vishnjakov. 2017. Melkogranulyarnaya randomizatsiya adresnogo prostranstva programmy pri zapuske [Fine-grained address space layout randomization on program load]. *Trudy ISP RAN* [ISP RAS Proceedings] 29(6):163–182. doi: 10.15514/ISPRAS-2017-29(6)-9.
8. Kosolapov, Y. V. 2020. On detecting code reuse attacks. *Autom. Control Comp. S.* 54:573–583. doi: 10.3103/S0146411620070111.

Received August 20, 2020

Contributor

Kosolapov Yury V. (b. 1982) — Candidate of Science (PhD) in technology, associate professor, Institute for Mathematics, Mechanics, and Computer Science named after I. I. Vorovich, Southern Federal University, 8a Milchakova Str., Rostov-on-Don 344090, Russian Federation; itaim@mail.ru

ИССЛЕДОВАТЕЛЬСКИЙ ПРОТОТИП КОГНИТИВНОЙ ГИБРИДНОЙ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ ДИАГНОСТИЧЕСКИХ РЕШЕНИЙ*

***С. Б. Румовская*[1], *И. А. Кириков*[2]**

Аннотация: Исследование когнитивных структур и механизмов специалиста (врача) позволит сделать процессы преодоления диагностической проблемы (ДП) видимыми, контрастными, а разработка на их основе систем с когнитивным моделированием ДП снизит число врачебных ошибок и повысит качество медицинских систем поддержки принятия решений. Рассмотрены когнитивная гибридная интеллектуальная диагностическая система (КГИДС), ее предметно-изобразительная модель и типовая архитектура, а также особенности лабораторной апробации на примере проблемы диагностики заболеваний поджелудочной железы.

Ключевые слова: когнитивная гибридная интеллектуальная диагностическая система; индивидуальное принятие диагностических решений; когнитивный образ объекта диагностики; оперативные образы объекта диагностики

DOI: 10.14357/08696527210406

1 Введение

По разным оценкам, примерно 50%–60% россиян хоть раз в жизни сталкивались с ошибками врачей [1, 2]. В среднем в 35% случаев имело место расхождение диагноза поликлинического и клинического. Специалисты в области психологии мышления [3] убеждены, что именно наличие двух способов представления информации (в виде последовательности символов и в виде картин-образов) обеспечивает сам феномен человеческого мышления. Образная система специалиста характеризуется высокой скоростью обработки образов. Sowa отмечает [4], что в человеческом мышлении огромную роль играют заготовки-схемы стандартных ситуаций, использование которых существенно ускоряет рассуждения. От оперативности реакции врача на возникающую проблемную диагностическую ситуацию зависит безопасность принимаемых решений. Визуализация позволяет перевести во внешне замечаемый план те содержания: образы, представления, и, соответственно, процессы с ними, — которые происходят у специалиста (врача) «в голове».

*Исследование выполнено при финансовой поддержке РФФИ (проект 19-07-00250А).

[1]Калининградский филиал Федерального исследовательского центра «Информатика и управление» Российской академии наук, sophiyabr@gmail.com

[2]Калининградский филиал Федерального исследовательского центра «Информатика и управление» Российской академии наук, baltbipiran@mail.ru

В работе рассмотрены:

(1) КГИДС как функциональная гибридная интеллектуальная система (ГиИС) поддержки принятия диагностических решений с имитацией интегрированного визуально-образного и вербально-знакового мышления эксперта и механизма решения проблем динамическим реструктурированием их как целого в систему — декомпозицию связанных задач как модель коллективного интеллекта, решающего ДП;

(2) ее предметно-изобразительная модель, типовая архитектура и алгоритм синтеза КГИДС, а также лабораторная апробация с КГИДС на примере проблемы диагностики заболеваний поджелудочной железы, в частности острого панкреатита.

2 Когнитивная гибридная интеллектуальная диагностическая система: определение и предметно-изобразительная модель

Когнитивная гибридная интеллектуальная диагностическая система определена на основе (1.5) из [5] как ГиИС, воспринимающая входные диагностические данные, выдающая выходные сигналы (диагноз) и находящаяся в некотором состоянии, релевантном состоянию коллективного диагностического процесса:

$$\alpha^u(t) =$$
$$= R^n_{19}\left({}^1X^0, \mathrm{МЕТ}^i\right) \wedge R^n_{12}\left({}^1X^0, {}^2\widehat{\mathbf{x}}^n_1\right) \wedge R^n_{12}\left({}^1X^0, {}^2\widehat{\mathbf{x}}^n_2\right) \wedge R^n_{12}\left({}^1X^0, {}^2\widehat{\mathbf{s}}^n\right) \wedge$$
$$\wedge {}^6R_{22}\left({}^2\widehat{\mathbf{s}}^n(t), {}^2\widehat{\mathbf{s}}^n(t+1)\right) \wedge {}^7R_{22}\left({}^2\widehat{\mathbf{x}}^n_1(t), {}^2\widehat{\mathbf{s}}^n(t)\right) \wedge {}^8R_{22}\left({}^2\widehat{\mathbf{s}}^n(t), {}^2\widehat{\mathbf{x}}^n_2(t)\right) \wedge$$
$$\wedge {}^{\psi\varphi}\ddot{R}^k_{11}\left({}^1X^0, {}^1X^n\right) \wedge {}^9R^n_{22}\left({}^2\widehat{\mathbf{x}}^n_1, {}^2X^n_1\right) \wedge {}^{10}R^n_{22}\left({}^2X^n_2, {}^2\widehat{\mathbf{x}}^n_2\right), \quad (1)$$

где $\mathrm{МЕТ}^i$ — интегрированный метод; ${}^2\widehat{\mathbf{x}}^n_1$ — вектор исходных данных ДП, передаваемый на вход одного или нескольких элементов гибрида α^u, решающих подзадачи из декомпозиции ДП; ${}^2\widehat{\mathbf{x}}^n_2$ — вектор выходных данных одного или нескольких элементов α^u — цель решения ДП; ${}^2\widehat{\mathbf{s}}^n$ — вектор состояния α^u, формирующийся из состояний «поведенческих» элементов с аналитическими, эволюционными, статистическими вычислениями и логическими рассуждениями, а также псевдосостояний элементов с нейро-, нечеткими вычислениями и рассуждениями на основе опыта; ${}^6R_{22}$, ${}^7R_{22}$ и ${}^8R_{22}$ — отношения функционирования гибрида, заданные в смежные моменты времени на множестве пар состояние–состояние, вход–состояние и на множестве пар состояние–выход соответственно; ${}^1X^n$ — множество знаков элементов α^h; ${}^{\psi\varphi}\ddot{R}^k_{11}$ — отношения интеграции элементов гибрида φ и ψ, знания которых участвуют в интеграции типа k (извлечение, включение и др.); ${}^2X^n_1$, ${}^2X^n_2$ — множество свойств «вход» и «выход» элементов из ${}^1X^n$ соответственно; ${}^9R^n_{22}$ и ${}^{10}R^n_{22}$ — отношения на

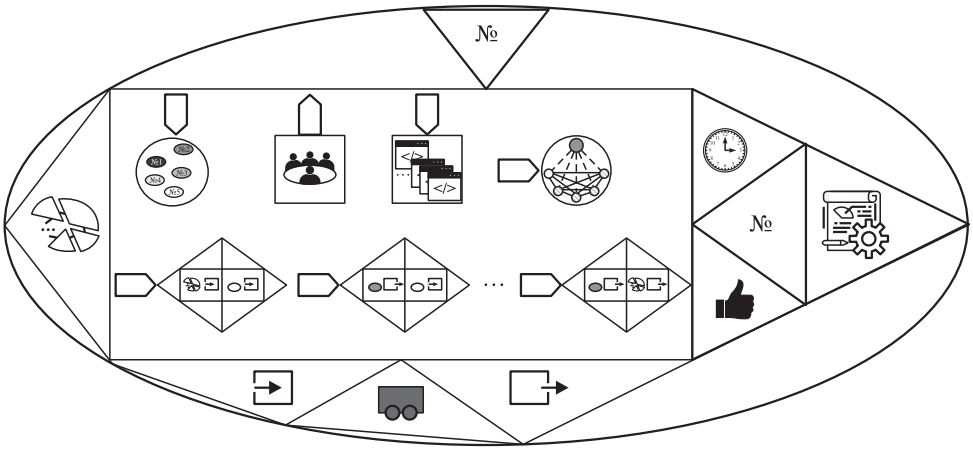


Рис. 1 Предметно-изобразительная модель КГИДС

множестве пар «вход ГиИС – входы элементов» и «выходы элементов – выход ГиИС» соответственно.

Гетерогенное модельное поле (ГМП) включает множество разнородных моделей (каждая реализуется как элемент гибрида), отображающих теоретические, профессиональные знания и опыт экспертов (врачей) и необходимых для учета, контроля, обсуждения и для высказывания своего мнения относительно решения подзадач из состава ДП, а также множество моделей, решающих подпроблемы формирования когнитивных образов (КО) экспертов-врачей узкой специализации, лиц, принимающих решения (ЛПР), и объекта диагностики [6].

Интегрированный метод в (1) — результат процесса синтеза метода решения (СМР) ДП. Синтез метода решения ДП можно определить как процесс моделирования рассуждений ЛПР, в результате которого над ГМП строится интегрированная модель, релевантная диагностической ситуации.

На рис. 1 представлена предметно-изобразительная модель КГИДС, релевантная (1). В центре модели КГИДС (овала) на рис. 1 располагается стрелка, образуя ролевое визуальное отношение «агрегат–действие», что подчеркивает деятельностный подход к репрезентации системы. Над ресурсной частью стрелки располагается свойство КГИДС — идентификатор системы.

Ресурсная (прямоугольная) составляющая действия (стрелки) по формированию системы включает следующие ролевые визуальные отношения:

– «действие–объект», где объект — ГМП; «действие–субъект», где правая роль — ресурс «коллектив экспертов»; «действие–объект», где правая роль у ресурса «множество программ»; «действие–результат», где правая роль — архитектура КГИДС, релевантная структуре ДП;

– «действие–свойство», которое выполняется после того, как установлена связь между входом КГИДС и входом одного из ее элементов. Данную связь символизируют сомкнутые основания треугольников — свойств. Каждое свойство состоит в ролевом визуальном отношении «свойство–ресурс» (РВО СР): слева — свойство «вход» и ресурс — КГИДС (прямоугольник). Аналогичное правое визуализирует высказывание «вход элемента КГИДС»;

– «действие–свойство» (на рис. 1 отображено многоточием — их число равно размерности ГМП), которое выполняется после того, как установлена связь между выходом одного элемента КГИДС и входом другого. Каждое свойство состоит в РВО СР: слева — свойство «выход» и ресурс — элемент. Аналогичное правое визуализирует высказывание «вход элемента КГИДС»;

– «действие–свойство», которое выполняется после того, как установлена связь между выходом элемента КГИДС и ее выходом. Каждое свойство состоит в РВО СР: слева — свойство «выход» и ресурс — элемент. Аналогичное правое визуализирует высказывание «выход КГИДС».

Под ресурсной частью стрелки расположены свойства КГИДС, которые схематизируют три ролевых визуальных отношения «агрегат–свойство» (слева направо): «агрегат–вход»; «агрегат–состояние»; «агрегат–выход». Треугольная часть стрелки отображает содержимое ролевого визуального отношения «действие–свойство»: «действие–время» («иметь время начала», «иметь время окончания»); «действие–имя» («иметь имя»); «действие–характеристика» («иметь характеристику»); «действие–оценка» («иметь оценку»). Пиктограмма характеристики здесь раскрывает методологический аспект конструирования КГИДС.

Рассмотрим далее архитектуру и алгоритм синтеза КГИДС, а также особенности лабораторной апробации.

3 Архитектура и алгоритм синтеза когнитивной гибридной интеллектуальной диагностической системы

Архитектура КГИДС релевантна структуре диагностической проблемы, для решения которой она предназначена. Типовая архитектура КГИДС (рис. 2) включает:

(1) интерфейс пользователя;

(2) функциональные элементы (ФЭ), решающие множество задач учета, контроля, а также диагностические подпроблемы (ДПП), решаемые экспертами и ЛПР;

(3) технологические элементы (ТЭ), решающие задачу предобработки информации и задачи формирования КО экспертов, ЛПР и объекта диагностики;

(4) хранилище предметно-изобразительных моделей.

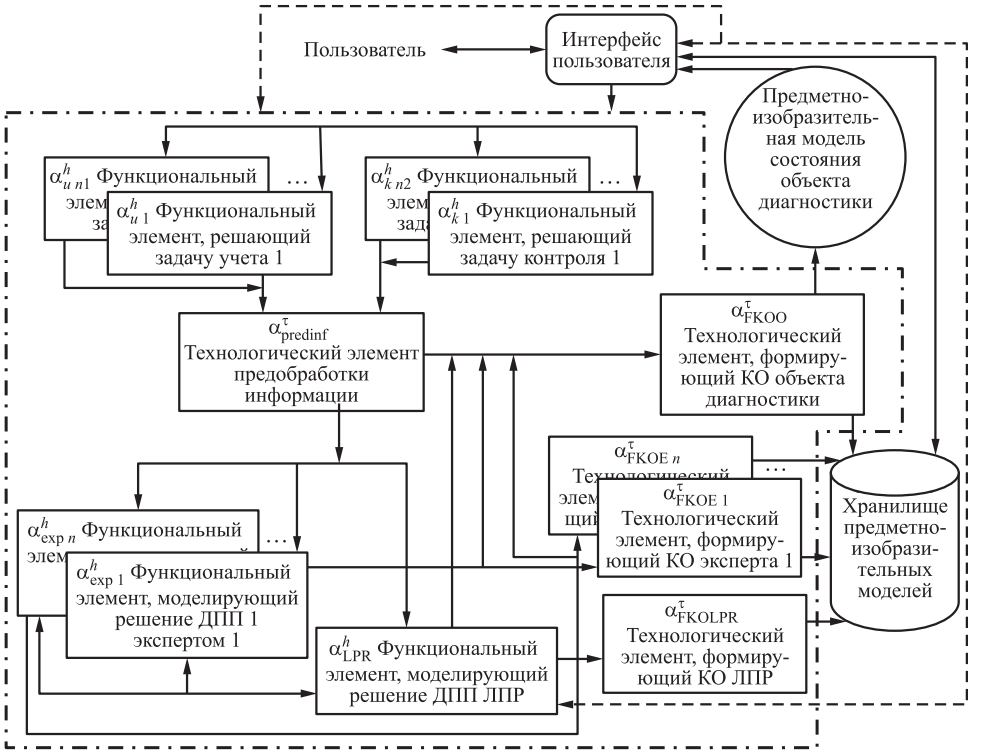


Рис. 2 Типовая архитектура КГИДС

Минимальная мощность множеств ФЭ, решающих задачи учета и контроля, равна 1. Элементы КГИДС взаимодействуют и обмениваются информацией, в том числе визуальной. Передавая друг другу частные решения в графическом, вербально-символьном или комбинированном виде, элементы КГИДС модифицируют изначальный вариант декомпозиции проблемы, конструируют интегрированное решение посредством запуска встроенного алгоритма синтеза КГИДС в ФЭ, моделирующий решение ДПП ЛПР.

Результат работы алгоритма модифицирует состав и связи элементов КГИДС и при необходимости запрашивает дополнение и/или обновление значений множества показателей состояния здоровья объекта, вводимых через интерфейс пользователя. Алгоритм может быть запущен при появлении дополнительной информации о состоянии здоровья объекта, введенной пользователем через интерфейс, что учитывает динамический характер ДП и синтез КГИДС, релевантной проблеме в момент ее решения. Технологические элементы, решающие задачи формирования КО экспертов, ЛПР и объекта диагностики, реализованы как статические продукционные экспертные системы (ЭС) с включением в правые части

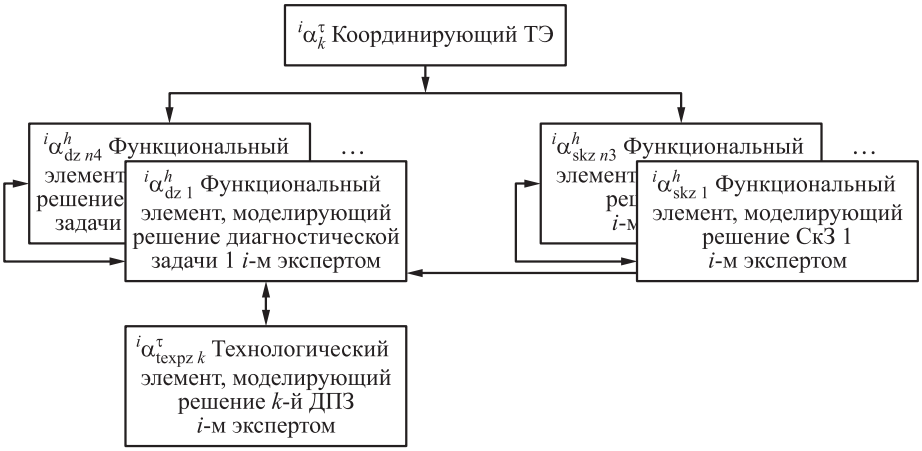


Рис. 3 Типовая структура сложного ФЭ, моделирующего решение ДПП экспертом

баз знаний пиктограмм и базисных визуальных символов, спроектированных аналогично ЭС, предложенной в [7].

Предметно-изобразительная модель отображается на интерфейсе пользователя. Результаты работы ТЭ, решающих задачи формирования КО экспертов, ЛПР и объекта диагностики, сохраняются в хранилище предметно-изобразительных моделей и могут быть при необходимости отображены на интерфейсе пользователя, а также использоваться моделями ГМП. Функциональные элементы, моделирующие решение ДП экспертами и ЛПР, могут быть составными и иметь архитектуру, показанную на рис. 3.

Архитектура на рис. 3 включает:

(1) координирующий ТЭ;

(2) множество ФЭ, решающих диагностические задачи, и ТЭ, решающие диагностические подзадачи (ДПЗ) («диагностическая задача» — более крупная, а «диагностическая подзадача» — более мелкая, связанная с решением крупной. Например, чтобы подтвердить предварительный диагноз, необходимо сделать и проанализировать ультразвуковое исследование брюшной полости, общие анализы и биохимию);

(3) множество ФЭ, решающих скомпенсированные затруднения (СкЗ) (часто встречающиеся трудности — диагностические задачи, для решения которых есть четкий и однозначный протокол обследования).

Метод решения диагностической проблемы конструируется в текущий момент времени ее решения. Пусть получены:

(1) декомпозиция $\widehat{p}_D = \langle P^h, {}^3R^n_{88} \rangle$ ДП в виде сети — ориентированного графа без петель. Вершины p^h_i, $i = [1, N_h]$, — однородные диагностические

задачи (функциональные и технологические), а ребра ${}^{3}r^{n}_{88}(p^{h}_{i}, p^{h}_{j}) \in \widehat{p}_{D}$, $i, j = [1, N_h]$, $i \neq j$, — отношения между ними;

(2) ГМП M^a и таблица гибридных стратегий T^u для декомпозиции $\widehat{p}_{D}$, которая тройкам ${}^{88}r^{n}_{3}(p^{h}_{i}, p^{h}_{j})$ задает взаимно однозначные соответствия — $\left\{ {}^{3}r^{n}_{88}\left(p^{h}_{\psi}, p^{h}_{\varphi}\right) \leftrightarrow {}^{\psi\varphi}\ddot{R}^{k}_{11} \right\}$, где ${}^{4}R^{n}_{88}$ — множество отношений соответствия целей и исходных данных задач из декомпозиции ДП, а также целей и исходных данных ДП;

(3) множества интерпретаторов I^a автономных моделей mod^{a}_{x} и межмодельных интерфейсов ${}^{xy}\zeta$ для всех mod^{x}_{a} и mod^{a}_{y}, для которых установлены отношения интеграции;

(4) соответствия моделей $\mathrm{mod}^{a} \in \mathrm{MOD}^{a}$ и интерпретаторов $I^{\eta} \in I^{a} | \eta = \{a, s, e, f, n, g, p\}$, где a — аналитические методы; s — статистические; e — ЭС; f — нечеткие системы; n — искусственные нейронные сети; g — генетические алгоритмы; p — методы рассуждения на основе опыта;

(5) алгоритмы преобразования представления информации с одного языка описания модели на другие;

(6) множество $S^x \in S$, $x \in [1, N_{\mathrm{ms}}]$ допустимых состояний $S^x = \{s^{x}_{\delta}\}$, $\delta \in [1, N_{sx}]$ для каждой модели mod^{a}_{x} и порядок на множестве P^h — множество $\mathrm{In} = \{\mathrm{In}^{1}, \ldots, \mathrm{In}^{\mathrm{NT}}\}$, где $\mathrm{In}^{1} = \{\mathrm{in}_{1}, \ldots, \mathrm{in}_{\mathrm{In}^{1}}\}$ и $\mathrm{In}^{\mathrm{NT}} = \{\mathrm{in}_{1}, \ldots, \mathrm{in}_{\mathrm{In}^{\mathrm{NT}}}\}$ — нижние индексы задач из P^h, решаемых в первую и последнюю очередь соответственно.

Выходная информация алгоритма:

(1) список Ls^{D} — множество отношений интеграции на моделях с учетом их состояния ${}^{1}R_{12}\left(\mathrm{mod}^{a}_{x}, s^{x}_{\delta}\right) \circ {}^{xy}\ddot{R}^{k}_{11}\left(\mathrm{mod}^{a}_{x}, \mathrm{mod}^{a}_{y}\right)$, $x, y \in [1, N_{\mathrm{ms}}]$;

(2) множество E матриц «модель–модель» $E_j \in E^{l}_{j}$, $l \in [1, \mathrm{NT}]$, строки которых — mod^{a} в состоянии s^{x}_{δ}, решающие задачу p^{h}_{ψ}, а столбцы — mod^{a}, решающие задачу p^{h}_{φ}.

Алгоритм конструирования КГИДС включает в себя следующие шаги:

1. Начало, присвоить $l = 1$.
2. Выбрать из $\widehat{p}_{D}$ множество троек $P^{h}_{\beta} = \left\{ {}^{3}r^{n}_{88}\left(p^{h}_{\psi}, p^{h}_{\varphi}\right) \right\}$, $\beta = [1, N^{l}_{h}]$, в которых каждая p^{h}_{φ} имеет очередность из In^{l}.
3. Присвоить $j = 1$, $E^{l} = \emptyset$.
4. Выбрать очередную j-ю пару из P^{h}_{β}.

5. Инициализировать для j-й пары матрицу E^l_j, помечая столбцы ω моделями mod^a_φ задачи p^h_ψ, а строки — моделями mod^a_ψ в состоянии s^x_δ задачи p^h_φ. Дополнить E^l матрицей E^l_j.
6. $j = N_h + 1$? Если «нет», то присвоить $j = j + 1$ и перейти к п. 4.
7. Присвоить $j = 1$.
8. Выбрать E^l_j из E^l. Активируются только те ее элементы, для которых между моделями заданы отношения интеграции. Используя ЭС, модели оцениваются. Оценки mod^a_ψ заносятся в знаменатели, а mod^a_φ — в числители элементов. Для начальных состояний моделей выбрать пару моделей с максимальным значением интегрированной оценки и занести ее в список Ls^D.
9. $j = N_h + 1$? Если «нет», то присвоить $j = j + 1$ и перейти к п. 8.
10. $l = \mathrm{NT}$? Если «нет», то присвоить $l = l + 1$ и перейти к п. 2, иначе дополнить Ls^D интерпретаторами и интерфейсами и сформировать базу знаний ФЭ, моделирующего решение ДПП ЛПР, из матриц E, чтобы он мог перестраивать интегрированную модель КГИДС в зависимости от ситуации и сочетать как символьные рассуждения, так и визуальные.

Лабораторная апробация архитектурных решений КГИДС для задачи диагностики заболеваний поджелудочной железы, в частности острого панкреатита, выполнена на материалах Калининградской областной клинической больницы. Гетерогенное модельное поле из 8 моделей, хранилище и интерфейс пользователя реализованы с помощью MATLAB-Simulink, графовой базы данных Dgraph и JavaScript.

4 Заключение

Разработка КГИДС с когнитивным моделированием ДП позволит подойти к сглаживанию противоречивых показаний аппаратуры, повысит качество и, соответственно, безопасность медицинских интеллектуальных систем поддержки принятия диагностических решений. Предложенные в данной работе КГИДС (1) синтезируют новый метод решения ДП при каждом ее возникновении, что релевантно ее меняющимся условиям и структуре; (2) оценивают совместную работу моделей, синтезируя метод как целостную систему; (3) качественно улучшают рассуждения за счет интеграции чувственного и рационального, единичного и всеобщего, а также презентации объекта диагностики врачу в многообразии свойств и отношений при сохранении ясности картины для перехода от одного действия к другому.

Литература

1. *Ласковец Е. А.* Статистика по видам врачебных ошибок. https://rospravomed.ru/otrasli-prava/meditsinskoe-pravo/vrachebnaya-oshibka/statistika-po-vidam-vrachebnyh-oshibok.

2. Анамнез ошибки. https://www.novayagazeta.ru/articles/2017/03/13/71761.
3. *Артищева Л. В.* Проблема образа психического состояния и субъективного (ментального) опыта человека в отечественной и зарубежной литературе // Мат-лы V Зимней школы по психологии состояний / Под ред. М. Г. Юсупова. — Казань: Изд-во Казанского ун-та, 2011. С. 5–9.
4. *Sowa J. F.* Conceptual structures — information processing in mind and machine. — The systems programming ser. — Reading, MA, USA: Addison-Wesley, 1984. 481 p.
5. *Румовская С. Б.* Исследование методов поддержки принятия коллективных диагностических решений и разработка инструментальных средств «Виртуальный консилиум» (на примере диагностики артериальной гипертензии): Дис. . . . канд. техн. наук. — М., 2017. 138 с.
6. *Румовская С. Б.* Редукция диагностической проблемы с когнитивной визуализацией ее элементов // Гибридные и синергетические интеллектуальные системы: Мат-лы V Всеросс. Поспеловской конф. с междунар. участием / Под ред. А. В. Колесникова. — Калининград: БФУ им. И. Канта, 2020. С. 242–251.
7. *Румовская С. Б., Колесников А. В., Литвин А. А.* Репрезентация методов решения подзадач разного типа из декомпозиции диагностической проблемы // Вестник Балтийского федерального университета им. И. Канта. Сер. Физико-математические и технические науки, 2020. № 2. С. 62–73.

Поступила в редакцию 31.08.21

RESEARCH PROTOTYPE OF A COGNITIVE HYBRID INTELLIGENT SYSTEM FOR SUPPORTING DIAGNOSTIC DECISION-MAKING

S. B. Rumovskaya and I. A. Kirikov

Kaliningrad Branch of the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation

Abstract: The study of the cognitive structures and mechanisms of a specialist (doctor) will make the processes of overcoming the diagnostic problem visible and contrasting. Systems with cognitive modeling of the diagnostic problem developed on their basis will reduce the number of medical errors and improve the quality of medical decision support systems. The paper considers the cognitive hybrid intelligent diagnostic system, its subject-visual model, and typical architecture as well as the features of laboratory testing by the example of the problem of diagnosing diseases of the pancreas.

Keywords: cognitive hybrid intelligent diagnostic system; individual diagnostic decision making; cognitive image of the diagnostic object; operational images of the diagnostic object

DOI: 10.14357/08696527210406

Acknowledgments

The reported study was funded by RFBR according to the research project No. 19-07-00250A.

References

1. Laskovets, E. A. Statistika po vidam vrachebnykh oshibok [Statistics on the types of medical errors]. Available at: https://rospravomed.ru/otrasli-prava/meditsinskoe-pravo/vrachebnaya-oshibka/statistika-po-vidam-vrachebnyh-oshibok (accessed September 22, 2021).
2. Anamnez oshibki [History of error]. Available at: https://www.novayagazeta.ru/articles/2017/03/13/71761 (accessed July 30, 2021).
3. Artishcheva, L. V. 2011. Problema obraza psikhicheskogo sostoyaniya i sub"ektivnogo (mental'nogo) opyta cheloveka v otechestvennoy i zarubezhnoy literature [The problem of the image of the mental state and the subjective (mental) experience of a person in domestic and foreign literature]. *Mat-ly V Zimney shkoly po psikhologii sostoyaniy* [5th Winter School on Psychology of Conditions Proceedings]. Kazan. 5–9.
4. Sowa, J. F. 1984. *Conceptual structures — information processing in mind and machine*. The systems programming ser. Reading, MA: Addison-Wesley. 481 p.
5. Rumovskaya, S. B. 2017. Issledovanie metodov podderzhki prinyatiya kollektivnykh diagnosticheskikh resheniy i razrabotka instrumental'nykh sredstv "Virtual'nyy konsilium" (na primere diagnostiki arterial'noy gipertenzii) [The research of methods of collective diagnostic decision support and development of instruments of the "Virtual council" (illustrated with the diagnostic of arterial hypertension)]. Moscow. PhD Diss. 138 p.
6. Rumovskaya, S. B. 2020. Reduktsiya diagnosticheskoy problemy s kognitivnoy vizualizatsiey ee elementov [Reduction of a diagnostic problem with cognitive visualization of its elements]. *Gibridnye i sinergeticheskie intellektual'nye sistemy: mat-ly V Vseross. Pospelovskoy konf. s mezhdunar. uchastiem* [Hybrid and Synergetic Intelligent Systems: 5th All-Russian Pospelovsky Conference with International Participation Proceedings]. Kaliningrad: Publishing House of Immanuel Kant Baltic Federal University. 242–251.
7. Rumovskaya, S. B., A. V. Kolesnikov, and A. A. Litvin. 2020. Reprezentatsiya metodov resheniya podzadach raznogo tipa iz dekompozitsii diagnosticheskoy problemy [Representation of methods for solving subtasks of different types from the decomposition of a diagnostic problem]. *Vestnik IKBFU. Physics, Mathematics, and Technology* 2:62–73.

Received August 31, 2021

Contributors

Rumovskaya Sophiya B. (b. 1985) — Candidate of Science (PhD) in technology, scientist, Kaliningrad Branch of the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation; sophiyabr@gmail.com

Kirikov Igor A. (b. 1955) — Candidate of Science (PhD) in technology, director, Kaliningrad Branch of the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation; baltbipiran@mail.ru

ОПТИМИЗАЦИЯ АППАРАТНОЙ ПОДДЕРЖКИ БЫСТРОГО ПРЕОБРАЗОВАНИЯ ФУРЬЕ В РЕКУРРЕНТНОМ СИГНАЛЬНОМ ПРОЦЕССОРЕ*

***Д. В. Хилько*[1], *Ю. А. Степченков*[2], *Ю. И. Шикунов*[3], *Ю. Г. Дьяченко*[4], *Г. А. Орлов*[5]**

Аннотация: Рассматривается поддержка быстрого преобразования Фурье (БПФ, *англ.* FFT — fast Fourier transform) в гибридной архитектуре рекуррентного обработчика сигналов (ГАРОС). Приводится анализ существующей реализации. Выявляются недостатки и их последствия. Предлагается оптимизированное решение, направленное на упрощение масштабирования как архитектуры, так и числа отсчетов БПФ.

Ключевые слова: цифровая обработка сигналов; быстрое преобразование Фурье; цифровой сигнальный процессор; Radix-2

DOI: 10.14357/08696527210407

1 Введение

В области цифровой обработки сигналов (ЦОС) приняты две наиболее распространенные процедуры: цифровая фильтрация и дискретное преобразование Фурье (ДПФ). «ДПФ позволяет анализировать, преобразовывать и синтезировать сигналы такими способами, которые невозможны при непрерывной (аналоговой) обработке» [1]. Дискретное преобразование Фурье используется практически во всех инженерных областях, а также в физике и технике.

Сама математическая процедура ДПФ очень неэффективна: требует проведения очень большого числа комплексных умножений, что делает прямое вычисление ДПФ нецелесообразным. Поэтому в 1965 г. Кули и Тьюки предложили алгоритм вычисления ДПФ [2], который известен как БПФ. Применение

*Исследование выполнено при поддержке Российского научного фонда (проект 19-11-00334).

[1]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, dhilko@yandex.ru

[2]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, YStepchenkov@ipiran.ru

[3]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, YIShikunov@gmail.com

[4]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, diaura@mail.ru

[5]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, orlov.jaja@gmail.com

БПФ сделало возможным проведение Фурье-анализа с помощью цифровых сигнальных процессоров (ЦСП).

Дальнейшее развитие алгоритма БПФ привело к появлению целого семейства алгоритмов: Radix-2, Radix-22, Radix-4 и др. Однако даже с учетом значительно более высокой скорости вычисления алгоритм БПФ все еще требует значительных временны́х затрат, особенно при увеличении разрешающей способности алгоритма (512 отсчетов и больше). Поэтому современный ЦСП должен предоставлять набор инструментов для эффективного вычисления БПФ на широком наборе разрешающих способностей.

Рассматриваемая в статье ГАРОС и ее ключевые особенности (управление потоком самодостаточных данных и рекуррентность) представлены в [3, 4]. Синтезированный на ее основе ПЛИС-прототип (ПЛИС — программируемая логическая интегральная схема) [5] представляет собой ЦСП общего назначения, а его валидация осуществлялась на задаче распознавания изолированных слов. В состав архитектуры были введены механизмы поддержки вычисления БПФ [4]. Однако при увеличении разрешающей способности БПФ оказалось, что накладные аппаратные расходы слишком велики.

Цель статьи — представление результатов оптимизации архитектуры ГАРОС и сокращения ее аппаратных затрат в рамках ПЛИС при сохранении эффективности реализации БПФ с произвольным масштабированием его числа отсчетов.

2 Текущая поддержка в гибридной архитектуре рекуррентного обработчика сигналов

2.1 Описание алгоритма быстрого преобразования Фурье

Алгоритм БПФ — эффективный способ вычисления ДПФ с числом отсчетов, равным натуральным числам во второй степени. Дискретное преобразование Фурье $X(k)$, где $k = 0, \ldots, N-1$, имеет вид:

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \left[\cos\left(\frac{2\pi nk}{N}\right) - i \sin\left(\frac{2\pi nk}{N}\right) \right] .$$

Поворотные коэффициенты размещены на единичной окружности в комплексной плоскости. Они симметричны и периодичны, что позволяет существенно сократить число умножений, требуемых для проведения ДПФ.

Алгоритм вычисления БПФ по основанию 2 (Radix-2) разделяет проведение ДПФ на серию 2-точечных ДПФ. Каждое такое преобразование называется типовой операцией «бабочка». Для работы алгоритма требуется, чтобы число отсчетов N было натуральной степенью двойки $N = 2^s$, $s \in \mathbb{N}$. Тогда для вычисления БПФ потребуется провести s стадий.

Результаты вычисления на каждой стадии могут быть сохранены в тех же самых ячейках памяти, которые изначально хранили исходные входные отсчеты.

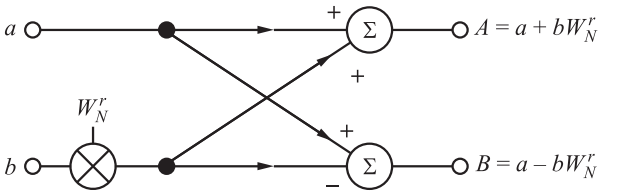


Рис. 1 Операция Radix-2 DIT «бабочка» [6, рис. 5.13]

Быстрое преобразование Фурье с прореживанием по времени (Decimation-in-time, DIT) вычисляет стадии, используя типовую операцию «бабочка», приведенную на рис. 1. Для корректной работы алгоритма необходимо загружать входные отсчеты в бит-реверсированном порядке.

Формулы вычисления комплексного БПФ DIT имеют вид:

$$\begin{aligned} R_a &= R_a + (R_a R_W - I_b I_w) = R_a + (C - D) = R_a + G; \\ I_a &= I_a + (R_b I_W + I_b R_w) = I_a + (E + F) = I_a + H; \\ R_b &= R_b + (R_a R_W - I_a I_w) = R_b - (C - D) = R_b - G; \\ I_b &= I_b + (R_a I_W + I_a R_w) = I_b - (E + F) = I_b - H, \end{aligned}$$

где R — действительная, а I — мнимая часть комплексного числа.

2.2 Описание существующей реализации

Текущая версия прототипа ГАРОС содержит средства аппаратной поддержки вычисления БПФ, которые обеспечивают выполнение 256-точечного Radix-2 DIT алгоритма в формате фиксированной точки с записью результатов на место входных данных (in-place-реализация).

Данные средства архитектуры характеризуются следующими особенностями:

- используется механизм упаковки четырех входных 16-битных данных в одном операнде, действительные и мнимые части — в разных разделах капсулы;
- упакованные данные хранятся в бит-реверсивном порядке;
- поворотные коэффициенты хранятся в блоке «Память констант секционная» (ПК_С): 128 действительных и 128 мнимых 16-битных констант;
- содержит 4 параллельных секции: 4 вычислительных блока (ВБ) и 4 копии ПК_С;
- в систему команд введена специальная инструкция «Butterfly», которая обеспечивает четырехстадийное вычисление Radix-2 DIT «бабочки».

Ключевой элемент аппаратной поддержки БПФ — инструкция «Butterfly». Эта инструкция на аппаратном уровне интерпретируется ВБ как готовый

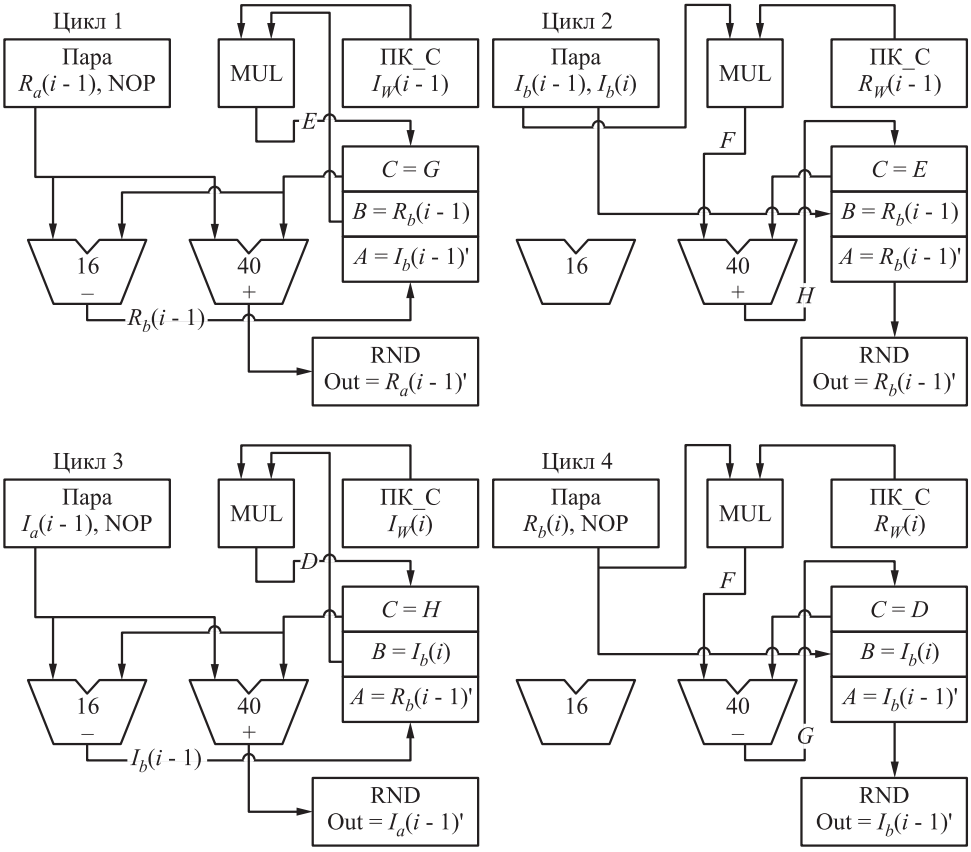


Рис. 2 Существующая схема инструкции «Butterfly»

четырехцикловый сценарий функционирования: каждому циклу выполнения инструкции определена своя схема вычислений. На рис. 2 представлена схема четырехцикловой инструкции: в установившемся режиме после наполнения конвейера «бабочка» вычисляется за 4 цикла.

Метрика Instruction Level Parallelism (ILP) для ВБ ГАРОС имеет значение 4 (1 умножитель, 1 блок сдвига и округления, 1 арифметико-логическое устройство 16-битное, 1 арифметическое устройство 40-битное). За 4 цикла ВБ может максимально выполнить 16 операций. Данная инструкция обеспечивает выполнение 14 операций. Коэффициент эффективности использования ILP составляет 87,5%.

Данное решение было апробировано на моделях архитектуры различного уровня, а также на ПЛИС-прототипе ГАРОС в процессе предваритель-

ной оценки производительности набора алгоритмов из BDTIMark2000 [7]. Несмотря на то что эффективность вычисления БПФ на одной секции ГАРОС сравнима с TMSC55x, данное решение требует аппаратной оптимизации.

2.3 Проблемы аппаратной поддержки быстрого преобразования Фурье в гибридной архитектуре рекуррентного обработчика сигналов

Основные проблемы аппаратной поддержки БПФ в ГАРОС:

(1) избыточные накладные расходы хранения поворотных коэффициентов, связанные с особенностями разделенной по секциям ПК_С;

(2) хранение отсчетов в капсуле в упакованных операндах.

Упаковка данных снижает избыточность тегированных данных, что является несомненным преимуществом. Однако на последних двух стадиях БПФ контроллеру буферной памяти (БП) приходится считывать и записывать части упакованных данных (вместо целого операнда) по разным адресам. Это приводит к усложнению схемы памяти и алгоритма вычисления адресов, росту накладных расходов и потенциальному снижению частоты ее работы;

(3) нецикличная схема считывания действительных и мнимых частей отсчетов.

Из рис. 3 видно, что входные отсчеты считываются в следующей последовательности:

Цикл 1: $R_a(i)$. Цикл 2: $I_b(i)$, а $I_b(i+1)$ уже должен быть в памяти операндов **заранее**!

Цикл 3: $I_a(i)$. Цикл 4: $R_b(i+1)$.

Как видно, действительные и мнимые части считываются непоследовательно и даже из разных «бабочек» в ходе одного прогона инструкции. Это осложняет алгоритм вычисления адресов в контроллере БП;

(4) цикл № 2 инструкции «Butterfly» требует специального режима распаковки и рассылки упакованных данных. В ГАРОС за распаковку и рассылку упакованных данных отвечает компонент «Распределитель». Особенность цикла № 2 заключается в том, что на вход вычислительного блока должны прийти две мнимые части i-й и $(i+1)$-й «бабочек». Это поведение выбивается из основной схемы рассылки данных, что требует ввода специального режима, который используется в ГАРОС только в рамках одного алгоритма;

(5) изменение числа отсчетов БПФ требует переработки капсулы.

Анализ проблем текущей версии средств аппаратной поддержки БПФ в целом свидетельствует о большой степени аппаратной избыточности. Поэтому оптимизация архитектуры ГАРОС является актуальной задачей.

Можно выделить два основных направления оптимизации архитектуры: модификация инструкции «бабочка» и модификация БП.

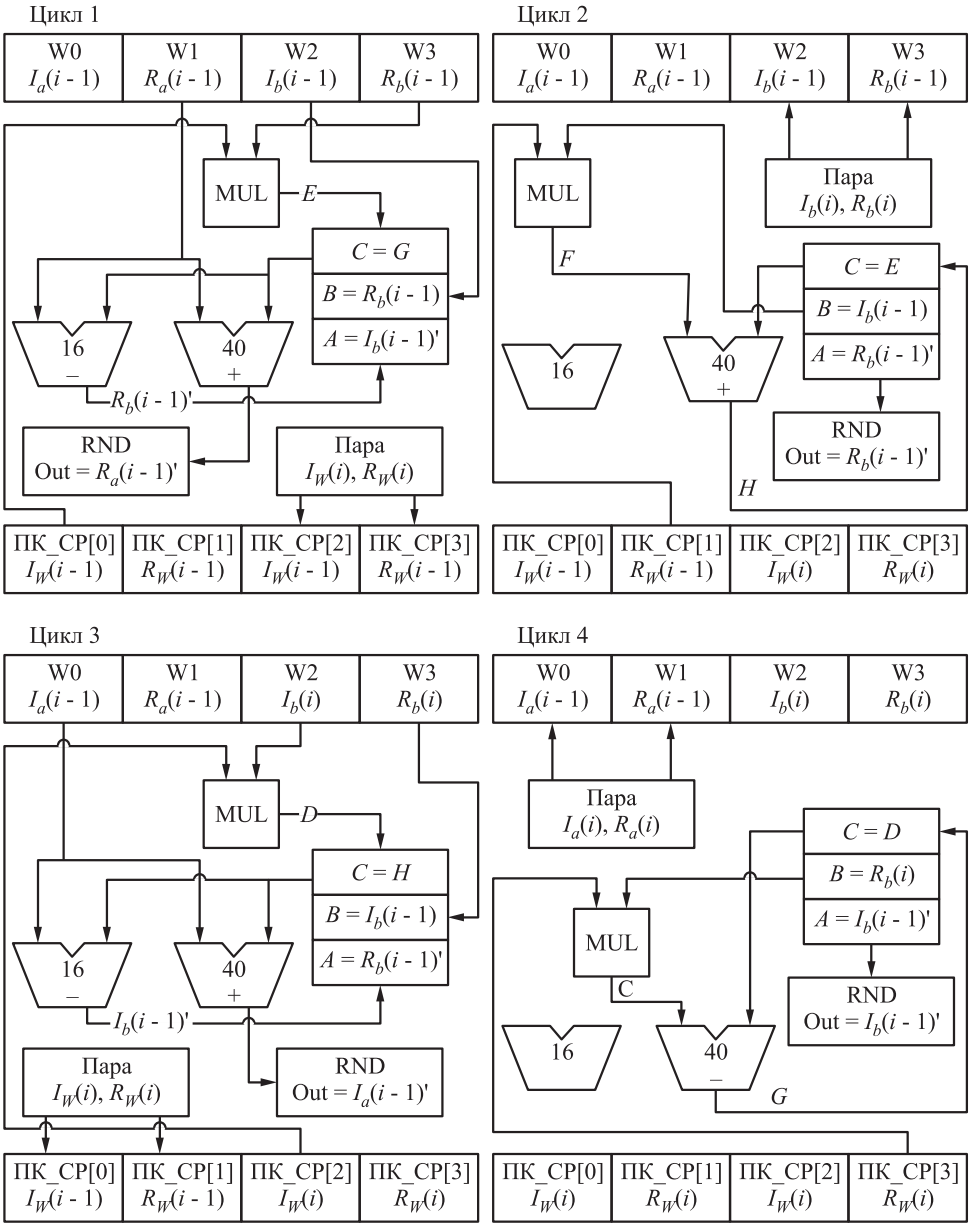


Рис. 3 Усовершенствованная схема инструкции «Butterfly»

3 Модифицированные средства аппаратной поддержки быстрого преобразования Фурье

3.1 Усовершенствование конвейеризованной инструкции

В предыдущем разделе был сделан вывод о необходимости переработки четырехстадийной инструкции «Butterfly». При этом технически ее характеристики не должны ухудшиться с точки зрения времени исполнения. Кроме того, переработанная схема инструкции должна также решать проблему 3, обозначенную в разд. 2.3. На рис. 3 представлены результаты переработки инструкции «бабочка».

В новой версии схемы инструкции поочередно считываются константы и данные, причем для циклов 1 и 3 считываются одни и те же константы, а данные считываются только для i-й бабочки (в отличие от предыдущей версии, где и для $(i+1)$-й). Таким образом, проблема 3 и, следовательно, проблема 4 решаются при внедрении данного решения.

С точки зрения использования ILP предлагаемое решение также обеспечивает выполнение 14 инструкций из 16 за 4 цикла и имеет коэффициент эффективности использования ILP, равный 87,5%.

3.2 Доработка вычислительных блоков

Для обеспечения возможности исполнения новой схемы инструкции «Butterfly» в состав вычислительных блоков ГАРОС необходимо ввести дополнительные входные пользовательские регистры W2 и W3. Таким образом регистровый файл будет расширен до четырех регистров. Для корректного выполнения БПФ исходной инициализации регистрового файла не требуется. Также в состав вычислительного блока входит память констант секционная регистровая (ПК_СР), имеющая объем в четыре 16-битных константы. В новой реализации необходимо обеспечить возможность записи сразу двух констант (ранее можно было записать только одну).

3.3 Модификация распределителя

Так как проблемы 3 и 4 решаются за счет изменения схемы инструкции, то из распределителя можно удалить поддержку специализированного режима распаковки операндов. Кроме того, особенности функционирования модифицированного контроллера БП требуют, чтобы распределитель на завершающей стадии вычисления алгоритма БПФ не замораживал конвейер, а генерировал пустые горсти операндов (что равносильно NOP-операциям).

3.4 Модификация буферной памяти

Самым существенным изменениям подверглась БП и ее контроллер. В состав БП необходимо ввести новый автономный банк памяти и блок управления

этой памятью, который включает в себя два работающих параллельно подблока: генератор адресов и контроллер чтения-записи.

1. **Блок памяти** состоит из блоков: IBlock и RBlock хранят мнимые и действительные части отсчетов преобразуемого сигнала соответственно; IwBlock и RwBlock хранят мнимые и действительные части поворотных коэффициентов W соответственно. Входные последовательности отсчетов должны быть записаны в бит-реверсированном порядке.

   Отсчеты имеют формат Q15, поэтому хранятся в 16-битных словах. Объем IBlock равен объему RBlock и составляет 1024 16-битных слова. Данные в IBlock и RBlock записываются со стороны управляющего устройства, поэтому в VHDL-реализации должен быть предоставлен соответствующий интерфейс доступа для записи и чтения данных блоков (так как in-place-реализация). Режим записи отсчетов как пакетный, так и одиночный.

   Поворотные коэффициенты также имеют формат Q15. Объем IwBlock равен объему RwBlock и составляет 512 16-битных слова. Эти коэффициенты являются константами и рассчитываются заранее. Запись констант осуществляется на этапе прошивки ПЛИС.

2. **Контроллер памяти для режима FFT** обеспечивает параллельное функционирование подблоков вычисления адресов и управления чтением и записью. Принимает данные от компонента «Импликатор» ГАРОС и передает их для записи в блок памяти.

   Инициализация блока осуществляется в момент считывания основным контроллером БП Acm: операнда, который имеет @Cdm = fft (т. е. перенастраивает режим памяти). После этого от операционного уровня архитектуры требуется подтверждение о том, что Acm: операнд получен и все нужные горсти сформированы, которое и инициирует выдачу операндов. После того как получено подтверждение от операционного уровня, начинается основной цикл функционирования.

   Размер Radix2 DIT БПФ определяется по значению поля количества итераций @Ci. Число стадий БПФ напрямую связано с количеством отсчетов по формуле «количество стадий» = $\mathrm{Log}_2(\mathrm{N})$. Тогда, например, для @Ci = 8 $N = 256$. Предлагаемая реализация БПФ контроллера в модели поддерживает настраиваемое количество секций исполнения алгоритма. Однако пока алгоритм рассчитан на 4 секции.

3. **Блок вычисления адресов** должен за 4 цикла вычислить адреса действительных и мнимых частей отсчетов, а также требуемых констант для текущих «бабочек». Особенность алгоритма заключается в том, что IBlock и RBlock адресуются одним и тем же адресом. Таким образом, необходимо вычислить 4 адреса констант и 4 адреса «бабочек».

   Особенностью схемы вычисления инструкции «Butterfly» является наличие четырех дополнительных циклов для запуска процесса. Однако для данного

блока в этом нет необходимости. Он должен сразу рассчитывать корректные адреса. Также данный блок должен работать «на шаг вперед» относительно блока чтения-записи, чтобы второй был всегда обеспечен новыми адресами.

Еще одна особенность алгоритма: адреса и параметры самых первых «бабочек» будут всегда идентичными вне зависимости от размера БПФ. Таким образом, первую порцию адресов нужно вычислять в процессе инициализации аппаратуры.

Наконец, для данного блока требуется особое поведение в завершающей стадии алгоритма, когда адреса для всех «бабочек» уже рассчитаны, но из-за отставания блока чтения-записи на 1 «бабочку», а также конвейера, блок должен продолжать работу вплоть до записи последнего выходного данного на последней стадии алгоритма.

4. **Блок управления чтением-записью** обеспечивает считывание по присланным адресам констант и отсчетов, требуемых для выполнения каждого конкретного цикла инструкции «бабочка». Кроме того, данный блок осуществляет запись выходных операндов, полученных от импликатора. Однако в этом процессе есть одна особенность. Запись выходных данных происходит по адресам «предыдущей бабочки», а значит, необходимо хранить эти адреса.

Рис. 4 Архитектура средств поддержки БПФ

Ранее было отмечено, что выходные данные от импликатора приходят с задержкой, т. е. данный блок должен рассчитывать реальные адреса записи выходных отсчетов наперед и хранить их в блоке FIFO (first in, first out), чтобы сохранить корректность in-place-реализации. Данный блок представляет собой конвейер длины 4; значит, для начала и завершения вычислений требуется наличие четырех дополнительных шагов для заполнения и опустошения FIFO соответственно.

Схема из четырех циклов выполнения инструкции «Butterfly» показывает, что последовательность поступления выходных отсчетов для одной секции следующая: Re, Re, Im, Im. Так как секций 4, получаем, что каждые 8 записей необходимо переключать блоки RBlock и IBlock соответственно. Результат работы блока на каждом шаге — два упакованных Apdi_x4: операнда, содержащих либо константы для четырех секций, либо отсчеты для четырех секций, т. е. константы поступают в рекуррентное операционное устройство стандартным механизмом формирования горстей. В процессе ожидания последних выходных отсчетов на завершении алгоритма блок не должен выдавать ничего. На рис. 4 представлена обновленная архитектура БП.

4 Результаты испытаний программной и аппаратной моделей

В работе [7] дана оценка скорости вычисления 256-точечного БПФ на одной секции, которая составила $2N \log_2 N$. В процессе испытаний тестовые запуски капсулы осуществлялись на четырех секциях ГАРОС — как на предыдущей версии средств аппаратной поддержки, так и на модифицированной.

Сравнительные результаты испытаний

Параметр	Старая версия	Новая версия
Объем памяти для хранения констант (16-битных слов)	$256 \times 4 = 1024$	1024
Поддерживаемые размеры окон БПФ (точек)	256	от 8 до 1024
Размер капсулы (операндов)	132 для отсчетов + 23	23
Объем памяти для отсчетов (16-битных слов)	528	512
Benchmark (циклов)	$2N \log_2 N/4 = 1024$	$2N \log_2 N/4 = 1024$
Overhead (циклов)	44	16
Итого (циклов)	1068	1040

В таблице приводятся сравнительные результаты для двух версий. Данные приведены для $N = 256$ точек.

5 Заключение

Полученные результаты подтверждают успешность оптимизации и модификации средств аппаратной поддержки БПФ в ГАРОС. Новая версия средств

не только обладает большей гибкостью и скоростью, но и позволяет сократить размер капсулы. Полученная реализация средств поддержки позволила решить все перечисленные в разд. 2.3 проблемы.

Проблема 1 была решена путем размещения поворотных коэффициентов в отдельном блоке памяти, причем его чтение осуществляется по стандартному в ГАРОС механизму за счет формирования на выходе пары упакованных операндов. Более того, в том же самом объеме памяти было размещено гораздо больше различных значений поворотных коэффициентов, за счет чего повысилась гибкость и масштабируемость алгоритма.

Проблема 2 была решена также путем размещения отсчетов в отдельном блоке памяти и формирования пары упакованных операндов на выходе при его чтении. Более того, это автоматически решило проблему 5, так как теперь модификация капсулы — это всего лишь реконфигурация размера вычисляемого БПФ.

Проблемы 3 и 4 были решены за счет существенной переработки инструкции «Butterfly».

В различных областях ЦОС требуется вычислять не только БПФ с прореживанием по времени, но и по частоте, а также обратный БПФ. Кроме того, алгоритм по основанию 2 менее быстрый, чем по основанию 4, поэтому в будущем желательно обеспечить реализацию и других алгоритмов из семейства БПФ. Другим направлением работ в данном направлении должна стать апробация IP-блоков БПФ для эффективной аппаратной реализации «бабочек» и еще большего повышения производительности ГАРОС на этом классе задач.

Литература

1. *Lyons R. G.* Understanding digital signal processing. — 3rd ed. — Pearson Education, 2011. 966 p.
2. *Cooley J., Tukey J.* An algorithm for the machine calculation of complex Fourier series // Math. Comput., 1965. Vol. 19. No. 90. P. 297–301.
3. *Степченков Ю. А., Дьяченко Ю. Г., Хилько Д. В., Петрухин В. С.* Рекуррентная потоковая архитектура: особенности и проблемы реализации // Проблемы разработки перспективных микро-и наноэлектронных систем, 2016. № 2. С. 120–127.
4. *Хилько Д. В., Степченков Ю. А., Шикунов Д. И., Шикунов Ю. И.* Рекуррентная потоковая архитектура: технические аспекты реализации и результаты моделирования // Проблемы разработки перспективных микро-и наноэлектронных систем, 2016. № 2. Р. 128–135.
5. *Stepchenkov Yu., Morozov N., Khilko D., Shikunov Yu., Orlov G.* Hybrid multi-core recurrent architecture approbation on FPGA // IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering Proceedings. — Piscataway, NJ, USA: IEEE, 2019. P. 1705–1708.
6. Mixed-signal and DSP design techniques / Ed. W. Kester. — Analog Devices Inc., 2003, 410 p.
7. *Stepchenkov Yu. A., Khilko D. V., Shikunov Yu. I., Orlov G. A.* DSP filter kernels preliminary Benchmarking for recurrent data-flow architecture // IEEE Conference of

Russian Young Researchers in Electrical and Electronic Engineering Proceedings. — Piscataway, NJ, USA: IEEE, 2021. P. 2040–2044.

Поступила в редакцию 22.09.21

HARDWARE SUPPORT OF FAST FOURIER TRANSFORM OPTIMIZATION IN A RECURRENT SIGNAL PROCESSOR

D. V. Khilko, Yu. A. Stepchenkov, Yu. I. Shikunov, Yu. G. Diachenko, and G. A. Orlov

Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The paper covers the fast Fourier transform (FFT) support in the hybrid recurrent signal processor architecture. An analysis of the existing implementation is presented. Disadvantages and their ramifications are identified. An optimized solution is proposed to ease the scaling of both the architecture and the number of FFT samples.

Keywords: digital signal processing; fast Fourier transform; digital signal processor; Radix-2

DOI: 10.14357/08696527210407

Acknowledgments

The research was supported by the Russian Science Foundation (project No. 19-11-0034).

References

1. Lyons, R. G. 2011. *Understanding digital signal processing*. 3rd ed. Pearson Education. 966 p.
2. Cooley, J., and J. Tukey. 1965. An algorithm for the machine calculation of complex Fourier series. *Math. Comput.* 19(90):297–301.
3. Stepchenkov, Yu. A., Yu. G. Diachenko, D. V. Khilko, and V. S. Petrukhin. 2016. Rekurrentnaya potokovaya arkhitektura: osobennosti i problemy realizatsii [Recurrent data-flow architecture: Features and realization problems]. *Problemy razrabotki perspektivnykh mikro- i nanoelektronnykh system* [Problems of perspective micro- and nanoelectronic systems development] 2:120–127.
4. Khilko, D. V., Yu. A. Stepchenkov, D. I. Shikunov, and Y. I. Shikunov. 2016. Rekurrentnaya potokovaya arkhitektura: tekhnicheskiye aspekty realizatsii i rezul'taty modelirovaniya [Recurrent data-flow architecture: Technical aspects of implementation and modeling results]. *Problemy razrabotki perspektivnykh mikro- i nanoelektronnykh system* [Problems of perspective micro- and nanoelectronic systems development] 2:128–135.

5. Stepchenkov, Yu., N. Morozov, D. Khilko, Yu. Shikunov, and G. Orlov. 2019. Hybrid multi-core recurrent architecture approbation on FPGA. *IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering Proceedings*. Piscataway, NJ: IEEE. 1075–1078.
6. Kesten, W., ed. 2003. *Mixed-signal and DSP design techniques*. Analog Devices Inc. 410 p.
7. Stepchenkov, Yu. A., D. V. Khilko, Yu. I. Shikunov, and G. A. Orlov. 2021. DSP filter kernels preliminary benchmarking for recurrent data-flow architecture. *IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering Proceedings*. Piscataway, NJ: IEEE. 2040–2044.

Received September 22, 2021

Contributors

Khilko Dmitri V. (b. 1987) — senior scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; dhilko@yandex.ru

Stepchenkov Yuri A. (b. 1951) — Candidate of Science (PhD) in technology, leading scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; YStepchenkov@ipiran.ru

Shikunov Yury I. (b. 1995) — engineer-researcher, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; yishikunov@gmail.com

Diachenko Yuri G. (b. 1958) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; diaura@mail.ru

Orlov Georgy A. (b. 1994) — engineer-researcher, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; orlov.jaja@gmail.com

КОМПЬЮТЕРНАЯ И ЭКОНОМИЧЕСКАЯ МОДЕЛИ ГЕНЕРАЦИИ НОВОГО ЗНАНИЯ: СОПОСТАВИТЕЛЬНЫЙ АНАЛИЗ*

И. М. Зацман[1]

Аннотация: Сопоставляются две модели генерации нового знания. Первая, спиральная модель, используется при описании процессов генерации нового знания в экономической сфере. Вторая модель используется в компьютерной лингвистике и медицинской информатике. Эта модель ориентирована на проектирование информационных технологий и автоматизированных систем, обеспечивающих генерацию нового знания в процессе семантического анализа текстов (далее — информационно-технологически ориентированная модель, или модель ИТО). Сопоставление этих двух моделей представлено в статье как поэтапная трансформация спиральной модели в модель ИТО. Основная цель статьи состоит в описании четырех этапов трансформации. На первом этапе в спиральной модели выделяются ментальная и информационная среды. Затем, на втором этапе, к ним добавляется цифровая среда с потенциальными источниками нового знания в виде базы текстовых данных, соответствующих цели его извлечения из текстов. Задается эталон новизны извлеченного знания. На третьем этапе формируется перечень из восьми процессов генерации нового знания и добавляется база для хранения форм представления концептов нового извлеченного знания. Трансформация спиральной модели в модель ИТО завершается на четвертом этапе интеграцией восьми процессов генерации нового знания. Этапы трансформации иллюстрируются примером извлечения лингвистами нового знания о значениях немецких модальных глаголов (НМГ).

Ключевые слова: спиральная модель; генерация нового знания; модель ИТО; семантический анализ текстов; трансформация спиральной модели

DOI: 10.14357/08696527210408

1 Введение

Чтобы сопоставить модель ИТО со спиральной моделью генерации знания, сначала опишем последнюю и концептуально близкие ей варианты, а затем, в следующем разделе, рассмотрим основные этапы ее трансформации в модель ИТО. В настоящее время спиральная модель генерации знания [1–3] стала одной из самых популярных в сфере экономики. Она носит объяснительный характер, является апостериорной и описывает только качественно уже произошедшие процессы генерации знания. В ней определены две категории знания:

*Исследование выполнено при финансовой поддержке РФФИ (проект 20-012-00166) с использованием ЦКП «Информатика» ФИЦ ИУ РАН.

[1]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, izatsman@yandex.ru

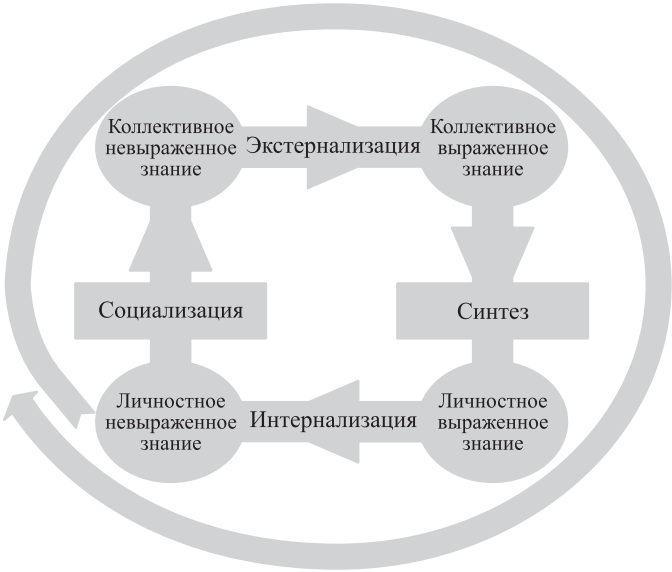


Рис. 1 Спиральная модель генерации знаний [4, с. 69]

индивидуальные (созданные одним экспертом) и коллективные (согласованные в коллективе экспертов). Каждая из них, в свою очередь, делится на две подкатегории: выраженное (явное) и невыраженное (неявное) знание. Таким образом, спиральная модель включает в себя следующие четыре вида знания:

(1) индивидуальное неявное знание,

(2) коллективное неявное знание,

(3) коллективное явное знание,

(4) индивидуальное явное знание.

В спиральной модели помимо этих четырех видов знания определены следующие четыре процесса перехода (преобразования) между ними: социализация, экстернализация, синтез и интернализация. По определению, каждый виток спирали как один этап генерации нового знания включает последовательность процессов: социализация → экстернализация → синтез → интернализация, после которого с процесса социализации начинается следующий этап генерации нового знания (рис. 1).

Согласно К. Братиану [5], «спиральная модель не содержит время в качестве явной переменной… Модель содержит время неявно, поскольку для любого преобразования оно требуется, но это неявное время невозможно измерить».

В работе [6] показано, что в этой модели явно или неявно присутствуют три измерения процесса генерации нового знания:

(1) измерение *социализации*, которое включает две номинативные категории (индивидуальное и коллективное знание);

(2) *временно́е* измерение, которое подразумевается в витках спирали генерации знания;

(3) измерение *эксплицитности*, которое включает две номинативные подкатегории (явное и неявное знание).

Есть варианты спиральной модели генерации знания, которые описывают его динамику с помощью большего числа номинативных категорий в измерении *социализации*, с одним или двумя явными временны́ми измерениями. Например, Ниссен [7] расширил спиральную модель, включив в нее два временны́х измерения: одно для этапов генерации/применения знания и второе для отражения процессов перехода между четырьмя видами знаний, перечисленными выше. В измерении *социализации* Ниссен определил три категории знания (индивидуальное, коллективное и организационное). Другая модель с тремя категориями знания в измерении *социализации* была предложена Вежбицким и Накамори [4, 8, 9]. Как и Ниссен, они также определяют три номинативные категории, но вместо организационного знания их модель включает конвенциональное (knowledge of humanity).

В работе [6] с использованием примеров на рис. 1 и 2 показано, что в концепции моделей, основанной на измерениях *социализации*, *эксплицитности* и *времени*, удается описать в лучшем случае только последовательность этапов генерации нового знания. Перечисленные модели [1–5, 7–9] по их концептуальному замыслу не предполагают возможность описания знания на входе и выходе каждого процесса перехода между его разными видами. Как следствие, эти модели не позволяют оценить рост знания на каждом этапе его генерации. Для описания знания на входе и выходе каждого процесса перехода между его разными видами, а также оценки степени роста нового знания в процессе семантического анализа текстов, а не апостериори была предложена другая модель [6, 10–14], которая получила англоязычное название «information-technology-oriented model — ITO model» [15, 16]. В русскоязычных работах используется краткое название «модель ИТО» [17].

Основная цель статьи состоит в сопоставлении спиральной модели и модели ИТО, которое осуществлено в статье в форме четырехэтапной трансформации первой модели во вторую. Следующий раздел содержит графическую иллюстрацию четырех этапов трансформации, подробно описанной в работах [15, 16]. Этапы трансформации иллюстрируются примером извлечения нового знания о значениях немецких модальных глаголов [18, 19].

2 Этапы трансформации

На первом этапе трансформации повернем рис. 1 на 90° по часовой стрелке и выделим в спиральной модели ментальную среду *невыраженного знания*,

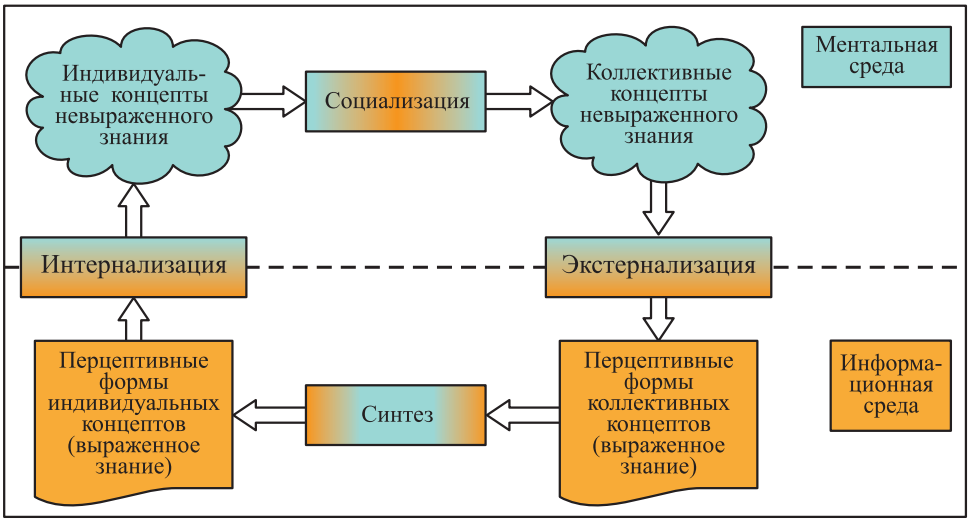


Рис. 2 Результат первого этапа трансформации спиральной модели [15, 16]

информационную среду *выраженного знания* и проведем границу между этими двумя подкатегориями каждой из двух категорий: как индивидуального, так и коллективного знания. Отметим, что процессы *социализации* и *синтеза* охватывают виды знания обеих сред, но, чтобы упростить рис. 2–5, разместим эти процессы в разных средах. Такое размещение этих процессов в средах соответствует начальной и конечной фазам каждого из них: начальная и конечная фазы процесса социализации относятся к ментальной среде, а процесса *синтеза* — к информационной. Процессы *интернализации* и *экстернализации* начинаются и завершаются в разных средах. Поэтому разместим их на границе между этими двумя средами (см. рис. 2).

Зададим цель генерации нового знания (например, извлечение нового знания в проекте по исследованию НМГ [18, 19]) и выберем эталон (критерий) его новизны (например, немецко-русский словарь [20]). Задание цели и выбор эталона новизны, как правило, предопределяют ту знаковую систему (в примере с НМГ — это знаковая система немецкого языка), которая будет использоваться, с одной стороны, для деления извлекаемого невыраженного знания на индивидуальные и коллективные концепты, а с другой стороны, для деления соответствующего ему выраженного знания[1] на перцептивные формы (см. рис. 2).

В примере с НМГ концепты — это значения модальных глаголов, определенные их дефинициями, а формы представления концептов — это сами модальные глаголы dürfen, können, mögen, müssen, sollen и wollen. Отметим, что од-

[1] Под выраженным знанием понимается знание, представленное в информационной среде [21].

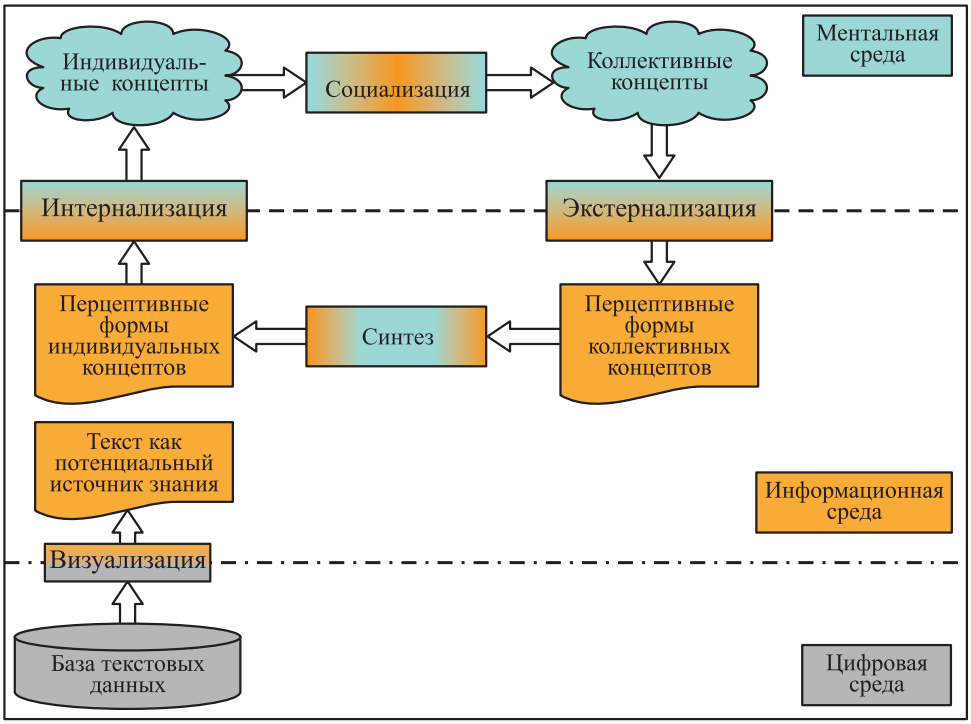


Рис. 3 Результат второго этапа трансформации [15, 16]

ной форме представления, например sollen, могут соответствовать более десяти концептов, так как модальные глаголы немецкого языка многозначны [22, 23].

На втором этапе трансформации к ментальной и информационной средам добавляется цифровая среда с базой текстовых данных как потенциальных источников нового знания, соответствующих цели его извлечения из текстов (в примере с НМГ — это база данных полных текстов нескольких сотен книг на немецком языке и их переводов на русский язык [15, 16, 18]), а также граница между информационной и цифровой средами. На рис. 3 показан новый процесс перехода, обозначенный словом «визуализация» на этой границе. Этот процесс необходим для преобразования двоичного кода текста предложения, найденного в базе текстовых данных по заданному критерию поиска (в примере поиск предложений ведется по НМГ), в перцептивный текст на экране, т. е. сенсорно воспринимаемый экспертом (в примере — лингвистом). На рис. 3 четыре процесса спиральной модели не связаны с процессом *визуализации*.

На третьем этапе добавляются еще три новых процесса перехода: *концептуализация*, *аннотирование* и *оцифровка* (см. рис. 4). Отметим, что аннотируются

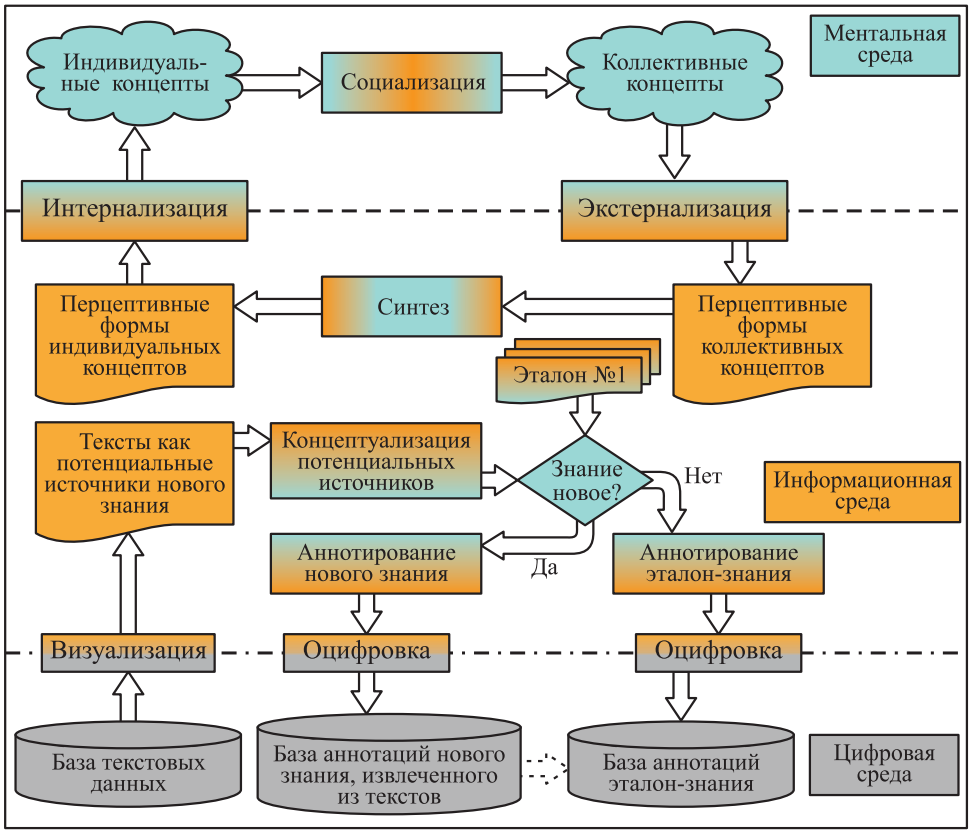


Рис. 4 Результат третьего этапа трансформации [15, 16]

только те предложения и их контекстные окружения, которые удовлетворяют заданному эталону новизны. По решению экспертов эталон может пополняться новым извлеченным знанием. В случае его пополнения все аннотации нового извлеченного знания переносятся в базу эталон-знания, что показано на рис. 4 пунктирной стрелкой. Отметим, что эталон новизны используется только после завершения процесса *концептуализации* предложений и их контекстных окружений как потенциальных источников нового знания.

После добавления этих трех процессов перехода (*концептуализации*, *аннотирования* и *оцифровки*) образуются два независимых контура генерации нового знания. Верхний контур соответствует спиральной модели с ее четырьмя процессами перехода и с делением знания на четыре вида, но без описания потенциальных источников нового знания. Нижний контур включает базу текстовых данных, четыре новых процесса перехода, которые связывают ее с базой для

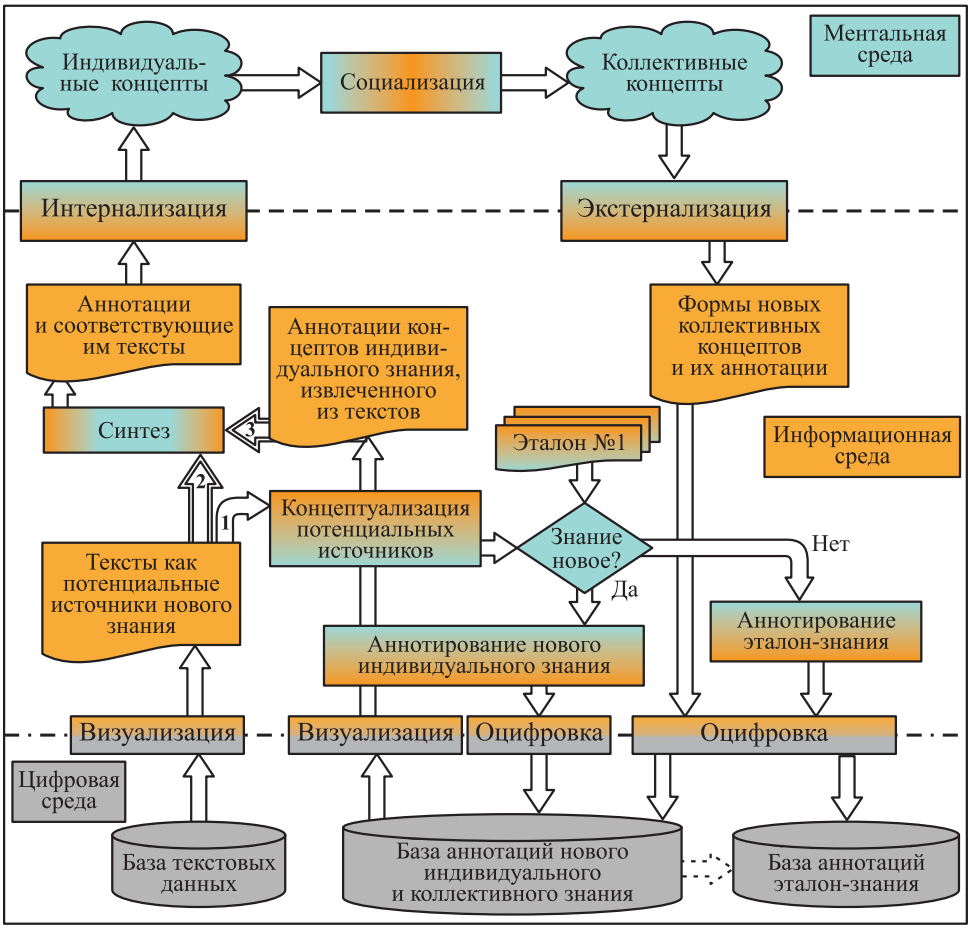


Рис. 5 Модель ИТО [15, 16]

хранения цифровых форм представления концептов нового извлеченного знания[1] (далее — база нового знания), но без явного деления извлеченного знания на индивидуальное и коллективное.

Трансформация спиральной модели в модель ИТО, объединяющую эти два контура, завершается на четвертом этапе. Их объединение выполнено с помощью разрыва связи между процессами *экстернализации* и *синтеза* с последующим изменением последовательности выполнения процессов перехода, их входа и выхода (см. рис. 5).

[1] Описание четырех новых процессов перехода, обеспечивающих целенаправленное извлечение экспертами нового знания из текстов, дано в работах [15–17].

В спиральной модели коллективное выраженное знание служит непосредственным входом для процесса *синтеза*, выходом которого является индивидуальное выраженное знание. В модели ИТО коллективное выраженное знание (знаковые формы перцептивных концептов) и их коллективные аннотации (этот процесс аннотирования на рис. 5 не показан) оцифровываются и загружаются в базу нового знания. Отметим, что процесс *концептуализации* выполняется для всех источников, что показано стрелкой с цифрой **1**. Однако на вход процесса *синтеза* подаются только те тексты, которые действительно оказались источниками нового знания, что показано стрелкой с двойным контуром и цифрой **2**. Их новизна выясняется только после завершения *концептуализации* предложений текста, и только они аннотируются с сохранением аннотаций в базе нового знания. После извлечения из нее эти аннотации становятся входом операции *синтеза*. Это показано стрелкой также с двойным контуром и цифрой **3**. Двойной контур стрелок на рис. 5 подчеркивает одновременность использования в операции *синтеза* текстов и аннотаций концептов нового индивидуального знания.

В модели ИТО предполагается, что в процессе *концептуализации* одного и того же текста разные эксперты могут сформировать разные аннотации. В этих случаях выполняется процесс *социализации*, во время которого согласовывается понимание предложений этого текста между экспертами, формируется коллективный концепт и соответствующая ему коллективная аннотация.

В процессе трансформации спиральной модели в модель ИТО были использованы следующие позиции, обеспечивающие целенаправленность генерации нового знания:

- цель извлечения нового знания из текстов и эталон его новизны, который может пополняться экспертами;
- знаковая система, которая обеспечивает деление извлекаемого невыраженного знания на индивидуальные и коллективные концепты, а их представления в информационной среде, т. е. выраженного знания, — на перцептивные формы представления концептов в информационной среде;
- база текстовых данных как потенциальных источников нового знания, соответствующих цели его извлечения из текстов.

3 Заключение

Кардинальное отличие модели ИТО от спиральной модели и ее вариантов состоит в том, что она является априорной и может служить основой создания информационных систем и технологий, обеспечивающих целенаправленную генерацию нового знания. На основе модели ИТО уже разработана информационная технология [10–17, 22, 23], которая на практике обеспечивает извлечение лингвистами нового знания о значениях НМГ [18, 19], что и ставилось целью проекта по гранту РФФИ № 20-012-00166. Сначала, в качестве эталона новизны извлеченного знания, в этом проекте использовалась версия словаря [20] до

его обновления с применением разработанной информационной технологии, что соответствует эталону № 1 на рис. 4 и 5.

База текстовых данных этого проекта содержит тексты нескольких сотен книг на немецком языке и их переводы на русский язык. В этих текстах есть 109 130 предложений, содержащих перечисленные выше шесть НМГ с их разными значениями. Эти предложения с их контекстным окружением и служат потенциальными источниками нового знания о значениях НМГ. По состоянию на 5 мая 2021 г. 5 366 из 109 130 предложений с их контекстным окружением были концептуализированы и из них были извлечены восемь новых значений НМГ, которые не были описаны ранее, но в 2021 г. они уже включены в словарь [20] по результатам применения разработанной информационной технологии.

В заключение отметим, что модель ИТО в настоящее время используется также при проектировании информационной технологии извлечения нового терминологического знания из текстов медицинских документов, относящихся к исследуемой болезни [15, 16, 24].

Литература

1. *Nonaka I.* The knowledge-creating company // Harvard Bus. Rev., 1991. Vol. 69. No. 6. P. 96–104.
2. *Nonaka I.* A dynamic theory of organizational knowledge creation // Organ. Sci., 1994. Vol. 5. No. 1. P. 14–37.
3. *Нонака И., Такеучи Х.* Компания — создатель знания / Пер. с англ. — М.: Олимп-бизнес, 2003. 384 с. (*Nonaka I., Takeuchi H.* The knowledge-creating company. — Oxford, NY, USA: Oxford University Press, 1995. 284 p.)
4. *Wierzbicki A. P., Nakamori Y.* Basic dimensions of creative space // Creative space: Models of creative processes for knowledge civilization age / Eds. A. P. Wierzbicki, Y. Nakamori. — Berlin: Springer Verlag, 2006. P. 59–90.
5. *Bratianu C.* A strategic view on the knowledge dynamics models used in knowledge management // 20th European Conference on Knowledge Management Proceedings. — Reading, U.K.: Academic Publishing International Ltd., 2019. Vol. 1. P. 185–192.
6. *Зацман И. М.* Проблемно-ориентированная верификация полноты темпоральных онтологий и заполнение понятийных лакун // Информатика и её применения, 2020. Т. 14. Вып. 3. С. 119–128.
7. *Nissen M. E.* Harnessing knowledge dynamics: Principled organizational knowing & learning. — London, U.K.: IRM Press, 2006. 278 p.
8. *Wierzbicki A. P., Nakamori Y.* Knowledge sciences: Some new developments // Z. Betriebswirt., 2007. Vol. 77. No. 3. P. 271–295.
9. *Nakamori Y.* Knowledge and systems science — enabling systemic knowledge synthesis. — London / New York: CRC Press, 2013. 234 p.
10. *Зацман И. М.* Стадии целенаправленного извлечения знаний, имплицированных в параллельных текстах // Системы и средства информатики, 2018. Т. 28. № 3. С. 175–188.

11. *Zatsman I.* Finding and filling lacunas in knowledge systems // 20th European Conference on Knowledge Management Proceedings. — Reading, U.K.: Academic Publishing International Ltd., 2019. Vol. 2. P. 1143–1151.
12. *Зацман И. М.* Целенаправленное развитие систем лингвистических знаний: выявление и заполнение лакун // Информатика и её применения, 2019. Т. 13. Вып. 1. С. 91–98.
13. *Zatsman I.* Finding and filling lacunas in linguistic typologies // 15th Forum (International) on Knowledge Asset Dynamics Proceedings. — Matera, Italy: Institute of Knowledge Asset Management, 2020. P. 780–793.
14. *Zatsman I.* Three-dimensional encoding of emerging meanings in AI-systems // 21st European Conference on Knowledge Management Proceedings. — Reading, U.K.: Academic Publishing International Ltd., 2020. P. 878–887.
15. *Zatsman I., Khakimova A.* New knowledge discovery for creating terminological profiles of diseases // 22nd European Conference on Knowledge Management Proceedings. — Reading, U.K.: Academic Publishing International Ltd., 2021. P. 837–846.
16. *Zatsman I.* A model of goal-oriented knowledge discovery based on human–computer symbiosis // 16th Forum (International) on Knowledge Asset Dynamics Proceedings. — Rome, Italy: Arts for Business Institute, 2021. P. 297–312.
17. *Зацман И. М.* Формы представления нового знания, извлеченного из текстов // Информатика и её применения, 2021. Т. 15. Вып. 3. С. 83–90.
18. *Добровольский Д. О., Зализняк Анна А.* Немецкие конструкции с модальными глаголами и их русские соответствия: проект надкорпусной базы данных // Компьютерная лингвистика и интеллектуальные технологии: По мат-лам Междунар. конф. «Диалог». — М.: РГГУ, 2018. Т. 17. С. 172–184.
19. *Добровольский Д. О., Зализняк Анна А.* О семантике немецкого глагола sollen // ВАПросы языкознания. — М.: Буки-Веди, 2020. С. 459–464.
20. Немецко-русский словарь: актуальная лексика / Под ред. Д. О. Добровольского. — М.: Лексрус, 2021 (в печати).
21. *Зацман И. М.* Кодирование концептов в цифровой среде // Информатика и её применения, 2019. Т. 13. Вып. 4. С. 97–106.
22. *Гончаров А. А., Зацман И. М., Кружков М. Г.* Эволюция классификаций в надкорпусных базах данных // Информатика и её применения, 2020. Т. 14. Вып. 4. С. 108–116.
23. *Гончаров А. А., Зацман И. М., Кружков М. Г.* Представление новых лексикографических знаний в динамических классификационных системах // Информатика и её применения, 2021. Т. 15. Вып. 1. С. 86–93.
24. *Зацман И. М.* Проблемно-ориентированная актуализация словарных статей двуязычных словарей и медицинской терминологии: сопоставительный анализ // Информатика и её применения, 2021. Т. 15. Вып. 1. С. 94–101.

Поступила в редакцию 11.10.21

COMPUTER AND ECONOMIC MODELS OF NEW KNOWLEDGE GENERATION: A COMPARATIVE ANALYSIS

I. M. Zatsman

Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: Two models of generating new knowledge are compared. The first, the spiral model, describes the processes of generating new knowledge in the economic sphere. The second model is used in computational linguistics and medical informatics. This model focuses on the design of information technologies and automated systems that ensure the generation of new knowledge during the semantic analysis of texts (referred to as the information technology-oriented model, or ITO model). The comparison of the two models is presented as a step-by-step transformation of the spiral model into the ITO model. The primary purpose of the paper is to describe four stages of transformation. At the first stage, mental and informational media are separated in the spiral model. Then, at the second stage, a digital medium is added to them with potential sources of new knowledge in the form of a database of text data corresponding to the goal of knowledge discovery from texts. The reference sample of the novelty of the discovered knowledge is being set. At the third stage, a list of eight processes of generating new knowledge is formed. A knowledge base is being added for storing forms of representation of concepts of newly discovered knowledge. The transformation of the spiral model into the ITO model is completed at the fourth stage by integrating eight processes of generating new knowledge. The stages of transformation are illustrated by an example of discovering new knowledge about the meanings of German modal verbs.

Keywords: spiral model; generation of new knowledge; ITO model; semantic analysis of texts; transformation of the spiral model

DOI: 10.14357/08696527210408

Acknowledgments

The reported study was funded by RFBR, project number 20-012-00166. The research was carried out using the infrastructure of the Shared Research Facilities "High Performance Computing and Big Data" (CKP "Informatics") of FRC CSC RAS (Moscow).

References

1. Nonaka, I. 1991. The knowledge-creating company. *Harvard Bus. Rev.* 69(6):96–104.
2. Nonaka, I. 1994. A dynamic theory of organizational knowledge creation. *Organ. Sci.* 5(1):14–37.

3. Nonaka, I., and H. Takeuchi. 1995. *The knowledge-creating company*. Oxford, NY: Oxford University Press. 284 p.
4. Wierzbicki, A. P., and Y. Nakamori. 2006. Basic dimensions of creative space. *Creative space: Models of creative processes for knowledge civilization age*. Eds. A. P. Wierzbicki and Y. Nakamori. Berlin: Springer Verlag. 59–90.
5. Bratianu, C. 2019. A strategic view on the knowledge dynamics models used in knowledge management. *20th European Conference on Knowledge Management Proceedings*. Reading, U.K.: Academic Publishing International Ltd. 1:185–192.
6. Zatsman, I. 2020. Problemno-orientirovannaya verifikatsiya polnoty temporal'nykh ontologiy i zapolnenie ponyatiynykh lakun [Problem-oriented verifying the completeness of temporal ontologies and filling conceptual lacunas]. *Informatika i ee Primeneniya — Inform. Appl.* 14(3):119–128.
7. Nissen, M. E. 2006. *Harnessing knowledge dynamics: Principled organizational knowing & learning*. London: IRM Press. 278 p.
8. Wierzbicki, A. P., and Y. Nakamori. 2007. Knowledge sciences: Some new developments. *Z. Betriebswirt.* 77(3):271–295.
9. Nakamori, Y. 2013. *Knowledge and systems science — enabling systemic knowledge synthesis*. London/New York: CRC Press. 234 p.
10. Zatsman, I. 2018. Stadii tselenapravlennogo izvlecheniya znaniy, implitsirovannykh v parallel'nykh tekstakh [Stages of goal-oriented discovery of knowledge implied in parallel texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(3):175–188.
11. Zatsman, I. 2019. Finding and filling lacunas in knowledge systems. *20th European Conference on Knowledge Management Proceedings*. Reading, U.K.: Academic Publishing International Ltd. 2:1143–1151.
12. Zatsman, I. 2019. Tselenapravlennoe razvitie sistem lingvisticheskikh znaniy: vyyavlenie i zapolnenie lakun [Goal-oriented development of linguistic knowledge systems: Identifying and filling of lacunae]. *Informatika i ee Primeneniya — Inform. Appl.* 13(1):91–98.
13. Zatsman, I. 2020. Finding and filling lacunas in linguistic typologies. *15th Forum (International) on Knowledge Asset Dynamics Proceedings.* Matera, Italy: Institute of Knowledge Asset Management. 780–793.
14. Zatsman, I. 2020. Three-dimensional encoding of emerging meanings in AI-systems. *21st European Conference on Knowledge Management Proceedings.* Reading, U.K.: Academic Publishing International Ltd. 878–887.
15. Zatsman, I., and A. Khakimova. 2021. New knowledge discovery for creating terminological profiles of diseases. *22nd European Conference on Knowledge Management Proceedings*. Reading, U.K.: Academic Publishing International Ltd. 837–846.
16. Zatsman, I. 2021. A model of goal-oriented knowledge discovery based on human–computer symbiosis. *16th Forum (International) on Knowledge Asset Dynamics Proceedings*. Rome, Italy: Arts for Business Institute. 297–312.
17. Zatsman, I. 2021. Formy predstavleniya novogo znaniya, izvlechennogo iz tekstov [Forms representing new knowledge discovered in texts]. *Informatika i ee Primeneniya — Inform. Appl.* 15(3):83–90.
18. Dobrovol'skij, D. O., and Anna A. Zalizniak. 2018. Nemetskie konstruktsii s modal'nymi glagolami i ikh russkie sootvetstviya: proekt nadkorpusnoy bazy dannykh

[German constructions with modal verbs and their Russian correlates: A supracorpora database project]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: po mat-lam Mezhdunar. konf. "Dialog"* [Computational Linguistics and Intellectual Technologies: Papers from the Annual Conference (International) "Dialogue"]. Moscow: RSHI. 17:172–184.

19. Dobrovol'skij, D. O., and Anna A. Zalizniak. 2020. O semantike nemetskogo glagola sollen [About the semantics of the German verb sollen]. *VAProsy yazykoznaniya* [Topics in the study of language]. Moscow: Buki-Vedi. 459–464.
20. Dobrovol'skiy, D. O., ed. 2021 (in press). *Nemetsko-russkiy slovar': aktual'naya leksika* [German–Russian dictionary: Actual vocabulary]. Moscow: Leksrus.
21. Zatsman, I. M. 2019. Kodirovanie kontseptov v tsifrovoy srede [Digital encoding of concepts]. *Informatika i ee Primeneniya — Inform. Appl.* 13(4):97–106.
22. Goncharov, A. A., I. M. Zatsman, and M. G. Kruzhkov. 2020. Evolyutsiya klassifikatsiy v nadkorpusnykh bazakh dannykh [Evolution of classifications in supracorpora databases]. *Informatika i ee Primeneniya — Inform. Appl.* 14(4):108–116.
23. Goncharov, A. A., I. M. Zatsman, and M. G. Kruzhkov. 2021. Predstavlenie novykh leksikograficheskikh znaniy v dinamicheskikh klassifikatsionnykh sistemakh [Representation of new lexicographical knowledge in dynamic classification systems]. *Informatika i ee Primeneniya — Inform. Appl.* 15(1):86–93.
24. Zatsman, I. 2021. Problemno-orientirovannaya aktualizatsiya slovarnykh statey dvuyazychnykh slovarey i meditsinskoy terminologii: sopostavitel'nyy analiz [Problem-oriented updating of dictionary entries of bilingual dictionaries and medical terminology: Comparative analysis]. *Informatika i ee Primeneniya — Inform. Appl.* 15(1):94–101.

Received October 11, 2021

Contributor

Zatsman Igor M. (b. 1952) — Doctor of Science in technology, Head of Department, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; izatsman@yandex.ru

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ, 2021. Т. 31. № 4. С. 97–113

ИНФОРМАЦИОННЫЕ АСПЕКТЫ ОБЕСПЕЧЕНИЯ БЕЗОПАСНОСТИ НА ТРАНСПОРТЕ: АНАЛИТИЧЕСКИЕ РАСЧЕТЫ

А. В. Борисов[1], *А. В. Босов*[2], *Д. В. Жуков*[3], *А. В. Иванов*[4]

Аннотация: В третьей статье цикла, посвященного описанию системы поддержки принятия решений (СППР) в области обеспечения безопасности пассажирских перевозок и противодействия противоправной деятельности на транспорте, концептуальная модель, предложенная в первой статье цикла и дополненная базовым функциональным содержанием во второй статье цикла, расширяется специализированными аналитическими постановками. Вначале формируется набор формальных понятий, обеспечивающих возможность трансформации оперативных задач пользователей системы в расчетные задачи над отобранными транспортными данными. Затем представлена серия содержательных расчетных задач, обеспечивающих информационную поддержку аналитика предметной области при решении оперативных задач, анализе инцидентов и ситуативном анализе. В заключении статьи анонсируется запланированное представление аналитических функций, связанных со сложным процессингом данных.

Ключевые слова: транспортная безопасность; система поддержки принятия решений; расчетная задача; транспортная информация; маршрут следования; транспортная зона

DOI: 10.14357/08696527210409

1 Введение

В работе [1], открывавшей цикл, продолженный затем в работе [2], представлены первые шаги исследования предметной области пассажирских перевозок с точки зрения организации информационного обеспечения в СППР, создаваемых в интересах обеспечения безопасности транспортных перевозок. Концептуальная модель [1], упорядочивающая такие базовые понятия, как пассажир,

[1]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, ABorisov@ipiran.ru

[2]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, AVBosov@ipiran.ru

[3]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, DZhukov@ipiran.ru

[4]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, AIvanov@ipiran.ru

перевозчик, поездка, рейс, транспортный пункт, определила ключевой информационный источник обсуждаемой СППР, содержащий факты перевозки конкретного пассажира из одного транспортного пункта в другой. Эти факты названы транспортной информацией, а объединяющий их ресурс — базой транспортной информации. В [2] реализован первый этап функционального наполнения СППР, обеспечивающий выполнение ее предназначения — информационную и аналитическую поддержку деятельности экспертов в сфере обеспечения безопасности на транспорте. Это этап поиска и отбора информации. В результате сформировано множество специфических запросов, характерных именно для пассажирских перевозок в связи с мероприятиями по обеспечению безопасности на транспорте или противодействию противоправной деятельности.

В данной статье круг постановок функциональных задач, поддерживаемых обсуждаемой СППР, расширяется существенно более специализированными задачами, требующими, в частности, выполнения некоторых расчетов с отобранными данными. Постановки потенциально содержательных задач анализа транспортной информации формировались в значительной степени из имеющегося опыта реализации прикладных проектов в исследуемой области. Однако исследовательский интерес здесь достаточно широк. Так, в работе [3] обезличенная транспортная информация используется для анализа характеристик поездок пассажиров городского транспорта, в [4] эти же данные используются для реконструкции поездок с учетом пересадок и последующего исследования пространственной и временно́й регулярности поездок, а в [5] — времени путешествий.

Некоторые транспортные задачи продвигаются дальше исследования транспортной информации. Так, в [6] для этого используется вполне естественный инструмент — данные GPS с телефона пассажира.

В данной работе акцент делается на проблематике обеспечения безопасности на транспорте: предложенная в [1] для этой области концептуальная модель во втором разделе статьи дополнена набором формальных понятий, обеспечивающих возможности трансформации аналитических задач в задачи поиска данных и выполнения расчетов. Ключевой здесь является модель, предлагаемая для формального описания понятия маршрута следования — последовательности перемещений пассажира или групп пассажиров между транспортными пунктами с темпоральной привязкой. Эта модель дает возможность сформулировать и предложить решения целой серии расчетных задач, способных оказывать значительную информационную поддержку пользователям СППР при решении реальных оперативных задач, анализе инцидентов и ситуативном анализе. Такие задачи представлены в третьем, ключевом разделе статьи. Это задачи построения маршрутов на основе имеющейся транспортной информации, их оценка и анализ, сравнение маршрутов отдельных пассажиров и их групп, ситуативный анализ маршрутов и событий.

В заключении статьи анонсируется содержание финального этапа аналитики обсуждаемой СППР.

2 Основные понятия методов аналитических расчетов

Описанные в [2] возможности первичного анализа транспортной информации обеспечивают отбор сведений о поездках — ключевой источник для дальнейшей аналитической работы. Транспортные данные, т. е. данные о поездках индивидуальных пассажиров или их групп, могут быть найдены и отобраны по довольно гибким критериям. Так формируется исходный материал для аналитической работы [7]. Для продолжения работы с этими данными нужны специализированные информационные модели, отражающие более глубокие свойства и процессы предметной области [8]. Набор формальных понятий, образующий такую модель, представлен ниже.

Базовое понятие используемой модели — поездка — для любого пассажира характеризуется с географической точки зрения *пунктами отправления и прибытия*. Такими пунктами могут быть железнодорожные вокзалы и прочие зарегистрированные места остановки пассажирских поездов, учитываемые в железнодорожном расписании; аэропорты и прочие места взлета/посадки пассажирских самолетов, учитываемые в авиарасписании; порты и прочие места остановки пассажирских судов; автовокзалы и прочие учитываемые места остановки пассажирских автобусов. Все такие пункты отправления/прибытия, сведения о которых автоматически включаются в состав транспортных данных, обобщены термином *транспортные пункты*.

Поскольку транспортных пунктов, в том числе географически близких, очень много, то требуется возможность уменьшения детализации. Для этого вводится *система транспортных зон* — разбиение всего множества транспортных пунктов (РФ и остальных государств) на непустые непересекающиеся множества. Для простоты будем считать, что любой транспортный пункт принадлежит в точности одной транспортной зоне.

Возможность формирования состава транспортных зон предполагается доступной эксперту — пользователю СППР. В соответствующем справочнике для каждой зоны экспертом должны быть определены несколько балльных характеристик, «*весов зоны*», характеризующих значение транспортных пунктов зоны при анализе перемещений пассажиров. Так, каждой транспортной зоне должны быть сопоставлены веса: W_{dep} — вес отправления и W_{arr} — вес прибытия. Семантика этих весов обсуждается ниже.

Определение 1. *Информацией* $\mathrm{TI}(x)$ *о перемещениях пассажира* x называется все множество перемещений пассажира, зарегистрированных в базе транспортной информации. Информация $\mathrm{TI}(x)$ представляет собой множество шестерок вида $(t^i_{\mathrm{dep}}, p^i_{\mathrm{dep}}, t^i_{\mathrm{arr}}, p^i_{\mathrm{arr}}, \mathrm{tt}^i, v^i)$, где t^i_{dep} — время отправления согласно i-му проездному документу; p^i_{dep} — транспортный пункт отправления; t^i_{arr} — время прибытия согласно i-му проездному документу; p^i_{arr} — транспортный пункт прибытия; tt^i — вид транспорта i-й поездки; v^i — идентификатор i-го рейса согласно проездному документу.

Информацией $\mathrm{TI}(x, T_1, T_2)$ *о перемещениях пассажира* x *на интервале времени* $[T_1, T_2]$ называется подмножество множества $\mathrm{TI}(x)$, для элементов которого либо $t^i_{\mathrm{dep}} \in [T_1, T_2]$, либо $t^i_{\mathrm{arr}} \in [T_1, T_2]$.

Заметим, что множества транспортной информации $\mathrm{TI}(x)$ и $\mathrm{TI}(x, T_1, T_2)$ не совпадают с маршрутом, т. е. упорядоченной последовательностью смен местоположений пассажира x, но могут использоваться для его восстановления или анализа. Для начальной характеризации этих множеств предлагается использовать следующие функции:

- $\#\mathrm{TI}(x)$ — общее число перемещений (купленных билетов) пассажира x;
- $\#\mathrm{TI}(x, T_1, T_2)$ — число перемещений пассажира x на интервале времени $[T_1, T_2]$;
- $\mathrm{TdepFirst}(x)$ — время первого отбытия пассажира x;
- $\mathrm{TarrFirst}(x)$ — время первого прибытия пассажира x;
- $\mathrm{TdepLast}(x)$ — время последнего отбытия пассажира x;
- $\mathrm{TarrLast}(x)$ — время последнего прибытия пассажира x.

Введенные формальные показатели позволяют перейти к расчетам качественных характеристик процессов перемещения пассажиров, зарегистрированных в базе транспортной информации.

2.1 Зонная активность

Вычисляемый показатель «*зонная активность*» AL — это интегральная числовая характеристика, описывающая все зарегистрированные перемещения по транспортным зонам с учетом их оперативного значения (количественно характеризуется весами W_{dep} и W_{arr}).

Далее описан рекуррентный алгоритм вычисления данной характеристики, достаточно простой для применения и последующего анализа выполненных вычислений. Результат его применения инвариантен относительно последовательности формирования транспортной информации.

Будем считать, что $\mathrm{AL} = \mathrm{AL}(k)$, где k — число зарегистрированных для данного пассажира в базе транспортной информации поездок, и

$$\mathrm{AL}(0) = 0\,, \quad \mathrm{AL}(k) = \mathrm{AL}(k-1) + \mathrm{WTr}(p^k_{\mathrm{dep}}, p^k_{\mathrm{arr}})\,,$$

где $\mathrm{WTr}(p_{\mathrm{dep}}, p_{\mathrm{arr}})$ — функция веса поездки из пункта p_{dep} в пункт p_{arr}.

Функция транспортной зоны $\mathrm{TZ}(p)$ ставит в соответствие транспортному пункту p транспортную зону, в которой находится данный пункт. В качестве функции WTr предлагается

$$\mathrm{WTr}\,(p_{\mathrm{dep}}, p_{\mathrm{arr}}) = \mathrm{Wdep}\,(\mathrm{TZ}(p_{\mathrm{dep}})) + \mathrm{Warr}\,(\mathrm{TZ}(p_{\mathrm{arr}}))\ .$$

Кроме того, можно вычислять изменения «зонной активности» $\mathrm{dAL}(x, T_1, T_2)$ пассажира x на интервале времени $[T_1, T_2]$, используя соотношения для AL, рассчитываемое по информации $\mathrm{TI}(x, T_1, T_2)$.

2.2 Мобильность

Вычисляемый показатель «*мобильность пассажира*» представляет собой группу числовых величин, характеризующих количественный и весовой «вклад» поездок за интересующий отрезок времени в соответствующие суммарные величины и их интенсивность. Под мобильностью пассажира x на отрезке времени $[T_1, T_2]$ будут подразумеваться следующие четыре характеристики.

1. *Мобильность по числу поездок* $\mathrm{MobNTr}(x, T_1, T_2)$ пассажира x на отрезке времени $[T_1, T_2]$ — функция

$$\mathrm{MobNTr}\,(x, T_1, T_2) = \frac{\#\mathrm{TI}\,(x, T_1, T_2)}{\#\mathrm{TI}(x)}\,.$$

   Величина $\mathrm{MobNTr}(x, T_1, T_2)$ всегда лежит в интервале $[0, 1]$ и определяет, какую долю в общем числе поездок занимают поездки на отрезке времени $[T_1, T_2]$.

2. *Мобильность по интенсивности поездок* $\mathrm{MobINTr}(x, T_1, T_2)$ пассажира x на отрезке времени $[T_1, T_2]$ — функция

$$\mathrm{MobINTr}\,(x, T_1, T_2) = \frac{(T_{\mathrm{Curr}} - \mathrm{TdepFirst}(x))\#\mathrm{TI}(x, T_1, T_2)}{(T_2 - T_1)\#\mathrm{TI}(x)}\,.$$

   Величина $\mathrm{MobINTr}(x, T_1, T_2)$ определяет отношение интенсивности поездок (числа поездок в единицу времени) на отрезке времени $[T_1, T_2]$ к средней интенсивности всех зарегистрированных поездок пассажира. Выполнение условия $\mathrm{MobINTr}(x, T_1, T_2) > 1$ означает, что на отрезке времени $[T_1, T_2]$ данный пассажир перемещался интенсивнее (чаще), чем в среднем.

3. *Мобильность по весу зон* $\mathrm{MobWTr}(x, T_1, T_2)$ пассажира x на отрезке времени $[T_1, T_2]$ — функция

$$\mathrm{MobWTr}\,(x, T_1, T_2) = \frac{\mathrm{dAL}\,(x, T_1, T_2)}{\mathrm{AL}(x)}\,.$$

   Величина $\mathrm{MobWTr}(x, T_1, T_2)$ всегда лежит в интервале $[0, 1]$ и определяет, какую долю в зонную активность пассажира внесли поездки на отрезке времени $[T_1, T_2]$.

4. *Мобильность по интенсивности весов зон* $\mathrm{MobIWTr}(x, T_1, T_2)$ пассажира x на отрезке времени $[T_1, T_2]$, вычисленная в текущий момент времени T_{Curr}, — функция

$$\mathrm{MobIWTr}\,(x, T_1, T_2) = \frac{(T_{\mathrm{Curr}} - \mathrm{TdepFirst}(x))\mathrm{dAL}(x, T_1, T_2)}{(T_2 - T_1)\mathrm{AL}(x)}\,.$$

Величина $\mathrm{MobIWTr}(x, T_1, T_2)$ показывает отношение роста зонной активности (баллов в единицу времени) при поездках на отрезке времени $[T_1, T_2]$ к среднему росту зонной активности всех зарегистрированных поездок пассажира. Выполнение условия $\mathrm{MobIWTr}(x, T_1, T_2) > 1$ означает, что на отрезке времени $[T_1, T_2]$ рост зонной активности у данного пассажира был выше, чем в среднем. С течением времени данный показатель мобильности меняется. Это придает ему динамический характер, отсутствующий у других показателей, что может оказаться более полезным.

2.3 Выявление пассажиров с прерывающимся маршрутом следования

Прерывающийся маршрут пассажира — это состояние транспортной информации о пассажире, которое не позволяет восстановить непрерывную последовательность его перемещений. Такое состояние может быть связано с наличием незарегистрированных перемещений пассажира (например, частным транспортом) или зарегистрированных перемещений, противоречащих друг другу.

Определение 2. *Перемещения* пассажира на интервале времени $[T_1, T_2]$, заданные транспортной информацией $\mathrm{TI}(x, T_1, T_2)$, называются *непрерывными*, если для всех N зарегистрированных перемещений одновременно выполняются следующие условия:

(1) перемещения удается упорядочить во времени, т. е. $t_{\mathrm{dep}_1} < t_{\mathrm{arr}_1} < t_{\mathrm{dep}_2} < \cdots$ $\cdots < t_{\mathrm{arr}_N}$;

(2) перемещения согласованы «по пространству», т. е. $p_{\mathrm{dep}_1} = p_{\mathrm{arr}_1}, \ldots$ $\ldots, p_{\mathrm{dep}_N} = p_{\mathrm{arr}_N}$.

Как вариант условие согласованности «по пространству» может иметь вид:

$$\mathrm{TZ}\left(p_{\mathrm{dep}_1}\right) = \mathrm{TZ}\left(p_{\mathrm{arr}_1}\right), \ldots, \mathrm{TZ}\left(p_{\mathrm{dep}_N}\right) = \mathrm{TZ}\left(p_{\mathrm{arr}_N}\right).$$

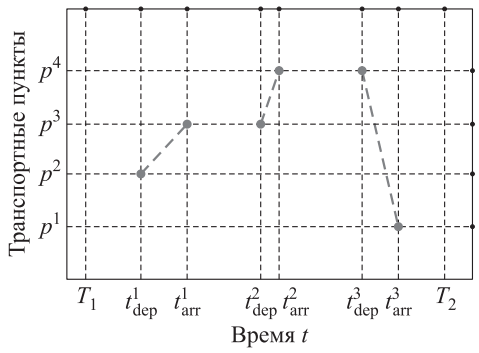


Рис. 1 Пример непрерывного перемещения

Такая интерпретация имеет смысл в том случае, когда транспортные зоны построены с учетом географической близости.

Перемещения, для которых не выполнено хотя бы одно из перечисленных условий, называются *разрывными*.

Графический пример непрерывного перемещения приведен на рис. 1.

Характеризуя далее разрывные перемещения на интервале времени $[T_1, T_2]$, заданные транспортной информацией $\mathrm{TI}(x, T_1, T_2)$, предлага-

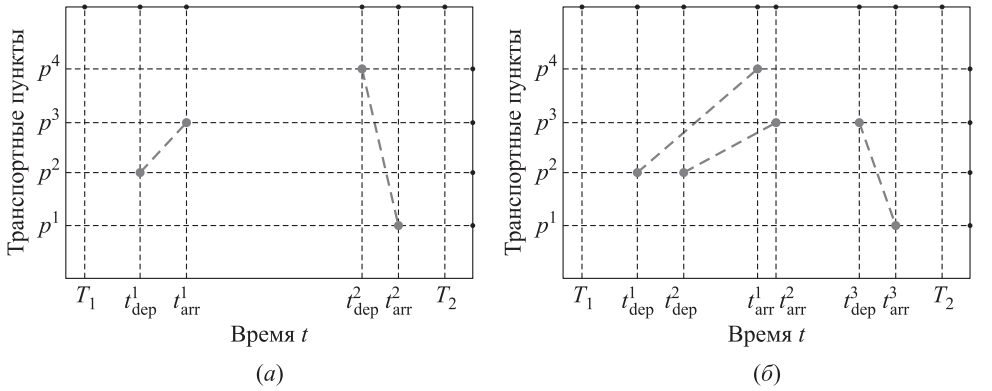


Рис. 2 Примеры простого (*а*) и ветвящегося (*б*) разрывных перемещений

ется выделять *простые разрывные перемещения* (или «*перемещения с выпавшим плечом*»), если условие 1 выполнено, а условие 2 нарушено, и *ветвящиеся разрывные перемещения*, если условие 1 нарушено.

Графические примеры простого и ветвящегося разрывного перемещений пассажира приведены на рис. 2. Ясно, что простое разрывное перемещение свидетельствует о возможной неполноте транспортной информации о пассажире x, в то время как ветвящееся разрывное перемещение означает, что пассажир x на интервале времени $[T_1, T_2]$ совершил несколько покупок билетов для совершения взаимоисключающих перемещений.

3 Задачи и методы аналитических расчетов

Напомним, что информация $\mathrm{TI}(x, T_1, T_2)$ содержит в себе все зарегистрированные перемещения на интервале времени $[T_1, T_2]$, а также, возможно, сведения о перемещении, начавшемся ранее T_1 или закончившемся после T_2. Для решения ряда сложных расчетных задач потребуется по информации $\mathrm{TI}(x, T_1, T_2)$ определять положение пассажира для любых моментов времени t: $T_1 \le t \le T_2$.

3.1 Детальное определение положения

Упорядочим по возрастанию все времена t^i_{dep} и t^i_{arr}, содержащиеся в информации $\mathrm{TI}(x, T_1, T_2)$. Подынтервалы $[t^i_{\text{arr}}, t^j_{\text{dep}}] \subset [T_1, T_2]$, а также $[T_1, t^j_{\text{dep}}]$ и $[t^i_{\text{arr}}, T_2]$, не содержащие в себе другие моменты t^i_{arr} и t^i_{dep}, характеризуются как *интервалы неподвижности*. Остальные интервалы, дополняющие множество интервалов неподвижности до всего интервала $[T_1, T_2]$, называются интервалами движения. Возможны два варианта определения положения пассажира на интервале неподвижности.

1. *Пространственно-детерминированный*. Пусть $[t^i_{\text{arr}}, t^j_{\text{dep}}] \subset [T_1, T_2]$ — некоторый интервал неподвижности пассажира x, а p^i_{arr} и p^i_{dep} — соответствующие пункты прибытия и отбытия. Будем считать, что на интервале $[t^i_{\text{arr}}, (t^i_{\text{arr}} + t^j_{\text{dep}})/2]$ пассажир x находится в пункте p^i_{arr}, а на интервале $[(t^i_{\text{arr}} + t^j_{\text{dep}})/2, t^j_{\text{dep}}]$ — в пункте p^i_{dep} (т. е. пассажир в каждый момент времени находится только в одном пункте, см. рис. 3, *а*, жирные линии).

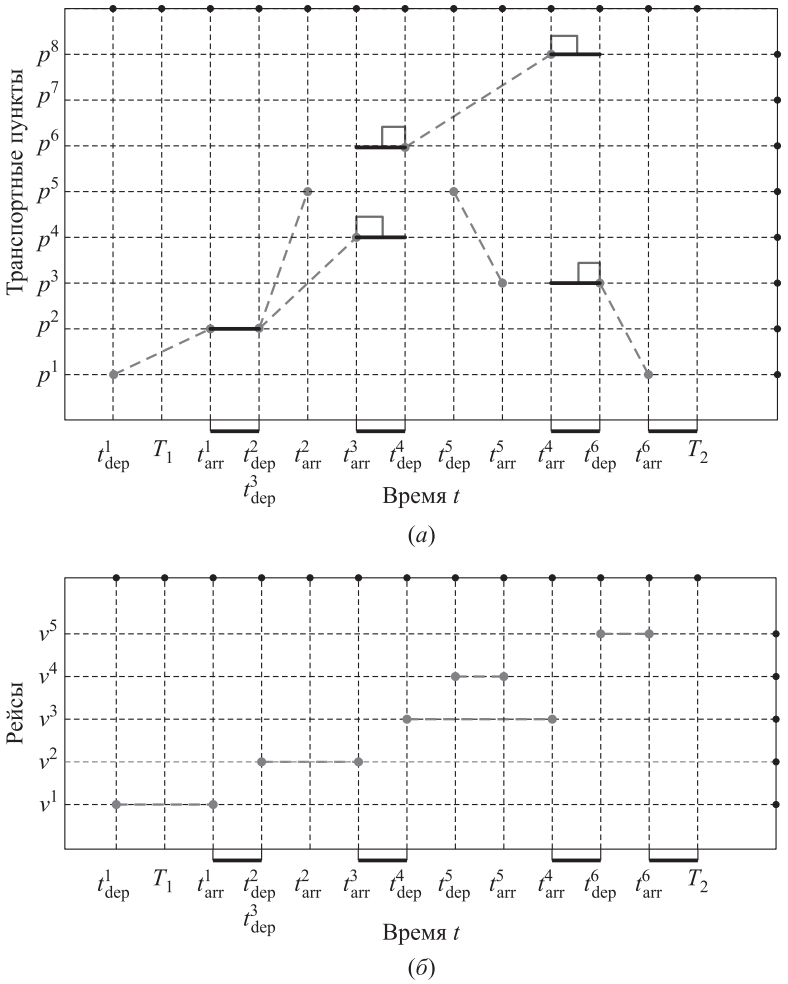


Рис. 3 Определение положения пассажира на интервале неподвижности (*а*) и пример задания положения пассажира во время перемещений (*б*)

2. *Пространственно-вероятностный*. Пусть $[t^i_{\text{arr}}, t^j_{\text{dep}}] \subset [T_1, T_2]$ — некоторый интервал неподвижности пассажира x, а p^i_{arr} и p^j_{dep} — соответствующие пункты прибытия и отправления. На интервале $[t^i_{\text{arr}}, t^j_{\text{dep}}]$ пассажир x находится во множестве пунктов $\{p^i_{\text{arr}}, p^j_{\text{dep}}\}$ (т. е. допускается нахождение пассажира в различных пунктах с определенной вероятностью, см. рис. 3, *а*, тонкие линии).

Определение положения пассажира во время перемещений осуществляется следующим образом. Пусть $[t^i_{\text{arr}}, t^j_{\text{dep}}] \subset [T_1, T_2]$ — некоторый интервал движения пассажира x. В любой фиксированный момент времени $t \in [t^i_{\text{arr}}, t^j_{\text{dep}}]$ пассажир x находится в поездке на всех рейсах $\{v^k\}$, для которых $t \in [t^k_{\text{arr}}, t^k_{\text{dep}}]$ (см. рис. 3, *б*). Множество $\{v^k\}$ определяется совокупностью приобретенных данным пассажиром билетов.

Таким образом, для детального определения положения пассажира необходимо последовательно решить следующие задачи:

– разбить интересующий интервал времени $[T_1, T_2]$ на временны́е интервалы неподвижности и интервалы движения;

– для каждого интервала неподвижности определить положение пассажира согласно пространственно-вероятностному варианту;

– для каждого интервала движения определить множества $\{v^k\}$ рейсов, на которых этот пассажир мог находиться в соответствии с приобретенными билетами.

3.2 Пересечение двух пассажиров по детальной информации

Под совместным местоположением двух пассажиров понимается совместное пребывание в транспортных пунктах/зонах и совместное перемещение в транспортных средствах. Трудоемкость решения задачи определения совместного местоположения относительно невысокая, так как эта задача по затратам вычислительных ресурсов не относится к разряду «комбинаторных». В связи с этим данная задача может быть решена явным образом в автоматизированном режиме.

Пересечением двух пассажиров x и y во времени и пространстве за время $[T_1, T_2]$ является пара (возможно, пустых) множеств $\boldsymbol{L}$ и $\boldsymbol{M}$. Множество $\boldsymbol{L}$ состоит из троек $(t^k_s, t^k_f, \text{TZ}^k)$, где $[t^k_s, t^k_f]$ — временной интервал k-й встречи x и y; TZ^k — множество транспортных зон k-й встречи x и y; множество $\boldsymbol{M}$ состоит из троек (t^i_s, t^i_f, v^i), где $[t^i_s, t^i_f]$ — интервал i-го совместного перемещения x и y; v^i — i-й совместный рейс x и y.

Множество $\boldsymbol{M}$ находится достаточно просто: необходимо выполнить запрос на определение совпадений рейсов для x и y за время $[T_1, T_2]$, т. е.

определение пятерок $(t^i_{\text{dep}}(x), p^i_{\text{dep}}(x), t^i_{\text{arr}}(x), p^i_{\text{arr}}(x), v^i(x)) \in \text{TI}(x, T_1, T_2)$ и $(t^j_{\text{dep}}(y), p^j_{\text{dep}}(y), t^j_{\text{arr}}(y), p^j_{\text{arr}}(y), v^j(y)) \in \text{TI}(y, T_1, T_2)$, для которых $v^i(x) = v^j(y)$ и $[t^i_{\text{dep}}(x), t^i_{\text{arr}}(x)] \cap [t^j_{\text{dep}}(y), t^j_{\text{arr}}(y)] \neq \emptyset$.

Множество $\boldsymbol{L}$ определяется следующим образом:

- для пассажиров x и y находится множество $\{([t^i_{\text{arr}}(x), t^i_{\text{dep}}(x)], \{p^i_{\text{dur}}(x)\})\}$, где $[t^i_{\text{arr}}(x), t^i_{\text{dep}}(x)]$ — все интервалы неподвижности пассажира x; $\{p^i_{\text{dur}}(x)\}$ — множества транспортных пунктов, в которых x может находиться на интервале $[t^i_{\text{arr}}(x), t^i_{\text{dep}}(x)]$ (полученные согласно пространственно-детерминированному или пространственно-вероятностному алгоритму); множество $\{([t^j_{\text{arr}}(y), t^j_{\text{dep}}(y)], \{p^j_{\text{dur}}(y)\})\}$ для пассажира y определяется аналогично;
- для множеств $\{p^i_{\text{dur}}(x)\}$ и $\{p^j_{\text{dur}}(y)\}$ вычисляются $\{\text{TZ}(p^i_{\text{dur}}(x))\}$ и $\{\text{TZ}(p^j_{\text{dur}}(y))\}$ — множества соответствующих транспортных зон;
- для всевозможных сочетаний пар $([t^i_{\text{arr}}(x), t^i_{\text{dep}}(x)], \{p^i_{\text{dur}}(x)\})$ пассажира x и $([t^j_{\text{arr}}(y), t^j_{\text{dep}}(y)], \{p^j_{\text{dur}}(y)\})$ пассажира y проверяется одновременное выполнение условий:

$$\left[t^i_{\text{arr}}(x), t^i_{\text{dep}}(x)\right] \cap \left[t^j_{\text{arr}}(y), t^j_{\text{dep}}(y)\right] = \left[t^k_s, t^k_f\right] \neq \emptyset;$$

$$\left\{\text{TZ}\left(p^i_{\text{dur}}(x)\right)\right\} \cap \left\{\text{TZ}\left(p^j_{\text{dur}}(y)\right)\right\} = \text{TZ}^k \neq \emptyset.$$

3.3 Совпадение перемещений и событий

Выявление по транспортной информации пассажиров, пребывавших в определенные моменты времени в определенных транспортных пунктах/зонах (т. е. совпадение перемещений пассажиров и фиксированных событий), представляет несомненный оперативный интерес. Приведем соответствующую постановку задачи и алгоритм ее решения.

Определение 3. *Событием* называется пара $(q^i, \{p^{ij}\}_j)$, где q^i — интервал времени i-го события (означенная дата), заданный дискретно с шагом в сутки/месяц/год; $\{p^{ij}\}_j$ — непустое множество транспортных пунктов (означенное место), в которых произошло i-е событие.

Причина того, что для события задается не один транспортный пункт, а некоторое множество, заключается в том, что истинное интересующее событие обычно связано с населенным пунктом, к которому привязано несколько транспортных пунктов. Например, если истинное событие произошло в Москве и известно, что интересующий пассажир прибыл в Москву на самолете, то вполне естественно, что множество транспортных пунктов будет содержать: {Шереметьево, Домодедово, Внуково, Быково}.

Исходное множество событий удобно представить с помощью понятия цилиндрического множества $C = \{q^i, \{p^{ij}\}_j\}_i$, которое будем предполагать уникально именованным и непустым.

Перемещения пассажира x полностью совпадают с событиями цилиндрического множества $C = \{q^i, \{p^{ij}\}_j\}_i$ если $\mathsf{P}(x, q^i) \cap \{p^{ij}\}_j \neq \emptyset \; \forall\, i$, т. е. согласно приобретенным проездным документам во все означенные даты q^i пассажир x мог находиться в означенных местах $\{p^{ij}\}_j$.

Задача определения совпадений перемещений и событий, связанная с цилиндрическим множеством C, решаемая для группы пассажиров G (или всех пассажиров, зарегистрированных в базе транспортной информации), заключается в нахождении всех пассажиров из G (или из всей базы), перемещения которых полностью совпадают с цилиндрическим множеством C.

Программная реализация решения данной задачи должна предусматривать следующие варианты заполнения транспортных пунктов в цилиндрических множествах:

- индивидуальный выбор транспортных пунктов;
- выбор всех пунктов, входящих в транспортную зону;
- выбор всех пунктов, входящих в транспортную зону с возможностью фильтрации по виду транспорта.

3.4 Выявление связей и маршрутов групп пассажиров

Решения группы задач по выявлению связей и маршрутов групп пассажиров представляются в следующей интуитивно понятной терминологии. Пассажир осуществляет последовательное по времени перемещение в системе транспортных пунктов/зон с учетом либо без учета видов транспорта и рейсов, которые образуют *маршрут*. Маршруты двух и более пассажиров могут пересекаться в том случае, если имеется полное или частичное совпадение последовательности нахождения пассажиров в транспортных пунктах/зонах (совпадение пребывания в пункте/зоне по времени не требуется). Такое *пересечение маршрутов* свидетельствует о наличии *связи пассажиров*.

Формальные понятия, привлекаемые к решению задачи, таковы.

Определение 4. *Звеном маршрута* называется упорядоченная тройка $(\mathrm{tp}^i, \mathrm{tt}^i, v^i)$, где tp^i — транспортный пункт (обязательное поле); tt^i — вид транспорта, с помощью которого покинули пункт (поле может быть пустым); v^i — рейс транспорта, с помощью которого покинули пункт (поле может быть пустым).

Маршрутом называется упорядоченная по времени совокупность звеньев $\{(\mathrm{tp}^i, \mathrm{tt}^i, v^i)\}_i$. В последнем звене маршрута поля (tt^i, v^i) обязательно отсутствуют. Порядок в совокупности соответствует хронологии пребывания в транспортных пунктах.

Зональным звеном маршрута называется упорядоченная совокупность $(\mathrm{tz}^i, \mathrm{tt}^i)$, где tz^i — транспортная зона (обязательное поле); tt^i — вид транспорта, с помощью которого покинули пункт (поле может быть пустым).

Зональным маршрутом называется упорядоченная по времени совокупность зональных звеньев $\{(\mathrm{tz}^i, \mathrm{tt}^i)\}_i$. В последнем звене маршрута поле tt^i обязательно отсутствует.

«Ручное» построение маршрута предполагает пошаговое задание полей в соответствии с используемыми определениями. Никаких ограничений на географическое расположение пунктов/зон (типа замкнутости) при «ручном» построении маршрута не накладывается. Выполняется пользователем СППР. Сформированные маршруты могут быть сохранены и использованы в дальнейшей работе. Для маршрута должна предусматриваться возможность графического представления (аналогичного приведенному на рис. 3).

Автоматическое построение маршрута пассажира по историческим данным, т. е. по имеющейся транспортной информации (по заданному промежутку времени) может осуществляться в автоматическом режиме. При этом используемая транспортная информация по конкретному пассажиру должна быть очищена от ветвящихся разрывных перемещений (см. п. 2.3). Операция «очистки» (т. е. выбора той или иной ветви-перемещения) должна выполняться экспертом вручную на основе автоматического решения задачи обнаружения ветвящегося разрывного перемещения.

Пусть $\mathrm{TI}(x, T_1, T_2)$ не содержит ветвящихся разрывных перемещений и транспортная информация упорядочена в хронологическом порядке. Будем использовать *склейку маршрутов* $\mathrm{MC}(x, T_1, T_2)$, состоящую из упорядоченного набора звеньев $\{(\mathrm{tp}^j, \mathrm{tt}^j, v^j)\}_j$. Склейка строится следующим образом по полям:

– tp^j в склейке соответствуют хронологическому появлению транспортных пунктов отбытия и прибытия p^i_{dep} и p^i_{arr}; при этом если $p^i_{\mathrm{dep}} = p^{i+1}_{\mathrm{arr}}$ (т. е. нет «выпавшего плеча»), то совпадающие пункты (прибытия и отбытия в следующую поездку) трактуются как один;

– (tt^j, v^j) при отсутствии «выпавшего плеча» формируются из транспортной информации; при наличии «выпавшего плеча» ($p^i_{\mathrm{dep}} \neq p^{i+1}_{\mathrm{arr}}$) поля (tt^j, v^j) остаются пустыми.

Пусть транспортный пункт ktp чаще всего встречается в множестве $\{p^i_{\mathrm{dep}}, p^i_{\mathrm{arr}}\}_i$. Такой пункт будем называть *узловым*. При наличии нескольких пунктов с максимальным числом вхождений в множество $\{p^i_{\mathrm{dep}}, p^i_{\mathrm{arr}}\}_i$ узловой пункт выбирается аналитиком из предоставленных вариантов вручную.

Пусть $\mathrm{MC}(x, T_1, T_2)$ — склейка маршрутов, полученная по транспортной информации $\mathrm{TI}(x, T_1, T_2)$, а ktp — узловой ТП. Множество отрезков склейки,

разбитой транспортным пунктом ktp, образует *автоматический набор маршрутов*, полученных из $\mathrm{MC}(x, T_1, T_2)$.

Автоматическое построение маршрута выполняется по запросам пользователя СППР. Предусматривается возможность сохранять сформированные маршруты. При построении маршрута пользователю доступны средства для удаления ветвящихся разрывных перемещений, выбора узлового пункта, доопределения «выпавшего плеча». Исходные данные для процесса при этом формируются из транспортной информации о перемещениях заданного пассажира x на заданном отрезке времени $[T_1, T_2]$. Для маршрута может быть сформировано графическое представление (см. рис. 3).

3.5 Сравнение маршрутов пассажиров

В рамках *качественного сравнения маршрутов* $M_1 = \{(\mathrm{tp}^j, \mathrm{tt}^j, v^j)\}_j$ *и* $M_2 = \{(\mathrm{TP}^i, \mathrm{TT}^i, V^i)\}_i$ определим следующие критерии.

Маршруты *совпадают*, если хронологически совпадают последовательности соответствующих транспортных пунктов в маршрутах, т. е. если $(\mathrm{tp}^1, \ldots, \mathrm{tp}^N) = (\mathrm{TP}^1, \ldots, \mathrm{TP}^N)$.

Маршруты *совпадают по способу передвижения*, если они совпадают географически, а также $(\mathrm{tt}^1, \ldots, \mathrm{tt}^N) = (\mathrm{TT}^1, \ldots, \mathrm{TT}^N)$.

Если $M_1 = M_2$, то маршруты будем считать *идентичными*.

Аналогично определяются критерии качественного сравнения зональных маршрутов $\mathrm{ZM}_1 = \{(\mathrm{tz}^j, \mathrm{tt}^j)\}_j$ и $\mathrm{ZM}_2 = \{(\mathrm{TZ}^i, \mathrm{TT}^i)\}_i$.

Зональные маршруты *географически совпадают*, если совпадают хронологические последовательности соответствующих зон в маршрутах, т. е. когда $(\mathrm{tz}^1, \ldots, \mathrm{tz}^n) = (\mathrm{TZ}^1, \ldots, \mathrm{TZ}^N)$.

Зональные маршруты *совпадают по способу передвижения*, если они совпадают географически, а также $(\mathrm{tt}^1, \ldots, \mathrm{tt}^N) = (\mathrm{TT}^1, \ldots, \mathrm{TT}^N)$.

Качественное сравнение маршрутов должно выполняться по запросу пользователя СППР для произвольных сохраненных наборов маршрутов. Результат сравнения маршрутов может быть представлен в графическом виде.

При сравнении маршрутов с географической точки зрения возможно выполнить *количественное сравнение маршрутов* и получить числовую характеристику их сходства. Так как в данной группе задач маршруты рассматриваются с географической точки зрения, то будем рассматривать географические маршруты в виде вектора $(\mathrm{tp}^1, \ldots, \mathrm{tp}^n)$ (т. е. упорядоченной последовательности транспортных пунктов), а зональные географические маршруты — в виде вектора $(\mathrm{tz}^1, \ldots, \mathrm{tz}^n)$ (упорядоченной последовательности транспортных зон). Определим функции сходства маршрутов $\mathrm{GM}_1 = (\mathrm{tp}^1, \ldots, \mathrm{tp}^n)$ и $\mathrm{GM}_2 = (\mathrm{TP}^1, \ldots, \mathrm{TP}^N)$.

Функция сходства маршрутов $\mathrm{GMS}(\mathrm{GM}_1, \mathrm{GM}_2)$:

$$\mathrm{GMS}(\mathrm{GM}_1, \mathrm{GM}_2) = \begin{cases} 0, & \text{если } \mathrm{GM}_1 \text{ не является подвектором } \mathrm{GM}_2 \\ & \text{и } \mathrm{GM}_2 \text{ не является подвектором } \mathrm{GM}_1\,; \\ \dfrac{\min(n, N)}{\max(n, N)} & \text{в противном случае}\,. \end{cases}$$

Легко проверить, что если маршруты отвечают критерию географического совпадения, то функция их сходства равна 1.

Функция сходства зональных маршрутов $\mathrm{GZMS}(\mathrm{GZM}_1, \mathrm{GZM}_2)$:

$$\mathrm{GZMS}(\mathrm{GZM}_1, \mathrm{GZM}_2) = \begin{cases} 0, & \text{если } \mathrm{GZM}_1 \text{ не является подвектором } \mathrm{GZM}_2 \\ & \text{и } \mathrm{GZM}_2 \text{ не является подвектором } \mathrm{GZM}_1\,; \\ \dfrac{\min(n, N)}{\max(n, N)} & \text{в противном случае}\,. \end{cases}$$

Легко проверить, что если зональные маршруты отвечают критерию географического совпадения, то функция их сходства равна 1.

Предполагается, что количественное сравнение маршрутов может выполняться для произвольных наборов сохраненных маршрутов.

Следующий вариант *сравнения двух множеств маршрутов* PS_1 *и* PS_2. Источник и способ их создания не важен — это могут быть два множества, созданных «вручную», множества маршрутов двух пассажиров, построенных автоматически по историческим данным и др. Результатом сравнения PS_1 и PS_2 являются результаты попарного сравнения «каждый с каждым» маршрутов из PS_1 и PS_2 согласно критериям сравнения маршрутов и функции сходства. При решении этой задачи эксперту предоставляются только положительные ответы (отрицательные и нулевые ответы опускаются).

Результатом *сравнения двух множеств зональных маршрутов* PZS_1 *и* PZS_2 являются результаты попарного сравнения «каждый с каждым» маршрутов из PZS_1 и PZS_2 согласно критериям сравнения зональных маршрутов и функции сходства. При решении этой задачи эксперту также предоставляются только положительные ответы.

4 Заключение

Исследование в области информационной поддержки обеспечения безопасности пассажирских перевозок и противодействия противоправной деятельности на транспорте, инициированное концептуальной моделью в [1], дополненное простыми поисковыми процедурами (аналитическими задачами первого типа) в [2],

в данной статье получило расширенную специальную функциональность, обеспеченную расчетными аналитическими задачами (задачами второго типа). Эти задачи уже представляют реальные инструменты, обеспечивающие информационную поддержку аналитика предметной области при решении оперативных задач, анализе инцидентов и ситуативном анализе. Теперь имеются все основания для обсуждения третьего типа задач, к которым относятся сложные вычислительные процедуры, требующие значительных вычислительных ресурсов, в том числе на предварительную подготовку данных. Характерным средством решения такого рода задач служат инструменты интерактивной аналитической обработки данных (OLAP, online analytical processing). Соответствующие постановки задач и методы их решения планируются к представлению в следующей, завершающей статье цикла.

Литература

1. *Борисов А. В., Босов А. В., Жуков Д. В., Иванов А. В., Сушко Д. В.* Информационные аспекты обеспечения безопасности на транспорте: онтология предметной области, модели и варианты использования // Системы и средства информатики, 2020. Т. 30. № 1. С. 126–134.
2. *Борисов А. В., Босов А. В., Жуков Д. В., Иванов А. В.* Информационные аспекты обеспечения безопасности на транспорте: поиск и отбор информации // Системы и средства информатики, 2021. Т. 31. № 2. С. 80–88.
3. *Hu Z., Jingen L., Bing H.* Analysis of passenger travel behavior based on public transportation OD data // 6th Conference (International) on Information and Education Technology Proceedings. — New York, NY, USA: Association for Computing Machinery, 2018. P. 275–280.
4. *Ouyang Q., Lv Y., Ren Y., Ma J., Li J.* Passenger travel regularity analysis based on a large scale smart card data // J. Adv. Transport., 2018. Vol. 2018. Art. 9457486. 11 p.
5. *Cristóbal T., Padrón G., Quesada-Arencibia A., Alayón F., García C.* Systematic approach to analyze travel time in road-based mass transit systems based on data mining // IEEE Access, 2018. Vol. 6. P. 32861–32873.
6. *Bellver P., Shuma J., Kirmse A., Udeshi T.* Extracting patterns from location history // 19th ACM SIGSPATIAL Conference (International) on Advances in Geographic Information Systems Proceedings. — New York, NY, USA: Association for Computing Machinery, 2011. P. 397–400.
7. *Попов И. А., Персианинов Ю. Н., Миронов Я. А., Дегтярев А. Ю.* Использование возможностей ЕИТКС ОВД в деятельности органов предварительного следствия в системе МВД России. — М.: Проспект, 2014. 112 с.
8. *van Slooten K.* Optimal information modeling techniques. — Idea Group Inc, 2001. 306 p.

Поступила в редакцию 20.02.21

INFORMATION ASPECTS OF SECURITY IN TRANSPORT: ANALYTICAL CALCULATIONS

A. V. Borisov, A. V. Bosov, D. V. Zhukov, and A. V. Ivanov

Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: In the third article of a series, devoted to the description of the decision support system in the field of ensuring the safety of passenger traffic and countering illegal activities in transport, the conceptual model proposed in the first article of a series and supplemented by the basic functional content in the second article of a series is expanded with specialized analytical statements. Firstly, a set of formal concepts is formed that provide the possibility of transforming the operational tasks of the system users into computational tasks for the selected transport data. Then, a series of computational problems is presented that provide information support for the analyst of the subject area in solving operational problems, analyzing incidents, and conducting situational analysis. Finally, a planned presentation of analytical functions associated with complex data processing is announced.

Keywords: transport safety; decision support system; design problem; transport information; route of travel; transport area

DOI: 10.14357/08696527210409

References

1. Borisov, A. V., A. V. Bosov, D. V. Zhukov, A. V. Ivanov, and D. V. Sushko. 2020. Informatsionnye aspekty obespecheniya bezopasnosti na transporte: ontologiya predmetnoy oblasti, modeli i varianty ispol'zovaniya [Information aspects of security in transport: Ontology of the subject area, models, and cases]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 30(1):126–134.
2. *Borisov, A. V., A. V. Bosov, D. V. Zhukov, and A. V. Ivanov*. 2021. Informatsionnye aspekty obespecheniya bezopasnosti na transporte: poisk i otbor informatsii [Information aspects of security in transport: Search and selection of information]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(2):80–88.
3. Hu, Z., L. Jingen, and H. Bing. 2018. Analysis of passenger travel behavior based on public transportation OD data. *6th Conference (International) on Information and Education Technology Proceedings*. New York, NY: Association for Computing Machinery. 275–280.
4. Ouyang, Q., Y. Lv, Y. Ren, J. Ma, and J. Li. 2018. Passenger travel regularity analysis based on a large scale smart card data. *J. Adv. Transport.* 2018:9457486. 11 p.
5. Cristóbal, T., G. Padrón, A. Quesada-Arencibia, F. Alayón, and C. García. 2018. Systematic approach to analyze travel time in road-based mass transit systems based on data mining. *IEEE Access* 6:32861–32873.

6. Bellver, P., J. Shuma, A. Kirmse, and T. Udeshi. 2011. Extracting patterns from location history. *19th Conference (International) on Advances in Geographic Information Systems Proceedings*. New York, NY: Association for Computing Machinery. 397–400.
7. Popov, I. A., Yu. N. Persianinov, Ya. A. Mironov, and A. Yu. Degtyarev. 2014. *Ispol'zovanie vozmozhnostey EITKS OVD v deyatel'nosti organov predvaritel'nogo sledstviya v sisteme MVD Rossii* [Using the capabilities of EITKS IAB in the activities of the preliminary investigation bodies in the system of the Ministry of Internal Affairs of Russia]. Moscow: Prospekt. 112 p.
8. van Slooten, K. 2001. *Optimal information modeling techniques*. Idea Group Inc. 306 p.

Received February 20, 2021

Contributors

Borisov Andrey V. (b. 1965) — Doctor of Science in physics and mathematics, principal scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; ABorosov@ipiran.ru

Bosov Alexey V. (b. 1969) — Doctor of Science in technology, principal scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; AVBosov@ipiran.ru

Zhukov Denis V. (b. 1979) — principal specialist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; DZhukov@ipiran.ru

Ivanov Alexey V. (b. 1976) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; AIvanov@ipiran.ru

СТРАТЕГИЯ ИССЛЕДОВАНИЙ И РАЗРАБОТОК В ОБЛАСТИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА III: ДОКТРИНА ГОСУДАРСТВЕННОЙ ПОДДЕРЖКИ США

А. В. Борисов[1]***, А. В. Босов***[2]***, Д. В. Жуков***[3]

Аннотация: Статья продолжает цикл работ, посвященных анализу влияния государственного управления на эффективность проведения исследований и разработок в области искусственного интеллекта (AI R&D, Artificial Intelligence Research and Development). В третьей части цикла рассматривается влияние государства на AI R&D на примере США. Дано краткое описание стратегического документа США в области AI R&D, включая компактное изложение его целей и задач, а также принципов реализации. Помимо этого проанализирован смежный документ Министерства обороны США. Представлен классификационный анализ направлений AI R&D в области обороны и безопасности, проводимых основными специализированными научно-исследовательскими организациями США.

Ключевые слова: искусственный интеллект; Министерство обороны США (DoD); Управление перспективных исследовательских проектов Министерства обороны США (DARPA); Агентство передовых исследований в сфере разведки (IARPA)

DOI: 10.14357/08696527210410

1 Введение

Статья представляет третью часть цикла работ, посвященного анализу государственного влияния на сферу исследований и разработок в области искусственного интеллекта. Предыдущие части [1, 2] были посвящены хронологии развития данной области в СССР и Российской Федерации и сравнительному анализу наукометрических показателей публикационной активности по AI R&D за последние 20 лет в России и мире. Эти обзоры сформировали представление о «начальных условиях» реализации «Национальной стратегии развития искусственного интеллекта на период до 2030 года» [3].

Наше государство было не первым, кто разработал и приступил к реализации стратегии подобного рода. Аналогичные документы были приняты к выполнению

[1]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, ABorisov@ipiran.ru

[2]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, AVBosov@ipiran.ru

[3]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, DZhukov@ipiran.ru

в 2016 г. в США и в 2017 г. в КНР. Цель данной части цикла — анализ государственного влияния в области AI R&D в США. Соединенные Штаты Америки являются не только сверхдержавой, родоначальницей средств вычислительной техники, но и инициатором развития отрасли AI R&D в целом.

Работа организована следующим образом. В разд. 2 дано краткое описание стратегического документа США [4], а также заявленных в нем целей, задач и принципов их реализации. К одной из исключительных государственных прерогатив относится функция обеспечения обороны, общественной и государственной безопасности. Раздел 3 статьи посвящен анализу программных документов Министерства обороны США в части AI R&D. Представлены основные направления применения технологий искусственного интеллекта в области обороны и безопасности, а также названы основные государственные специализированные научно-исследовательские организации, формирующие спектр AI R&D: Управление перспективных исследовательских проектов Министерства обороны США (DARPA, Defence Advanced Research Projects Agency) и Агентство передовых исследований в сфере разведки (IARPA, Intelligence Advanced Research Projects Activity). На основе доступных источников проведен анализ направлений прикладных исследований данных организаций в области искусственного интеллекта. Раздел 4 посвящен внедрению приложений, использующих искусственный интеллект, в правоохранительных органах США: ФБР и полиции. Заключительные выводы по данной части цикла представлены в разд. 5.

2 Государственный подход к исследованиям и разработкам в области искусственного интеллекта в США

Исследования в области искусственного интеллекта проводятся в США давно. Американское правительство выделяло и выделяет на эти цели весьма большие средства. Например, в 2015 г. общий объем затрат США на AI R&D составил примерно 1,1 млрд долларов, что с учетом инфляции сейчас эквивалентно 80 млрд рублей. При этом размер частных инвестиций в сферу AI R&D точно назвать сложно, однако он, без сомнения, больше государственного. Известно, что объем венчурного (рискового) капитала, направленного на стартапы в области AI R&D, в период с 2013 по 2014 гг. вырос в 4 раза. Для сравнения: на проведение всех фундаментальных исследований в 2020 г. (включая РАН, государственные исследовательские центры, НИИ, Сколково и др.) до начала пандемии COVID-19 планировалось выделить 150 млрд рублей.

Понимание стратегической важности AI R&D в США нашло свое отражение в разработанной Национальным советом по науке и технике США (NSTC) в 2016 г. документа «Национальный стратегический план по исследованиям и разработке искусственного интеллекта» [4] (далее — Стратегия). В нем выделены направления AI R&D, которые будут получать государственное финансирование. Документ не фокусировался на AI R&D исключительно в военных целях.

Основной целью данных исследований было названо получение новых знаний и технологий в области AI R&D. Для ее достижения были определены следующие направления и ключевые принципы деятельности (стратегии, в терминах документа):

(1) **долговременные инвестиции в AI R&D:** горизонт ожидания результатов — 5–10 лет; «высокий риск – большие выгоды»; развитие методологий, ориентированных на данные, для получения новых знаний; улучшение возможностей восприятия у систем AI; исследование теоретических возможностей и ограничений AI; проведение исследований в области AI общего назначения; создание масштабируемых систем AI; поддержка исследований AI, подобного человеческому; создание более совершенных и надежных роботов; развитие аппаратной базы для улучшенного AI; разработка систем AI для модернизированной аппаратной базы;

(2) **разработка эффективных методов взаимодействия человека и AI:** поиск новых алгоритмов взаимодействия «человек–AI»; использование AI для развития человеческих способностей; развитие технологии визуализации и интерфейсов «человек–AI»; создание более эффективных систем лингвистической обработки;

(3) **выявление и решение этических, законодательных и социальных проблем использования AI:** повышение уровня справедливости, прозрачности и учета; построение систем AI, подчиненных этическим нормам; разработка архитектуры этического AI;

(4) **обеспечение безопасности использования систем AI:** повышение объяснимости и прозрачности; увеличение доверия; улучшение качества верификации и подтверждения; увеличение защищенности против атак различной природы; достижение долгосрочной безопасности систем AI и выравнивание качества;

(5) **разработка открытых массивов данных для обучения и тестирования систем AI:** создание и публикация широкого спектра массивов данных для нужд разработчиков AI; создание ресурсов для тренировок и тестов с учетов коммерческих и общественных интересов; разработка библиотек и пакетов с открытым кодом;

(6) **метрологическое обеспечение технологий искусственного интеллекта путем создания стандартов и критериев:** создание широкого спектра стандартов по AI; разработка технологических тестов; повышение доступности тестовых стендов для AI; вовлечение сообщества разработчиков и исследователей AI в процесс создания стандартов и тестов;

(7) **изучение и учет потребностей национальных разработчиков AI.**

В качестве своих составляющих стратегия включает следующие стратегические планы:

– Федеральный план исследований и разработок в области Big Data [5];

– Федеральный план исследований и разработок в области кибербезопасности [6];

– Национальную стратегию исследования защиты персональных данных [7];

– Национальный инициативный план по исследованиям в области нанотехнологий [8];

– Национальную компьютерную стратегическую инициативу [9];

– Инициативу по исследованию мозга с помощью инновационных нейротехнологий [10];

– Национальную робототехническую инициативу [11].

Согласно разработанной Стратегии, США планируют достичь следующих преимуществ путем внедрения технологий искусственного интеллекта:

– увеличить экономическое благосостояние (развитие производства, логистики, финансов, транспорта, сельского хозяйства, торговли, коммуникаций, науки и технологий);

– улучшить возможности обучения и качество жизни (образование, медицина, юриспруденция, персонализированные сервисы);

– укрепить государственную и общественную безопасность.

Летом 2019 г. по первым результатам выполнения Стратегии был опубликован отчет [12], содержащий исчерпывающий анализ результатов трех лет ее выполнения. По результатам первого этапа реализации Стратегии была скорректирована организационная структура для ее реализации. Она представлена в табл. 1.

Результатом анализа AI R&D стало определение еще одного, восьмого, направления в этой области:

(8) **расширение государственно-частного партнерства для ускорения развития.** Выделение этого направления связано с осознанием важности тезиса Стратегии об «устойчивом финансировании AI R&D в сотрудничестве с академическими, производственными международными партнерами и союзниками и другими негосударственными юридическими лицами для достижения технологических прорывов в области AI и смежных технологиях для их быстрейшей трансформации в реальные средства укрепления национальной и экономической безопасности США». Частно-государственное партнерство позволит эффективно перераспределять ресурсы и знания для скорейшего достижения результатов.

Таблица 1 Скорректированная версия организационной структуры реализации Стратегии

Общие положения в AI R&D	**AI R&D**	**Приложения**
Долговременные инвестиции (1-е направление)	Аналитика данных	Сельское хозяйство
	Восприятие	Коммуникации и связь
Взаимодействие «человек–AI» (2-е направление)		Оборона
	Теоретические ограничения	Образование
Этические, законодательные, социальные аспекты AI (3-е направление)	Общий (генеральный) AI	Финансы
	Масштабируемый AI	Государственные службы
Безопасность (4-е направление)		Здравоохранение
	AI человеческого уровня	Юриспруденция
Массивы данных (5-е направление)	Робототехника	Логистика
	Аппаратное обеспечение	Производство
Стандарты и критерии AI (6-е направление)	Автономные системы AI, взаимодействующие с человеком	Маркетинг
Изучение потребностей применения AI (7-е направление)	Расширение интеллектуальных возможностей человека	Персональные услуги
		Наука и инженерия
	Обработка естественных языков	Безопасность
Частно-государственное партнерство (8-е направление)	Интерфейсы и визуализация	Транспорт

3 Направления исследований и разработок в области искусственного интеллекта для обороны и безопасности

Министерство обороны США (DoD, Department of Defence) в рамках реализации Национальной оборонной инициативы США 2018 г. разработало и опубликовало документ [13], содержащий краткие сведения о стратегии страны по использованию технологий искусственного интеллекта в военных целях. Подразделением DoD, ответственным за выполнение этой стратегии, назван Объединенный центр искусственного интеллекта (JAIC, Joint Artificial Intelligence Center). В качестве организаций, ведущих работы по созданию и адаптации разработок в области AI для нужд DoD, указаны Управление перспективных исследовательских проектов Министерства обороны США (DARPA), Агентство передовых исследований в сфере разведки (IARPA), национальные исследовательские (в исходном тексте — *academic*) центры, частные компании, международные объединения.

В документе указаны следующие общие направления применения искусственного интеллекта для достижения стратегического военного преимущества.

Улучшение ситуационной осведомленности и принятия решений. Технологии AI, примененные для решения задач восприятия/осмысления, таких как анализ изображений, могут извлекать полезную информацию из сырых начальных данных и тем самым повышать ситуационную осведомленность командного состава. Искусственный интеллект может генерировать и помогать руководству исследовать оперативные сценарии, чтобы оно впоследствии могло осуществить оптимальный выбор для одновременного достижения поставленной цели и минимальных рисков как для вооруженных сил, так и для гражданского населения.

Повышение безопасности эксплуатации оборудования. Искусственный интеллект обладает серьезным потенциалом повышения безопасности эксплуатации военной техники (самолетов, кораблей, наземных транспортных и боевых средств) в сложных быстроменяющихся ситуациях, предупреждая операторов о скрытых опасностях.

Внедрение упреждающего обслуживания и снабжения — использование AI для прогнозирования выхода из строя критических частей, автоматической диагностики и планирования технического обслуживания на основе статистических данных и состояния оборудования. Подобная же технология будет использоваться для упреждающего управления поставками запасных частей и оптимизации уровня запасов. Эти преимущества обеспечат поддержку запасов на требуемом уровне, а также помогут в устранении неполадок и сделают возможным более быстрое развертывание средств и адаптацию затрат при снижении издержек.

Оптимизация деловых процессов. Искусственный интеллект будет использоваться с целью сокращения времени, затрачиваемого на ручные, повторя-

ющиеся и частые задачи. Оставляя людям задачи надзора за выполнением автоматизированных функций, AI имеет потенциал к сокращению числа и стоимости ошибок, увеличению выхода и степени приспособления, для направления ресурсов DoD на выполнение более приоритетных задач и достижение важнейших целей.

Ниже представлена информация о проектах в области AI R&D, проводимых «головными» специализированными исследовательскими организациями, упомянутыми в [13], — DARPA [14] и IARPA [15].

Управление DAPRA объявило в сентябре 2018 г. о многолетних инвестициях в размере более 2 млрд долл. США в новые и существующие программы, называемые кампанией «AI Next». Ключевые составляющие кампании предполагают автоматизацию критически важных бизнес-процессов Министерства обороны, таких как:

- проверка соответствия требованиям безопасности или аккредитация программных систем для оперативного развертывания;
- повышение надежности систем искусственного интеллекта;
- повышение безопасности и отказоустойчивости машинного обучения и технологий искусственного интеллекта;
- снижение зависимости от нехватки данных и производительности.

Управление перспективных исследовательских проектов Министерства обороны США создаст мощные возможности для DoD, опираясь на следующие принципы.

1. **Обеспечение новых возможностей.** Технологии AI применяются регулярно для реализации проектов DARPA, включая более 60 существующих программ, связанных с анализом сложных кибератак в реальном времени, обнаружением мошеннических изображений, построением динамических цепочек убийств, для ведения войн во всех областях, технологиях человеческого языка, мультимодальном автоматическом распознавании целей и др. Управление DARPA будет продвигать технологии AI для обеспечения автоматизации критически важных бизнес-процессов DoD. Один из таких процессов — аккредитация программных систем перед эксплуатационным развертыванием. Автоматизация этого процесса аккредитации с использованием известных AI и других технологий теперь представляется возможной.
2. **Обеспечение надежного искусственного интеллекта.** Технологии AI продемонстрировали большое значение для таких разнообразных задач, как анализ спутниковых изображений, предупреждение о кибератаках, логистика и др. В то же время режимы отказа/аварий технологий AI плохо изучены. Управление DARPA работает над решением этой проблемы, уделяя особое внимание исследованиям и разработкам — как аналитическим, так и эмпирическим.

3. **Учет фактора состязательности при разработке искусственного интеллекта.** Самый мощный инструмент AI на сегодня — это машинное обучение (ML, machine learning). Системы ML могут быть легко обмануты изменениями, которые никогда не обманут человека (так называемые *adverse samples*). Данные, используемые для обучения таких систем, могут быть повреждены, и само программное обеспечение (ПО) уязвимо для кибератак. Задачи обеспечения устойчивости ПО и защиты банков данных, используемых для обучения, планируется решать совместно и оперативно.
4. **Создание высокопроизводительного искусственного интеллекта.** Повышение производительности компьютеров за последнее десятилетие позволило добиться успеха в машинном обучении в сочетании с большими наборами данных и библиотеками ПО. Повышение производительности при снижении энергопотребления необходимо для развертывания центрального и тактического звеньев. Управление DARPA продемонстрировало аналоговую обработку алгоритмов AI с ускорением в 1000 раз и эффективностью энергопотребления в 1000 раз по сравнению с современными цифровыми процессорами, а также исследует аппаратные разработки для AI. DARPA также прикладывает усилия для снижения требований к маркированным данным обучения (качеству и объемам).
5. **Создание искусственного интеллекта следующего поколения.** Алгоритмы машинного обучения, обеспечивающие распознавание лиц, и автомобили с автоматическим вождением были изобретены более 20 лет назад. Управление DARPA заняло ведущее место в новаторских исследованиях по разработке следующего поколения алгоритмов AI, которые превратят компьютеры из инструментов в партнеров по решению проблем. Исследование DARPA направлено на то, чтобы дать возможность рассуждать (делать умозаключения).

Агентство передовых исследований в сфере разведки (IARPA) было создано в 2006 г. в целях проведения междисциплинарных исследований, определения новых возможностей и создания прорывных технологий в области разведки. Задачи IARPA должны решаться, основываясь на технических и эксплуатационных знаниях, которыми обладают органы разведки. Таким образом, миссия IARPA заключается в прогнозировании долгосрочных потребностей разведывательного сообщества и предоставлении ему технических и исследовательских возможностей.

В табл. 2 представлена классификация направлений исследований DARPA и IARPA, в той или иной степени базирующихся на AI R&D или касающихся AI R&D. Следует отметить, что доля подобных проектов у IARPA от общего объема проводимых исследований составляет 45%–55%, в DARPA этот показатель ниже. В данную классификацию не включены работы, связанные с различными беспилотными аппаратами, а также AI R&D в области химии, материаловедения и пр.

Таблица 2 Классификация AI R&D, проводимых DARPA и IARPA

Отрасль AI	Направление исследований	Проекты DARPA	Проекты IARPA
1. Системы обработки текстовой/речевой информации	– Глубокая обработка текста, выявление неочевидного смысла, выделение оперативно значимой информации – Интеллектуальная обработка малоресурсных языковых запросов с целью повышения ситуационной осведомленности – Системы распознавания речи на любом человеческом языке – Изучение использования метафор различными культурами, чтобы понять культурные нормы различных народов – Раскрытие социальных целей членов групп по их сообщениям – Разработка методов извлечения детализируемой семантической информации с акцентом на событиях в форме «кто–что–кому–когда-то»	DEFT, LORELEI	Babel, Metaphor, SCIL, BETTER, MATERIAL
2. Системы анализа и обработки изображений	– Разработка средств оценивания, манипуляций и подделки цифровых изображений – Аналитическая обработка видео- и аудиоинформации – Разработка системы распознавания спутниковых фотографий. – Разработка технологии автоматической генерации точных трехмерных моделей объектов с реальными физическими свойствами из множества источников данных изображений – Разработка ПО автоматического обнаружения движения для многокамерной среды потокового видео – Повышение производительности систем распознавания лиц – Системы обнаружения антропогенной деятельности	MefiFor	Aladdin Video, Amon-Hen, CORE3D, DIVA, Janus, SMART
3. Создание биометрических систем	– Разработка высокоточных систем сопоставления, получения биометрических сигнатур по данным низкого качества		BEST
4. Системы обработки специальных сигналов и данных	– Внедрение ML-методов для обработки сигналов радиочастотного спектра – Системы глубокой аналитической обработки данных геолокации – Разработка методов непрерывного анализа данных радиоэлектронной разведки	RFMLS	Finder, Mercury

Продолжение табл. 2 *на с.* 123

Таблица 2 (*продолжение*) Классификация AI R&D, проводимых DARPA и IARPA

Отрасль AI	Направление исследований	Проекты DARPA	Проекты IARPA
5. Системы оптимизации бизнес-процессов	– Разработка «блочного» инструментария описания проектов – Разработка математических основ и средств вычисления для управления и оптимизации процессов проектирования	FUN Design, TRADES	
6. Системы интеллектуальной обработки больших данных	– Альтернативные интерпретации событий по разнородным зашумленным конфликтующим источникам – Обработка реферативной информации с выделением причинно-следственных связей и объяснительных моделей – Обнаружение «сложных» событий и их формальное описание/классификация – Анализ сообщений средств массовой информации, обнаружение фактов подделок, фабрикаций и манипуляций, разработка средств для выполнения этих операций – Разработка и тестирование методов получения точных прогнозов для значительных достижений науки и техники по результатам открытых публикаций – Разработка средств получения актуальных данных из нескольких разрозненных быстропоявляющихся и меняющихся источников – Системы непрерывного анализа общедоступных данных для выявления/прогнозирования социально значимых явлений – Разработка гибридных систем геополитического прогнозирования	AIDA, Big Mechanism, KAIROS, SemaFor	ForeST, FUSE, KDD, OSI, HFC
7. Организация человеко-машинных систем	– Определение целей и ситуационных знаний партнеров-людей, прогнозирования их потребностей/действий для оптимизации работы команды «человек–компьютер» – Обеспечение «симметричной» связи с людьми (интонация, жесты, мимика) для улучшения обмена «сложной» информацией – Оптимизация операций, проводимых людьми и различными автономными аппаратами – Оптимизация человеческих адаптивных рассуждений, тестирование и проверка воздействий, повышающая производительность рассуждений – Разработка пользовательских сред, которые более динамичны, безопасны, проверяемы, переносимы и эффективны, чем текущие предложения	ASIST, CwC, CAML,	SHARP, VirtUE

Продолжение табл. 2 *на с.* 124

Таблица 2 (*продолжение*) Классификация AI R&D, проводимых DARPA и IARPA

Отрасль AI	Направление исследований	Проекты DARPA	Проекты IARPA
8. Обнаружение атак различного рода, компьютерная безопасность	– Разработка автоматизированных средств обнаружения распределенных атак, сбора корректных контекстных данных, коллективная выработка защитных действий – Системы безопасности компьютерных сетей – Разработка методов запуска конечными пользователями сомнительного с точки зрения безопасности ПО – Гомоморфные вычислительные методы шифрования с сокращением накладных расходов, разработка широкого спектра безопасных распределенных приложений, использующих передовые методы криптографии – Обнаружение атак с использованием биометрических представлений (искажение или скрытие биометрической идентичности) – Разработка систем тестирования и обнаружения внутренних угроз и нарушителей – Разработка средств создания и обнаружения «троянов» в библиотеках и средствах разработки AI	CHASE	ATHENA, CAUSE, STONESOUP, HECTOR, Odin, SCITE, TROjAI
9. Разработка средств построения моделей	– Разработка средств построения эмпирических моделей неспециалистами в статистике – Разработка средств моделирования в системах с отсутствием полных подтвержденных адекватных моделей – Повышение точности прогнозирования событий широкого спектра с учетом мнений экспертов	D3M, SD2	ACE
10. Совершенствование ML-методов и алгоритмов	– Повышение уровня объяснимости ML-моделей – Предпроектная оценка возможностей ML-систем, оценка достоверности из характеристик – Разработка ML-методов, базирующихся на малых обучающих выборках – Разработка ML-методов, базирующихся на реинжиниренге алгоритмов мозга	XAI, FunLoL LwLL	MICrONS
11. Совершенствование средств разработки ПО	– Создание автоматизированных средств создания, отладки, проверки, обслуживания и понимания ПО, борьба с программными уязвимостями	MUSE	

Окончание табл. 2 *на с.* 125

Таблица 2 (*окончание*) Классификация AI R&D, проводимых DARPA и IARPA

Отрасль AI	Направление исследований	Проекты DARPA	Проекты IARPA
12. Совершенствование аппаратно-программной базы	– Разработка реконфигурируемого высокопроизводительного аппаратно-программного обеспечения – Обеспечение беспроводным устройствам доступа к электромагнитному спектру, создание программно-определяемых радиостанций – Средства разработки микросхем/процессоров военного назначения – Разработка высокопроизводительных энергоэффективных компьютеров с использованием свойств сверхпроводимости	SDH, SC2, SDCPS	C3
13. Построение математических моделей мыслительных процессов человека	– Построение моделей групповых человеческих предубеждений – Построение когнитивных моделей восприятия смысла и представления концептуальных знаний, основанных на нейробиологии – Идентификация реальных пользователей по поведению в виртуальном мире (онлайн-игры, социальные сети) – Разработка игр для обучения распознавать и смягчать когнитивные искажения – Анализ инстинктивного (вазомоторного) поведения человека для оценки возможности доверия в условиях стресса/обмана – Разработка и тестирование систем, которые используют краудсорсинг и структурированные аналитические методы для улучшения аналитических рассуждений – Разработка и эмпирическая оценка систематических подходов к контрфактурному прогнозированию и извлеченным урокам – Разработка методов постоянного пассивного «зондирования» персонала с целью оценки эффективности его работы	UGB	ICArUS, KRNS, Reynard, Sirius, TRUST, CREATE, FOCUS, MOSAIC
14. Использование нового математического аппарата	– Использование квантовых эффектов для эффективного решения сложных задач комбинаторной оптимизации		QEO

4 Использование технологий искусственного интеллекта Федеральным бюро расследований США и полицией

Согласно открытым источникам [16], на момент весны 2019 г. Федеральное бюро расследований США (ФБР) уже использовало следующие элементы AI:

1. **Система распознавания лиц.** Федеральное бюро расследований обратилось к AI для разработки технологии идентификации следующего поколения (NGI, Next Generation Identification). Одной из его возможностей стало сравнение изображений для выявления тех лиц, которые связаны с преступной деятельностью, с использованием базы данных Межгосударственной фотосистемы. Использование ПО для распознавания лиц в ФБР имеет успехи. Федеральное бюро расследований также изучает коммерчески доступное ПО для видеонаблюдения, в частности Rekognition от Amazon.
2. **Система идентификации по отпечаткам пальцев.** Федеральное бюро расследований использует уже существующую систему IAFIS (Integrated Automated Fingerprint Identification System) прежде всего для сопоставления отпечатков пальцев из известной базы данных дактокарт, а система нового поколения расширила эту возможность. Новая система учитывает возможные отпечатки ладоней. Федеральное бюро расследований часто сталкивалось со стертыми или иным образом измененными отпечатками пальцев, которые мешали идентификации. Чтобы обойти эту сложность, агентство заказало разработку технологии AI, которая могла бы выполнять идентификацию при наличии подобных преднамеренных или непреднамеренных изменений.

Другие возможности NGI включают в себя:

– хранилище особого значения (RISC, Repository for Individuals of Special Concern) — это программа раннего выявления разыскиваемых и других опасных лиц с использованием базы данных отпечатков пальцев; предназначено для обеспечения сотрудников правоохранительных органов на месте и мгновенный доступ через мобильное устройство;

– файл расследования Friction Ridge — содержит все соответствующие изображения и события, связанные с человеком, для повышения точности поиска, включая отпечатки пальцев, RISC, дополнительные отпечатки пальцев и отпечатки ладоней в рамках Национальной системы отпечатков ладоней (NPPS, National Palm Print System);

– сервис Rap Back — регулярно сообщает о деятельности людей, находящихся под государственным расследованием или надзором, а также занимающих должности доверенного лица, чтобы избежать необходимости повторять проверку данных;

– «Холодное дело/Неизвестный мертвец» — позволяет агентам под прикрытием и следователям правоохранительных органов использовать базу данных NGI и алгоритмы поиска для идентификации неизвестных умерших;

– Iris Pilot — запущенный в конце 2013 г. пилотный проект по использованию человеческой радужки для биометрической идентификации в исправительных учреждениях, на границе, в программах условного освобождения, для мобильной идентификации на местах и в расследованиях с использованием видеодоказательств.

3. **Система проверки соответствия ДНК.** Федеральное бюро расследований признало значение профилей ДНК на ранних этапах уголовного расследования и в итоге создало базу данных под названием «Объединенная система индекса ДНК» (CODIS, Combined DNA Index System). До того как появились алгоритмы машинного обучения, правоохранительные органы посылали образцы в специальные лаборатории для их обработки, а затем отправляли профиль ДНК в ФБР для сопоставления. Весь процесс был медленным, занимал дни или недели в зависимости от отставания судебно-медицинской экспертизы и сложности поиска соответствия. Лабораторий, способных генерировать профили ДНК, было немного, и в некоторых областях отставание от дел могло достигать 5 лет. Кроме того, профили ДНК содержат сложный набор данных, и по состоянию на февраль 2019 г. CODIS имеет почти 14 млн профилей, поэтому стандартным компьютерным программам требуется время, чтобы просмотреть все данные и определить один профиль, если он там есть. Во многих случаях долгое ожидание может быть отложено, правосудие не состоится.

   Новая технология может изменить все это. Правоохранительные органы используют полностью автоматизированные и портативные машины Rapid DNA для ускорения процесса создания профилей ДНК из мазков на щеках, сокращая его до 90 мин вместо дней или недель. С помощью ПО для машинного обучения полиция может брать образцы, создавать профиль и сопоставлять данные в течение двух часов, если у них есть доступ к CODIS. Это именно то, что ФБР хочет сделать, руководствуясь Актом о быстрой ДНК от 2017 г. Среди прочего, это позволяет агентству создать сеть из этих машин, чтобы получить доступ к CODIS. В 2018 г. Национальный институт стандартов и технологий (NIST, National Institute of Standards and Technology) провел оценку работы комплексов быстрой ДНК. Результаты показали, что, хотя машины способны работать самостоятельно, модифицированный анализ с вмешательством человека дал результаты с более высокой точностью.

4. **Системы кибербезопасности** ФБР служат хранилищем большого объема данных, распределенных по многим полевым офисам, поэтому, как и в любой крупной организации, кибербезопасность становится серьезной проблемой. Федеральное бюро расследований заключило контракт с компанией ECS по кибербезопасности на управление безопасностью своей сети с помощью искусственного интеллекта. Компания ECS будет использовать машинное обучение, чтобы группы экспертов по безопасности проходили через сети ФБР, выявляли слабые места, сканировали потенциальные проблемные области и настраивали персонал для оценки и управления требованиями безопасности.

5. **Системы идентификации внутренних нарушителей.** Начиная с 1980-х гг. ФБР подвергалось многочисленным случаям инсайдерских угроз. Противо-

действие в этой области ФБР видит в более активном подходе к внутренним угрозам, запустив две программы, использующие аналитику данных для решения этой проблемы. Одна из них — Javelin, которая следит за внутренними проступками, нарушениями безопасности и внутренним шпионажем. Другая — это платформа анализа внутренних угроз (InTAP), которая просматривает большие объемы данных, чтобы найти шаблоны, указывающие на подозрительные действия и потенциальные угрозы для организации. Обе эти программы находятся в стадии реализации.

6. **Менеджмент бизнес-процессов.** Федеральное бюро расследований — это крупная организация, включающая много отделений по всему миру, в которых круглосуточно работают более 51 000 чел. Старые способы административного управления не всегда соответствуют растущим требованиям времени. Чтобы повысить операционную эффективность, ФБР заключило контракт с компанией-разработчиком ПО Pega Systems. Pega поможет ФБР оптимизировать операции и сократить расходы с помощью двух настраиваемых бизнес-решений: Pega Government Platform и Pega Robotic Process Automation. Платформа будет использоваться для разработки масштабируемых и гибких приложений для удовлетворения потребностей бизнес-процессов. Компонент автоматизации заключается в освобождении персонала от повторяющихся задач посредством автоматизации.

Согласно открытым источникам [17], на момент весны 2019 г. полиция США уже использовала многие элементы технологий AI. Искусственный интеллект и робототехнические приложения правоохранительных органов подразделяются на четыре большие категории (табл. 3):

(1) прогнозирование и анализ (Big Data и машинное обучение);

(2) распознавание (лиц в разнородной мультимедийной информации; голоса);

(3) патрулирование (автономные устройства, привязка к местности);

(4) коммуникации (обеспечение комплексной связи и доступа к информации и сервисам).

Внедрение AI направлено на усиление следующих качеств правоохранительных органов: масштабности, оперативности, эффективности, возможности удаленного влияния.

С помощью AI предполагается бороться со следующими видами негативных воздействий (атак):

– **цифровые атаки**: автоматический фишинг, обнаружение и использование киберуязвимостей и пр.,

– **политические атаки**: распространение фальшивых новостей и создание фальшивых средств массовой информации для появления путаницы или конфлик-

Таблица 3 Распределение приложений по категориям и степени готовности

Категория	Концепт	Прототип	Опытная эксплуатация	Боевое внедрение
Прогнозирование и анализ	Текстовый анализ и построение умозаключений Повышение чистоты в проведении расследований	Моделирование ситуаций с использованием агентов Предсказание протестов и преступлений Контекстный анализ интеллекта	Профилактические меры Системы цифровой криминалистики Идентификация подозрительного поведения	Идентификация юридически значимой информации Предсказание преступлений
Распознавание	Идентификация машин Анализ видео- и аудиоинформации	Выделение информации из ежедневных сводок о преступлениях Лицевая и текстовая биометрия	Анализ аудио- и видеоинформации из мест заключения Справочные машины Анализ и распознавание голоса в системах связи Системы наблюдения за преступниками	
Коммуникации		Аудиопереводчик	Коммуникационные роботы	
Патрулирование		Роботы для патрулирования периметра	Патрульные дроны (места заключения, граница) Дроны-наблюдатели Генерируемые AI ленты новостей, телеканалы прямого эфира	

тов, подмена и фальсификация инструментов, позволяющих манипулировать видео и ставить под угрозу доверие политическим деятелям, и пр.,

– **физические атаки**: использование вооруженных дронов с функцией распознавания, контрабанда и пр.

В качестве уже используемых элементов AI в [17] упоминаются:

– дроны и роботы: полицейские роботы, наблюдение в тюрьмах, роботы безопасности Knightscope для аэропортов, дроны в правоохранительных органах;

– системы распознавания лиц в правоохранительных органах с использованием компьютерного зрения и систем машинного обучения;
– системы наружного наблюдения с использованием AI: роботизированные птицы-наблюдатели, умные очки, камеры с поддержкой AI, ЭЭГ-повязки, униформа с поддержкой AI, браслеты с поддержкой AI.

5 Заключение

Анализируя российскую «Национальную стратегию развития искусственного интеллекта на период до 2030 года» [3], ее американский аналог, а также другие рассмотренные выше документы США в области AI R&D, можно сделать следующие выводы.

1. Российская стратегия разработана коммерческой организацией (пусть и с государственным участием) — Сбербанком, в то время как американский стратегический план — государственной организацией: Национальным советом по науке и технике США (NSTC, National Science and Technology Council).
2. Декларируемые цели примерно одни и те же: «*обеспечение роста благосостояния и качества жизни населения, обеспечение национальной безопасности и правопорядка, достижение устойчивой конкурентоспособности российской экономики, в том числе лидирующих позиций в мире в области искусственного интеллекта*» в [3] и «*создание новых знаний и технологий в области AI, обеспечивающих обществу ряд позитивных преимуществ при минимизации негативных воздействий*» в [4]. Тем не менее цели США выглядят более абстрактно и фундаментально (создание новых знаний и технологий) по сравнению с достаточно утилитарной, но сложно проверяемой российской версией (обеспечение роста благосостояния и качества жизни).
3. В качестве «адресатов» российская стратегия упоминает государственные программы РФ и отдельных субъектов, федеральные и региональные проекты, плановые и программно-целевые документы госкорпораций и компаний, акционерных обществ с государственным участием и стратегические документы иных организаций в части AI. Американский стратегический план устанавливает набор целей для исследований AI R&D, финансируемых из федерального бюджета, как исследований, проводимых в рамках правительства, так и исследований, которые проводятся за пределами правительства, например в академических кругах. В 2019 г. США внесли в свой план дополнительный пункт о расширении государственно-частного партнерства.
4. Координацию деятельности участников реализации российской стратегии осуществляет Правительственная комиссия по цифровому развитию, использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности. В США аналогичную

функцию выполняет Подкомитет R&D в области сетевых и информационных технологий (NITRD, Networking and Information Technology Research and Development) Национального совета по науке и технике США.

5. Американский стратегический план детально более проработан, так как опирается на совокупность стратегических документов, регламентирующих R&D по частным направлениям [5–11], согласованно разработанных с основным документом [4].

6. Согласно доступным открытым источникам, планы AI R&D в области обороны и безопасности детально проработаны. При этом акцент в них сделан не на внедрение уже имеющихся интеллектуальных технологий, а на фундаментальные исследования, которые дадут стратегический прикладной выигрыш в будущем. Из анализа конкурсов/проектов следует, что государство не финансирует проекты, сомнительные с точки зрения их выполнимости (сомнение в существовании решения поставленной задачи, некорректность или размытость в постановке задачи, отсутствие современной или перспективной аппаратно-программной базы для решения задачи).

7. В [3] говорится о наличии большого числа открытых библиотек и банков данных, предназначенных для AI R&D. Этот факт не подлежит сомнению и порождает соблазн использовать готовые библиотеки и банки в качестве базы создания новых AI-продуктов и технологий. Однако директивные и проектные документы DARPA содержат серьезное предупреждение об опасности такого шага в виде тезиса о состязательности в области AI R&D. Это означает, что перед созданием AI-приложений в области обороны и безопасности процедуры открытых библиотек AI R&D следует изначально считать недоверенными (содержащими вредоносный код, ошибки, неэффективные алгоритмы), а банки обучения — сфабрикованными (*adverse sampled*).

Деятельность индивидуумов, организаций и государств в конкурентной среде имеет свой целью извлечение прибыли или достижение конкурентных преимуществ. Польза последних не всегда очевидна и сложна в сравнении из-за невозможности их монетарной оценки. Для государства именно достижение конкурентных преимуществ служит стратегической целью. В связи с этим большая фундаментальная научная направленность американского стратегического плана и смежных документов AI R&D выглядит более предпочтительной по сравнению с преимущественным внедрением имеющихся AI-технологий в экономике и социальной сфере. Особую жизненную важность конкурентные преимущества приобретают в таких государственных областях, как оборона и обеспечение общественной и государственной безопасности. Из американских документов также можно сделать вывод о том, что в случае дилеммы «прибыль – конкурентное преимущество» предлагается безальтернативный выбор в пользу последнего, обеспечивающего американский государственный и национальный приоритет.

Литература

1. *Борисов А. В., Босов А. В., Жуков Д. В.* Стратегия исследований и разработок в области искусственного интеллекта I: Основные понятия и краткая хронология // Системы и средства информатики, 2021. Т. 31. № 1. С. 57–68.
2. *Борисов А. В., Босов А. В., Жуков Д. В.* Стратегия исследований и разработок в области искусственного интеллекта II: Сравнительный анализ наукометрических показателей в мире и в Российской Федерации // Системы и средства информатики, 2021. Т. 31. № 2. С. 89–107.
3. Национальная стратегия развития искусственного интеллекта на период до 2030 года. Указ Президента РФ от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации».
4. The National Artificial Intelligence Research and Development Strategic Plan. — National Science and Technology Council, Networking and Information Research and Development Sub-committee, 2016. www.nitrd.gov/pubs/national_ai_rd_strategic_plan.pdf.
5. Federal Big Data Research and Development Strategic Plan, 2016. www.nitrd.gov/pubs/bigdatardstrategicplan.pdf
6. Federal Cybersecurity Research and Development Strategic Plan, 2016. www.nitrd.gov/cybersecurity/publications/2016_Federal_Cybersecurity_Research_and_Development_Strategic_Plan.pdf.
7. National Privacy Research Strategy, 2016. www.nitrd.gov/pubs/NationalPrivacyResearchStrategy.pdf.
8. National Nanotechnology Initiative Strategic Plan, 2014. www.nano.gov/sites/default/files/pub_resource/2014_nni_strategic_plan.pdf.
9. National Strategic Computing Initiative Strategic Plan, 2016. www.whitehouse.gov/sites/whitehouse.gov/files/images/NSCI% 20Strategic% 20Plan.pdf.
10. Brain Research through Advancing Innovative Neurotechnologies (BRAIN), 2013. www.whitehouse.gov/BRAIN.
11. National Robotics Initiative, 2011. www.whitehouse.gov/blog/2011/06/24/developing-next-generation-robots.
12. The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update: A Report by the Select Committee on Artificial Intelligence of the National Science and Technology Council, 2019. www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf.
13. Summary of the 2018 Department of Defense Artificial Intelligence Strategy, 2018. media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/summary-of-dod-ai-strategy.pdf.
14. DARPA: Our research. www.darpa.mil/our-research.
15. IARPA: Research programs. www.iarpa.gov/index.php/research-programs.
16. Artificial Intelligence at the FBI — 6 Current Initiatives and Projects. www.emerj.com/ai-sector-overviews/artificial-intelligence-fbi/.
17. Artificial intelligence in olicing — use-cases, ethical concerns, and trends. www.emerj.com/ai-sector-overviews/artificial-intelligence-in-policing/.

Поступила в редакцию 20.02.21

RESEARCH AND DEVELOPMENT STRATEGY IN THE FIELD OF ARTIFICIAL INTELLEGENCE III: UNITED STATES GOVERNMENT SUPPORT DOCTRINE

A. V. Borisov, A. V. Bosov, and D. V. Zhukov

Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The article continues the cycle of works devoted to the analysis of the impact of public administration on the effectiveness of research and development in the field of artificial intelligence (AI R&D). The third part of the cycle is a study of government influence on AI R&D using the United States as an example. A brief description of the U.S. strategic document in the field of AI R&D is given including a compact statement of its goals and objectives as well as implementation principles. In addition, a related document of the US Department of Defense was analyzed. A classification analysis of AI R&D directions in the field of defense and security carried out by the main specialized research organizations of the United States is presented.

Keywords: artificial intelligence (AI); US Department of Defense (DoD); Defense Advanced Research Projects Agency (DARPA); Intelligence Advanced Research Projects Agency (IARPA)

DOI: 10.14357/08696527210410

References

1. Borisov, A. V., A. V. Bosov, and D. V. Zhukov. 2021. Strategiya issledovaniy i razrabotok v oblasti iskusstvennogo intellekta I: Osnovnye ponyatiya i kratkaya khronologiya [Reseach and development strategy in the field of artificial intelligence I: Basic concepts and brief chronology]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(1):57–68.
2. Borisov, A. V., A. V. Bosov, and D. V. Zhukov. 2021. Strategiya issledovaniy i razrabotok v oblasti iskusstvennogo intellekta II: Sravnitel'nyy analiz naukometricheskikh pokazateley v mire i v Rossiyskoy Federatsii [Reseach and development strategy in the field of artificial intelligence II: Comparative analysis of scientometric indicators in the world and in the Russian Federation]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 31(2):89–107.
3. O razvitii iskusstvennogo intellekta v Rossiyskoy Federatsii: ukaz Prezidenta ot 10.10.2019 No. 490 [About strategy of scientific and technological development of the Russian Federation. Presidential Decree No. 490 dated 10.10.2019]. Available at: http://static.kremlin.ru/media/events/les/ru/AH4x6HgKWANwVtMOfPDhcbRpvd1HCCsv.pdf (accessed October 5, 2021).
4. National Science and Technology Council, Networking and Information Research and Development Sub-committee. 2016. The National Artificial Intelligence Research and Development Strategic Plan. Available at: www.nitrd.gov/pubs/national_ai_rd_strategic_plan.pdf (accessed October 5, 2021).
5. Federal Big Data Research and Development Strategic Plan. Available at: www.nitrd.gov/pubs/bigdatardstrategicplan.pdf (accessed October 5, 2021).

6. Federal Cybersecurity Research and Development Strategic Plan. 2016. Available at: www.nitrd.gov/cybersecurity/publications/2016_Federal_Cybersecurity_Research_and_Development_Strategic_Plan.pdf (accessed October 05, 2021).
7. National Privacy Research Strategy. 2016. Available at: www.nitrd.gov/pubs/NationalPrivacyResearchStrategy.pdf (accessed October 05, 2021).
8. National Nanotechnology Initiative Strategic Plan. 2014. Available at: www.nano.gov/sites/default/files/pub_resource/2014_nni_strategic_plan.pdf (accessed October 5, 2021).
9. National Strategic Computing Initiative Strategic Plan. 2016. Available at: www.whitehouse.gov/sites/whitehouse.gov/files/images/NSCI%20Strategic%20Plan.pdf (accessed October 5, 2021).
10. Brain Research through Advancing Innovative Neurotechnologies (BRAIN). 2013. Available at: www.whitehouse.gov/BRAIN (accessed October 5, 2021).
11. National Robotics Initiative. 2011. Available at: www.whitehouse.gov/blog/2011/06/24/developing-next-generation-robots (accessed October 5, 2021).
12. The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update. 2019. A Report by the Select Committee on Artificial Intelligence of the National Science and Technology Council. Available at: www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf (accessed October 5, 2021).
13. Summary of the 2018 Department of Defense Artificial Intelligence Strategy. 2018. Available at: media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/summary-of-dod-ai-strategy.pdf (accessed October 5, 2021).
14. DARPA: Our research. Available at: www.darpa.mil/our-research (accessed October 05, 2021).
15. IARPA: Research programs. Available at: www.iarpa.gov/index.php/research-programs (accessed October 5, 2021).
16. Artificial Intelligence at the FBI — 6 Current Initiatives and Projects. Available at: www.emerj.com/ai-sector-overviews/artificial-intelligence-fbi/ (accessed October 5, 2021).
17. Artificial intelligence in olicing — use-cases, ethical concerns, and trends. Available at: www.emerj.com/ai-sector-overviews/artificial-intelligence-in-policing/ (accessed October 5, 2021).

Received February 20, 2021

Contributors

Borisov Andrey V. (b. 1965) — Doctor of Science in physics and mathematics, principal scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; ABorosov@ipiran.ru

Bosov Alexey V. (b. 1969) — Doctor of Science in technology, principal scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; AVBosov@ipiran.ru

Zhukov Denis V. (b. 1979) — principal specialist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; DZhukov@ipiran.ru

УСИЛЕННЫЙ АЛГОРИТМ ТОКЕНИЗАЦИИ ДЛЯ ЗАЩИТЫ ПЕРСОНАЛЬНЫХ ДАННЫХ*

А. А. Грушо[1], *Д. В. Смирнов*[2], *Е. Е. Тимонина*[3], *С. Я. Шоргин*[4]

Аннотация: Одним из методов обезличивания персональных данных (ПД) является токенизация. Этот метод представляет собой взаимно однозначную замену фрагментов ПД на случайные элементы некоторого множества. Одна из слабостей защиты ПД с помощью токенизации заключается в возможности статистической оценки вероятностей встречаемости защищаемых фрагментов ПД. Предложен метод усиления алгоритмов токенизации, позволяющий преодолеть указанную слабость. Усиленный алгоритм токенизации по сложности незначительно отличается от других алгоритмов. При этом усиленный алгоритм может быть использован как в случаях токенизации с помощью замен алфавитов, описывающих различные фрагменты ПД, так и в случаях, когда ПД разбиваются на фрагменты одинаковой длины и преобразуются во фрагменты той же длины, но в других алфавитах.

Ключевые слова: информационная безопасность; обезличивание персональных данных; токенизация; математическая статистика

DOI: 10.14357/08696527210411

1 Введение

Проблема защиты ПД актуальна во многих областях цифровой обработки информации и обмена данными. Сведения о ПД и транзакциях в банковской сфере представляют интерес для многих преступных группировок. Криминальные структуры собирают ПД всюду, где они появляются и недостаточно защищены. Большое значение защита ПД имеет при обмене медицинскими данными.

Нормативная база и стандарты по защите ПД разработаны достаточно хорошо. В РФ всем известен Федеральный закон № 152 о защите ПД [1]. Нормативным актом по выполнению требований защиты ПД служит Приказ ФСТЭК № 21 [2]. Информация о ПД, которую необходимо защищать, состоит

*Работа частично поддержана РФФИ (проект 18-29-03081).

[1]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, grusho@yandex.ru

[2]ПАО Сбербанк России, dvlsmirnov@sberbank.ru

[3]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, eltimon@yandex.ru

[4]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, sshorgin@ipiran.ru

из прямых и косвенных признаков идентификации. Примеры прямых признаков идентификации ПД:

- ФИО;
- девичья фамилия;
- адрес проживания;
- номер мобильного телефона.

Примеры косвенных признаков идентификации ПД:

- данные о рождении;
- данные о семье;
- личные данные (хобби, происхождение доходов, остатки на счетах, медицинский диагноз, свидетельства о медицинских обследованиях);
- данные о карьере.

Наиболее распространенный метод защиты ПД — обезличивание. Этот метод используется для предотвращения несанкционированного использования ПД при сохранении возможностей дальнейшей обработки информации, связанной с ПД. В РФ действуют приказы по обезличиванию ПД:

- Роскомнадзора № 996 [3],
- Росстата № 165 [4],
- Минздрава РФ № 341н [5].

2 Анализ защищенности при обезличивании

Для обезличивания обычно используют замену различных фрагментов ПД на куски случайных последовательностей. Такой заменой может быть подстановка вместо букв алфавита букв другого алфавита или замена сочетаний из k букв на фиксированные случайные сочетания такой же длины. Описанный подход называется токенизацией [6]. Продукт, реализующий токенизацию, Protegrity Vaultless Tokenization [7], признан экспертами одним из лучших в мире. Он использует несколько алфавитов, предназначенных для различных типов данных, каждый алфавит использует свою замену. Это означает, что преобразование частей ПД происходит в разных информационных пространствах [8].

Использование замен на несколько подряд идущих символов значительно увеличивает размер базы данных для хранения заменяемых сочетаний и требует дополнительных средств защиты такой базы данных.

Во всех подобных методах следует иметь в виду следующую слабость. Все перечисленные виды замен используют сочетания знаков, которые встречаются в ПД, т. е. распределение вероятностей на ПД совпадает с распределением вероятностей на последовательностях, получаемых после замены. Следовательно, существует инвариант, не зависящий от типа замены. Таким инвариантом

являются вероятности встречаемости сочетаний знаков, подвергаемых заменам. Эти вероятности встречаемости можно оценить статистически. При наличии инсайдеров такие оценки можно сделать по большому числу ПД из хранилища данных. Существуют также другие методы. Это означает, что есть возможность использовать незащищенную часть информации и скомпрометировать частично ПД [9, 10].

3 Защита токенов при токенизации

Для защиты токенов от статистических атак можно предложить следующий метод, не требующий предварительной статистической обработки. Пусть таблица, где в первой строке записаны защищаемые фрагменты ПД, которые преобразуются в токены (вторая строка таблицы), состоит не из двух строк, а из матрицы с одинаковыми длинами столбцов. Иными словами, каждому защищаемому фрагменту ПД, или комбинации знаков длины k, может быть поставлено в соответствие несколько комбинаций случайных знаков длины k.

4 Алгоритм усиления токенизации

Введем следующие обозначения. Пусть $\mathcal{A} = \{a\}$ — алфавит записи ПД, который будет заменен на алфавит $\mathcal{B} = \{b\}$, состоящий из случайных символов, аналогичных символам $\mathcal{A}$. Тогда мощности обоих алфавитов равны $|\mathcal{A}| = |\mathcal{B}| = L$. Параметр T описывает единое время информационной системы (ИС), в котором выделено начальное значение t_0. Все обезличенные данные в ИС имеют одинаковый формат: (ПД, t, данные), где t означает время обезличивания ПД. Защите подлежат только ПД, остальные параметры передаются в неизменном виде. Определим параметр τ, имеющий значение промежутка времени действия одной замены алфавита $\mathcal{A}$. Тогда для любого момента времени t существует неотрицательное целое число p такое, что $t \in [t_0 + p\tau, t_0 + (p+1)\tau)$.

Пусть задана управляющая равновероятная последовательность $\overline{\gamma}$ на сервере, где должна проводиться токенизация и обращение токенов в ПД. Обозначим элементы этой последовательности γ_n, $n = 1, 2 \ldots$, и пусть эти элементы принимают значения 1 и 2. Определим последовательность $\overline{v} = \{v_p\}$ следующим образом:

$$v_p = v_{p-1} + \gamma_p\,, \quad p = 1, 2, \ldots, \quad v_0 = 0\,.$$

Таблица соответствия алфавита $\mathcal{A}$ алфавиту $\mathcal{B}$ в математике называется подстановкой, будем обозначать ее через π. Осуществим сдвиг строки $B = (b_1, \ldots, b_L)$, $b_m \in \mathcal{B}$, в подстановке π на одну единицу вправо по циклу, т. е. последний элемент строки B переходит на первое место. Построенная строка $B^{(1)}$ по-прежнему состоит из символов алфавита $\mathcal{B}$ и содержит все символы $\mathcal{B}$.

Таким образом, построена новая подстановка $\pi^{(1)}$, в которой присутствуют элементы строки первоначальной подстановки π, но номера символов в строке $B^{(1)}$ вычисляются из номеров строки B по формуле:

$$b_m^{(1)} = b_{m-1} (\mathrm{mod}\, L)\,, \;\; m = 1, \ldots, L\,,$$

где нумерация мест идет слева направо. Аналогично сдвиг строки B подстановки π на p единиц вправо по циклу позволит построить новую подстановку $\pi^{(p)}$, в которой

$$b_m^{(p)} = b_{m-p} (\mathrm{mod}\, L)\,.$$

Построим связь времени обезличивания с подстановкой π, по которой осуществляется обезличивание. Поскольку процедура обезличивания сопровождается меткой времени t, то из того, что $t \in [t_0 + p\tau, t_0 + (p+1)\tau)$ можно однозначно вычислить значение параметра p. Из случайной последовательности $\overline{\gamma}$ вычисляем значение γ_{p+1}. Отсюда вычисляется значение v_{p+1}, причем значение v_p — это суммарный сдвиг в ранее построенной подстановке с элементами $b_m^{(v_p)} = b_{m-v_p} (\mathrm{mod}\, L)$. Это означает, что обезличивание поступивших данных происходит по подстановке, в которой $b_m^{(v_{p+1})} = b_{m-v_{p+1}} (\mathrm{mod}\, L)$, т. е. новая подстановка π^* отличается от предыдущей сдвигом второй строки по циклу вправо по $\mathrm{mod}\, L$ на один или на два шага в зависимости от значения γ_{p+1}.

Использование неравномерного сдвига при неизвестной первоначальной подстановке позволяет полностью скрыть закономерности реальных сдвигов первоначальной подстановки π, т. е. не позволяет восстановить реальную величину сдвига уже через несколько шагов. Если используется обычный алфавит, то реальную величину сдвига также невозможно восстановить. Одна подстановка действует в течение времени τ. Невозможность восстановления реальной величины сдвига не позволяет накапливать статистику и использовать статистический метод для компрометации ПД.

5 Эффективность алгоритмов обезличивания персональных данных и восстановления персональных данных из обезличенных данных

Эффективный алгоритм обезличивания строится следующим образом. Алфавит $\mathcal{A}$ упорядочивается, по крайней мере лексикографически. Это позволяет быстро находить защищаемый фрагмент ПД в первоначальной подстановке. Поскольку первоначальная подстановка считается заданной в явном виде, то необходимо найти действующий в данный момент сдвиг. Определение значения p реализуется быстро. Предыдущие преобразования ПД позволяют легко вычислить необходимую величину сдвига. В самом деле, γ_p и v_{p-1} известны, что дает возможность вычислить параметр сдвига для нужной подстановки. Этот сдвиг для данного фрагмента быстро вычисляется либо по первой строке A, либо по

второй строке B. Совокупность указанных простых операций позволяет быстро найти нужное значение для преобразования ПД.

Подстановка, обратная для первоначальной, строится следующим образом. Случайные элементы алфавита $\mathcal{B}$ должны быть упорядочены для того, чтобы обеспечить быстрый поиск в подстановке, обратной к первоначальной. Для подстановки, обратной к первоначальной, значение ПД просто находится по известному случайному элементу.

Рассмотрим алгоритм восстановления ПД для обезличенных в момент t данных. Время t порождает значение p, которое участвует в формировании величины сдвига. Самая сложная операция в данном алгоритме — это вычисление v_p, которое, в отличие от прямого алгоритма, трудно сопрячь с предыдущим преобразованием. Некоторым упрощением является запоминание для последовательности p чисел 1 и 2 в последовательности $\overline{\gamma}$. Пусть элемент $b \in B$, созданный в момент времени t, необходимо обратить. Элемент b соответствует некоторому сдвигу первоначальной подстановки. Тогда по подстановке, обратной к первоначальной, можно найти b. По b находим соответствующее $a \in A$. Искомое a^* отстоит от a на величину сдвига, который определяется величиной v_p. Отсюда и из первоначальной подстановки находим искомое значение a^*.

Таким образом, при достаточно быстром нахождении v_p сложности прямого и обратного алгоритмов почти не отличаются.

Пример. Пусть $\mathcal{A} = \{a, b, c\}$ и $\mathcal{B} = \{3, 2, 1\}$, $\tau = 1$, для $p = 0, 1, 2, \ldots$ задана последовательность $\overline{\gamma} = (1, 2, 2, 1, \ldots)$, $t_0 = 0$. Построим прямую π и обратную π^{-1} к первоначальной подстановке:

$$\pi = \begin{pmatrix} a\ b\ c \\ 3\ 2\ 1 \end{pmatrix};\ \ \pi^{-1} = \begin{pmatrix} 1\ 2\ 3 \\ c\ b\ a \end{pmatrix}.$$

В соответствии с параметром p рассмотрим моменты времени $t_1 \in [0, 1)$, т. е. $p = 0$, $t_2 \in [1, 2)$, т. е. $p = 1$, $t_3 \in [2, 3)$, т. е. $p = 2$, и т. д. Построим начальный участок последовательности $\overline{v} = (0, 1, 3(\mathrm{mod}\,3) = 0, 2, \ldots)$. Тогда подстановки замен следующие.

1. Для $p = 0$ имеем $v_0 = 0$ и

$$\pi = \begin{pmatrix} a\ b\ c \\ 3\ 2\ 1 \end{pmatrix};\ \ \pi^{-1} = \begin{pmatrix} 1\ 2\ 3 \\ c\ b\ a \end{pmatrix}.$$

2. Для $p = 1$ имеем $v_1 = 1$ и

$$\pi^{(v_1)} = \pi^{(1)} = \begin{pmatrix} a\ b\ c \\ 1\ 3\ 2 \end{pmatrix};\ \ \left(\pi^{(1)}\right)^{-1} = \begin{pmatrix} 1\ 2\ 3 \\ a\ c\ b \end{pmatrix}.$$

3. Для $p = 2$ имеем $v_2 = 0$ и

$$\pi^{(v_2)} = \pi = \begin{pmatrix} a & b & c \\ 3 & 2 & 1 \end{pmatrix}; \quad \pi^{-1} = \begin{pmatrix} 1 & 2 & 3 \\ c & b & a \end{pmatrix}.$$

Построим алгоритм обращения токенов. Сначала для t_1 вычислим p, получим $p = 0$. Пусть ПД были преобразованы в число 2. Число 2 в обращении π^{-1} первоначальной подстановки π определяет букву b, т. е. ПД $= b$.

Для t_2 находим $p = 1$ и $v_1 = 1$. Так как число 2 переходит в обращении первоначальной подстановки в ПД $= b$, необходимо провести в строке A сдвиг, равный 1. Отсюда символ c и есть фрагмент ПД.

Для t_3 имеем $p = 2$ и $v_2 = 0$. Тогда 2 в обращении π^{-1} первоначальной подстановки π определяет букву b, т. е. ПД $= b$.

6 Заключение

Для обезличивания ПД часто используется токенизация, состоящая в замене по некоторой случайной подстановке фрагментов ПД. В этом случае вероятности появления значений заменяемых фрагментов ПД совпадают с вероятностями появлений самих фрагментов ПД. Статистические оценки вероятностей этих значений могут скомпрометировать ПД.

Для усиления защищенности ПД от описанных статистических атак предложено использовать множество подстановок фрагментов ПД. Такие подстановки получаются из исходной с помощью сдвигов, зависящих от времени обезличивания и случайной управляющей последовательности.

Сложность построенных алгоритмов не намного отличается от сложности простейшего преобразования токенизации по подстановке.

Литература

1. О персональных данных: Федеральный закон от 27.07.2006 № 152-ФЗ (последняя редакция). http://www.consultant.ru/document/cons_doc_LAW_61801.
2. Об утверждении состава и содержания организационных и технических мер по обеспечению безопасности персональных данных при их обработке в информационных системах персональных данных: Приказ ФСТЭК России от 18 февраля 2013 г. № 21. 19 с. https://fstec.ru/normotvorcheskaya/akty/53-prikazy/691.
3. Об утверждении требований и методов по обезличиванию персональных данных: Приказ Роскомнадзора от 05.09.2013 № 996. 7 с. https://legalacts.ru/doc/prikaz-roskomnadzora-ot-05092013-n-996-ob.
4. Об утверждении Методологических положений по формированию массивов деперсонифицированных микроданных годового структурного обследования по форме федерального статистического наблюдения N 1-предприятие «Основные сведения о деятельности организации» общего пользования для представления пользователям в аналитических целях: Приказ Росстата от 19.04.2013 № 165. 9 с. https://docs.cntd.ru/document/499020817/.

5. Об утверждении Порядка обезличивания сведений о лицах, которым оказывается медицинская помощь, а также о лицах, в отношении которых проводятся медицинские экспертизы, медицинские осмотры и медицинские освидетельствования: Приказ Министерства здравоохранения Российской Федерации от 14.06.2018 № 341н. 7 с. http://publication.pravo.gov.ru/Document/View/0001201808090005.
6. Tokenization 101: Understanding the Basics. https://www.wexinc.com/insights/blog/corporate-payments-edge/credit-card-tokenization-basics.
7. Protegrity Vaultless Tokenization. https://s3.amazonaws.com/ptymarketingcollateral/Vaultless_Tokenization_FAQs_APRIL_12.pdf.
8. *Грушо А. А., Забежайло М. И., Смирнов Д. В., Тимонина Е. Е.* Модель множества информационных пространств в задаче поиска инсайдера // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 65–69.
9. *Denning D., Akl S., Heckman M., et al.* Views for multilevel database security // IEEE T. Software Eng., 1987. Vol. SE-13. P. 129–140.
10. *Грушо А. А., Грушо Н. А., Забежайло М. И., Смирнов Д. В., Тимонина Е. Е.* Параметризация в прикладных задачах поиска эмпирических причин // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 62–66.

Поступила в редакцию 23.09.21

ENHANCED TOKENIZATION ALGORITHM FOR PERSONAL DATA PROTECTION

***A. A. Grusho*[1], *D. V. Smirnov*[2], *E. E. Timonina*[1], *and S. Ya. Shorgin*[1]**

[1]Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation
[2]Sberbank of Russia, 19 Vavilov Str., Moscow 117999, Russian Federation

Abstract: Tokenization is one of the methods of depersonalizing personal data. This method is a bijective replacement of fragments of personal data with random elements of a certain set. One of the weaknesses of personal data protection through tokenization is the possibility of statistically assessing the probabilities of the occurrence of protected fragments of personal data. The paper proposes a method of enhancing tokenization algorithms which allows overcoming this weakness. The enhanced tokenization algorithm is slightly different in complexity from other algorithms. At the same time, the enhanced algorithm can be used both in cases of tokenization by replacing alphabets describing various fragments of personal data and in cases where personal data are divided into fragments of the same length and converted into fragments of the same length but in other alphabets.

Keywords: information security; depersonalization of personal data; tokenization; mathematical statistics

DOI: 10.14357/08696527210411

Acknowledgments

The paper was partially supported by the Russian Foundation for Basic Research (project 18-29-03081).

References

1. O personal'nykh dannykh: Federal'nyy zakon 152-FZ [About personal data: Federal law 152-FZ]. July 27, 2006. Available at: http://www.consultant.ru/document/cons_doc_LAW_61801 (accessed September 15, 2021).
2. FSTEC Rossii. February 18, 2013. Ob utverzhdenii sostava i soderzhaniya organizatsionnykh i tekhnicheskikh mer po obespecheniyu bezopasnosti personal'nykh dannykh pri ikh obrabotke v informatsionnykh sistemakh personal'nykh dannykh: Prikaz No. 21 [On approval of the composition and content of organizational and technical measures to ensure the security of personal data during their processing in personal data information systems: Order No. 21]. 19 p. Available at: https://fstec.ru/normotvorcheskaya/akty/53-prikazy/691 (accessed September 15, 2021).
3. Roskomnadzor. September 5, 2013. Ob utverzhdenii trebovaniy i metodov po obezlichivaniyu personal'nykh dannykh: Prikaz No. 996 [On approval of requirements and methods for the depersonalization of personal data: Order No. 996]. 7 p. Available at: http://rkn.gov.ru/docs/doc_981.tiff (accessed September 15, 2021).
4. Rosstat. April 19, 2013. Ob utverzhdenii Metodologicheskikh polozheniy po formirovaniyu massivov depersonifitsirovannykh mikrodannykh godovogo strukturnogo obsledovaniya po forme federal'nogo statisticheskogo nablyudeniya N 1-predpriyatie "Osnovnye svedeniya o "deyatel'nosti organizatsii" obshchego pol'zovaniya dlya predstavleniya pol'zovatelyam v analiticheskikh tselyakh: Prikaz No. 165 [On adoption of Methodological provisions for the formation of arrays of depersonalized microdata of the annual structural inspection in the form of federal statistical observation N 1-enterprise "Main data on the activity of the organization" for general use to provide users with analytical purposes: Order No. 165]. 9 p. Available at: https://docs.cntd.ru/document/499020817/ (accessed September 15, 2021).
5. Ministerstva zdravookhraneniya Rossiyskoy Federatsii. June 14, 2018. Ob utverzhdenii Poryadka obezlichivaniya svedeniy o litsakh, kotorym okazyvayetsya meditsinskaya pomoshch', a takzhe o litsakh, v otnoshenii kotorykh provodyatsya meditsinskie ekspertizy, meditsinskie osmotry i meditsinskie osvidetel'stvovaniya: Prikaz No. 341n [On approval of the statement of the Order on depersonalization of information about persons to whom medical care is provided as well as about persons in respect of whom medical expertise, medical examinations, and medical certifications are performed: Order No. 341n]. 7 p. Available at: http://publication.pravo.gov.ru/Document/View/0001201808090005 (accessed September 15, 2021).
6. Tokenization 101: Understanding the basics. Available at: https://www.wexinc.com/insights/blog/corporate-payments-edge/credit-card-tokenization-basics/ (accessed September 15, 2021).
7. Protegrity Vaultless Tokenization. Available at: https://s3.amazonaws.com/ptymarketingcollateral/Vaultless_Tokenization_FAQs_APRIL_12.pdf (accessed September 15, 2021).

8. Grusho, A. A., M. I. Zabezhailo, D. V. Smirnov, and E. E. Timonina. 2017. Model' mnozhestva informatsionnykh prostranstv v zadache poiska insaydera [The model of the set of information spaces in the problem of insider detection]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):65–69.
9. Denning, D., S. Akl, M. Heckman, *et al.* 1987. Views for multilevel database security. *IEEE T. Software Eng.* SE-13:129–140.
10. Grusho, A. A., N. A. Grusho, M. I. Zabezhailo, D. V. Smirnov, and E. E. Timonina. 2018. Parametrizatsiya v prikladnykh zadachakh poiska empiricheskikh prichin [Parametrization in applied problems of search of the empirical reasons]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):62–66.

Received September 23, 2021

Contributors

Grusho Alexander A. (b. 1946) — Doctor of Science in physics and mathematics, professor, principal scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; grusho@yandex.ru

Smirnov Dmitry V. (b. 1984) — business partner for IT security department, Sberbank of Russia, 19 Vavilov Str., Moscow 117999, Russian Federation; dvlsmirnov@sberbank.ru

Timonina Elena E. (b. 1952) — Doctor of Science in technology, professor, leading scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; eltimon@yandex.ru

Shorgin Sergey Ya. (b. 1952) — Doctor of Science in physics and mathematics, professor, principal scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; sshorgin@ipiran.ru

АНАЛИЗ НЕПРЕРЫВНОСТИ ПОЛЬЗОВАТЕЛЬСКОЙ СЕССИИ В БЕСПРОВОДНЫХ СИСТЕМАХ ТЕРАГЕРЦЕВОГО ДИАПАЗОНА*

***В. А. Бесчастный*[1], *Д. Ю. Острикова*[2], *В. С. Шоргин*[3], *Д. А. Молчанов*[4], *Ю. В. Гайдамака*[5]**

Аннотация: В настоящее время внимание производителей оборудования и операторов связи обращено на сети следующего, шестого (6G), поколения, работающие на частотах терагерцевого (ТГц) диапазона. Однако использование диапазона чрезвычайно высоких частот приводит к существенным потерям при распространении сигнала, для сокращения которых приходится использовать меньшую ширину диаграммы направленности как на стороне базовой станции (БС), так и на стороне пользовательского устройства (ПУ), что позволяет добиться высоких коэффициентов усиления сигнала. В результате эти системы оказываются подвержены влиянию не только динамической блокировки линии прямой видимости до БС, но и микромобильности, т. е. незначительного изменения ориентации устройства относительно своего центра. Негативные эффекты этих явлений можно смягчить, используя функциональные возможности множественного подключения, или принципа мультисвязности, позволяющего пользователям подключаться к нескольким БС одновременно, а затем переключаться между ними в случае потери сигнала. В работе представлена математическая модель для анализа вероятности успешного завершения сессии в условиях блокировки прямой видимости и микромобильности, построенная с учетом особенностей распространения сигнала в ТГц-диапазоне, плотности развертывания БС, а также функциональных возможностей множественного подключения. Полученные результаты показывают, что положительный эффект от множественного подключения наблюдается при одновременном использовании до 5 каналов связи и существенно зависит от длительности допустимого интервала отсутствия связи, вызванного процедурой поиска луча.

Ключевые слова: терагерцевый диапазон; множественное подключение; микромобильность; отсутствие связи; вероятность блокировки; поиск луча

DOI: 10.14357/08696527210412

*Публикация выполнена при поддержке Программы стратегического академического лидерства РУДН и при финансовой поддержке РФФИ (проекты 19-07-00933 и 20-07-01064).

[1]Российский университет дружбы народов, beschastnyy-va@rudn.ru

[2]Российский университет дружбы народов, ostrikova-dyu@rudn.ru

[3]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, vshorgin@ipiran.ru

[4]Университет Тампере, Финляндия, dmitri.moltchanov@tuni.fi

[5]Российский университет дружбы народов; Федеральный исследовательский центр «Информатика и управление» Российской академии наук, gaydamaka-yuv@rudn.ru

1 Введение

В настоящее время, когда процесс стандартизации технологии пятого поколения «новое радио» (*англ.* 5G New Radio, 5G NR), использующей частоты миллиметрового диапазона, почти завершен, активно ведутся исследования новой технологии радиоинтерфейса для систем 6G [1]. По общему мнению, этот новый интерфейс будет работать в нижней части терагерцевого (0,3–3 ТГц) диапазона частот.

Для предотвращения серьезных потерь при распространении сигнала в системах ТГц-диапазона будут использованы массивные фазированные антенные решетки (ФАР) с сотнями элементов как на стороне БС, так и на стороне ПУ. Такие ФАР способны создавать диаграммы направленности излучения антенны с шириной луча всего несколько градусов и даже меньше [2], что позволит значительно снизить межсотовую интерференцию и повысить коэффициент усиления сигнала в направлении передачи и приема. Однако это также создаст ряд уникальных проблем для разработчиков системы. В частности, к усилившемуся по сравнению с системами миллиметрового диапазона эффекту блокировки прямой видимости (Line of Sight, LoS) добавится эффект от микромобильности, т. е. от незначительного изменения наклона устройства в руке пользователя по вертикальной и горизонтальной осям, а также от небольшого смещения по этим осям [3]. Для восстановления активного соединения, нарушенного в результате микромобильности, требуется процедура поиска луча, занимающая временны́е и энергетические ресурсы системы. Производительность ТГц-систем при наличии сбоев из-за микромобильности пользователя была исследована в [4], где авторы продемонстрировали обменные соотношения между долей времени в условиях отсутствия связи и достигнутой пропускной способностью канала.

Одна из возможностей избежать блокировки сигнала в системах терагерцевого и миллиметрового диапазонов — использование стандартизированной 3GPP-процедуры множественного подключения, или мультисвязности [5], в соответствии с которой ПУ может поддерживать два или более одновременных соединения с соседними БС и использовать их в случае потери текущего канала. В контексте систем миллиметрового диапазона процедура множественного подключения была изучена в работах [6, 7]. Использование принципа мультисвязности для ТГц-систем было исследовано в [8], где авторы продемонстрировали его положительное влияние на спектральную эффективность. Однако принцип мультисвязности можно также использовать и для борьбы с эффектами микромобильности.

Проведенные до сих пор исследования принципа мультисвязности как в миллиметровом, так и в ТГц-диапазоне концентрировались на средних показателях производительности, таких как вероятность сброса сессии или средняя спектральная эффективность. Однако на практике сетевым приложениям для нормальной работы часто требуется непрерывное подключение к сети. Решить эту проблему значительно сложнее, поскольку для этого требуется фиксировать из-

менение состояния пользователя во времени. Именно на решение этой проблемы и направлена данная работа, в которой построена математическая модель для оценки непрерывности пользовательской сессии в условиях блокировки прямой видимости и микромобильности пользователя при наличии множественных подключений, а также получено выражение для плотности распределения времени непрерывного соединения, учитывающее конкретные условия распространения сигнала в ТГц-диапазоне, плотность развернутых базовых станций ТГц-диапазона частот (ТГц-БС), конфигурацию антенн, механизм поиска луча и степень мультивязности (число одновременно подключенных БС). Затем для заданного интервала времени, который приложение может провести в условиях отсутствия связи, приводится выражение для расчета вероятности успешного завершения сессии.

2 Системная модель

В работе расположение ТГц-БС соответствует точечному пуассоновскому процессу (Poisson Point Process, PPP) в $\mathbb{R}^2$ с плотностью λ_A. Высота ТГц-БС постоянна и равна h_T. Пользовательское устройство случайным образом расположено в зоне обслуживания ТГц-БС; следовательно, плотность распределения расстояния до ближайшей i-й ТГц-БС описывается как [9]

$$f_i(x) = \frac{2(\pi\lambda_A)^i}{(i-1)!}\, x^{2i-1} e^{-\pi\lambda_A x^2}, \quad x > 0, \;\; i = 1, 2, \ldots$$

Предположим, что связь возможна только при наличии прямой видимости между ПУ и ТГц-БС. Линия прямой видимости может блокироваться телом человека, т. е. люди в пешеходной зоне вокруг ПУ действуют как потенциальные блокаторы с плотностью λ_B и движутся в соответствии с моделью случайного направления (Random Direction Model, RDM) с постоянной скоростью v. Блокаторы моделируются как цилиндры с радиусом r_B. Предположим, что рост людей постоянен и равен h_B, высота расположения ПУ постоянна и равна h_U, причем $h_v \leq h_B$. В соответствии с [10] время между последовательными событиями блокировки распределено экспоненциально. Таким образом, процесс событий блокировки прямой видимости до i-й ТГц-БС является однородным пуассоновским с интенсивностью $\mu_{B,i}$:

$$\mu_{B,i} = \int_0^\infty f_i(x) \frac{(x[h_B - h_U] + r_B[h_T - h_U])}{(2r_B\lambda_B v)^{-1}(h_T - h_U)}\, dx\,. \tag{1}$$

Значение SNR (Signal to Noise Ratio, отношение сигнал–шум) на ПУ в состоянии LoS может быть записано как [11]

$$S(x) = P_T G_T G_U \left[\frac{x^{-\zeta_T} e^{-Kx}}{N_0} \right],$$

где ζ_T — коэффициент затухания сигнала; N_0 — тепловой шум; P_T — излучаемая мощность антенны БС; G_T и G_U — коэффициенты усиления антенн БС и ПУ; K — коэффициент абсорбции.

Аналогично [4] будем описывать процесс микромобильности как броуновское движение со случайным процессом смещения по осям $x(t)$ и $y(t)$, а также вращениями вокруг этих осей $\varphi(t)$ и $\theta(t)$ соответственно. Смещения по оси z и повороты относительно этой оси не учитываются, так как показано, что их вклад в эффект расхождения лучей ничтожно мал [4]. Таким образом, распределение времени до наступления микромобильности совпадает с распределением времени первого достижения одной из границ броуновской частицей:

$$f_{T_A}(t) = f_\theta(t)\left[1 - F_\phi(t)\right] + f_\phi(t)\left[1 - F_\theta(t)\right],$$

компоненты которого определяются как

$$f_{(\cdot)}(t) = \frac{M_{\Phi\Theta}}{\sqrt{4\pi D_{(\cdot)} t^3}} \exp\left(-\frac{-M_{\Phi\Theta}^2}{4D_{(\cdot)} t} \right),$$

где $D_{(\cdot)}$ — коэффициент диффузии; $M_{\Phi\Theta} = (102\pi/360)(1/N_T + 1/N_U)$.

Предположим, что как в случае блокировки прямой видимости (состояние nLoS), так и в случае расхождения лучей ТГц-БС инициируют итеративный алгоритм поиска луча, для которого время поиска составляет $T_B = (N_U + N_T)\sigma$, где N_U и N_T — число элементов антенных решеток соответственно на ПУ и ТГц-БС; σ — время однократного переключения решетки.

Чтобы оценить влияние мультисвязности на непрерывность пользовательской сессии, предположим, что ПУ поддерживает $N = 1, 2\ldots$ каналов с ближайшими ТГц-БС. Поскольку отношение сигнал–шум обратно пропорционально расстоянию между ТГц-БС и ПУ, ближайшая ТГц-БС обеспечивает наилучшее качество канала с точки зрения усредненного по времени SNR. Процедура поиска луча предполагает выбор ТГц-БС с наибольшим значением SNR на данный момент, т. е. ближайшую ТГц-БС с прямой видимостью. В рассматриваемой модели поиск луча запускается в момент, когда активное соединение потеряно из-за микромобильности или блокировки LoS.

В следующем разделе представлена математическая модель для анализа непрерывности пользовательской сессии в беспроводных системах ТГц-диапазона.

3 Математическая модель

Интервалы наличия и отсутствия связи в рассматриваемой системе образуют альтернирующий процесс восстановления. Идея основана на применении теории

полумарковских процессов [12], позволяющей характеризовать альтернирующий процесс восстановления матрицей интервальных переходных вероятностей $\mathbf{\Phi}(t)$, состоящей из условных вероятностей переходов между состояниями за время t и распределений длительностей этих переходов.

Пространство состояний полумарковского процесса можно разделить на пять подмножеств:

(1) состояния активного соединения с БС, $j = 1, \ldots, N$;

(2) состояния поиска луча при блокировке LoS, $j = N + 1, \ldots, 2N$;

(3) состояния поиска луча при микромобильности, $j = 2N + 1, \ldots, 3N$;

(4) состояния временного отсутствия связи, $j = 3N + 1, \ldots, 4N$;

(5) нулевое поглощающее состояние.

Процесс описывается вектором вероятностей начальных состояний $\boldsymbol{\pi} = [\pi_i]$, $i = 0, \ldots, 4N$, и матрицей переходных вероятностей $\mathbf{U} = [u_{ij}]$, $i, j = 0, \ldots, 4N$.

Согласно [13], вероятность нахождения в состоянии блокировки LoS с i-й БС равна

$$p_{N,i} = 1 - \int\limits_0^\infty \frac{2(\pi\lambda_A)^i}{(i-1)!} \frac{x^{2i-1}}{e^{\pi\lambda_A x^2}} e^{-2xr_B\lambda_B(h_B-h_U)/(h_A-h_U)} dx\,.$$

Длительность блокировки прямой видимости до i-й станции можно найти как период занятости системы массового обслуживания (СМО) $M/\mathrm{GI}/\infty$ [14]:

$$F_{B,i}(t) = 1 - \left(\int\limits_0^t [1 - F_{B,i}(t-z)] \left| de^{-\mu_{B,i}F_T(z)} \right| + \right.$$
$$\left. + [1 - F_T(t)] \left[1 - \int\limits_0^t \frac{1 - F_{B,i}(t-z)}{e^{\mu_{B,i}F_T(z)}} \mu_{B,i}\, dz \right] \right),$$

где $F_T(t) = H(t - 2r_B/v)$ — функция распределения (ФР) времени пересечения блокатором зоны блокировки, представленная в виде функции Хэвисайда. Тогда ФР времени одновременной блокировки LoS до всех N станций можно найти как минимум из длительностей блокировок на каждой из станций:

$$F_{\mathrm{NL}}(t; i) = 1 - [1 - F_{B,i}(t)] \prod_{j=1, j\neq i}^{N} \left[1 - F_{B_j^*}(t) \right],$$

где $F_{B_j^*}(t)$ — время остаточной блокировки LoS на j-й станции.

Используя тот же принцип нахождения минимума из случайных величин, матрицу $\mathbf{U}$ можно представить в следующем виде:

$$u_{i,N+i} = 1 - \int_0^\infty \int_y^\infty f_{T_A}(y) f_{T_L,i}(x)\, dxdy\,, \;\; i = 1, \ldots, N\,;$$

$$u_{i,2N+i} = \int_0^\infty \int_y^\infty f_{T_A}(y) f_{T_L,i}(x)\, dxdy\,, \;\; i = 1, \ldots, N\,;$$

$$u_{N+i,j} = (1 - H(T_B)) \prod_{k=1}^{j-1} p_{N,k}\,, \;\; i, j = 1, \ldots, N\,,\; j \le i\,;$$

$$u_{2N+i,i} = (1 - H(T_B)) \prod_{k=1}^{i-1} p_{N,k}\,, \;\; i = 1, \ldots, N\,;$$

$$u_{N+i,3N+i} = (1 - H(T_B)) \prod_{k=1, k\neq i}^{N} p_{N,k}\,, \;\; i = 1, \ldots, N\,;$$

$$u_{N+i,0} = H(T_B)\,, \;\; i = 1, \ldots, 2N\,;$$

$$u_{3N+i,2N+i} = F_{\mathrm{NL}}(T_N; i) \prod_{k=1}^{i-1} p_{N,k}\,, \;\; i = 1, \ldots, N\,;$$

$$u_{3N+i,0} = 1 - F_{\mathrm{NL}}(T_N; i), \;\; i = 1, \ldots, N\,;$$

$$u_{0,0} = 1\,, \;\; i = 1, \ldots, N\,;$$

$$u_{i,j} = 0 \text{ иначе,}$$

где $f_{T_A}(t)$ — распределение времени до наступления микромобильности; $f_{T_L,i}(t)$ — экспоненциальное распределение времени до наступления блокировки LoS до i-й станции с параметром $\mu_{B,i}$, представленным в (1).

Аналогично можно найти распределения длительностей переходов в виде матрицы $\mathbf{F}(t) = [f_{ij}(t)]$, $i, j = 0, \ldots, 4N$:

$$f_{i,N+i}(t) = f_{T_L,i}(t)\,(1 - F_{T_A}(t))\,, \;\; i = 1, \ldots, N\,;$$

$$f_{i,2N+i}(t) = f_{T_A}(t)\,(1 - F_{T_L,i}(t))\,, \;\; i = 1, \ldots, N\,;$$

$$f_{N+i,j}(t) = \delta(t - T_B)\,, \;\; i = 1, \ldots, N\,;$$

$$f_{2N+i,j}(t) = \delta(t - T_B)\,, \;\; i = 1, \ldots, N\,;$$

$$f_{N+i,3N+i}(t) = \delta(t - T_B)\,, \;\; i = 1, \ldots, N\,;$$

$$f_{N+i,0} = \delta(t - T_N)\,, \;\; i = 1, \ldots, 2N\,;$$

$$f_{3N+i,2N+i}(t) = f_{\mathrm{NL}}(t; i) F_{\mathrm{NL}}(T_N; i)\,, \;\; i = 1, \ldots, N\,;$$

$$f_{3N+i,0}(t) = f_{\mathrm{NL}}(t + T_N; i) F_{\mathrm{NL}}(T_N; i)\,, \;\; i = 1, \ldots, N\,;$$

$$f_{i,j}(t) = 0 \text{ иначе}\,,$$

где $\delta(t)$ — дельта-функция Дирака; T_N — длительность допустимого интервала отсутствия связи.

Найдя матрицы переходных вероятностей и распределений длительностей условных переходов, можно перейти к промежуточной диагональной матрице $\mathbf{\Psi}^* = \left[\psi_{ij}^*\right]$, $i, j = 0, \ldots, 4N$, элементы которой выражаются как

$$\psi_{ii}^*(s) = \frac{1}{s}\left(1 - \sum_{j=0}^{4N} u_{ij} f_{ij}^*(s)\right), \quad i = 0, \ldots, 4N;$$
$$\psi_{ij}^*(s) = 0 \text{ иначе}.$$

Далее, применяя обратное преобразование Лапласа к выражению

$$\mathbf{\Phi}^*(s) = \left[\mathbf{I} - \mathbf{U} \circ \mathbf{F}^*(s)\right]^{-1} \mathbf{\Psi}^*(s),$$

где $\mathbf{U} \circ \mathbf{F}$ — произведение Адамара; $\mathbf{I}$ — единичная матрица, можно численно найти исходные значения интервально-переходных вероятностей матрицы $\mathbf{\Phi}(t)$.

Наконец, определим распределение времени до обрыва соединения и вероятность успешного завершения пользовательской сессии как

$$f_U(t) = 1 - \sum_{i=1}^{4N} \pi_i \Phi_{i0}(t); \quad p_U = \int_0^{T_C} f_U(t)\, dt,$$

где T_C — полное время сессии.

4 Численный анализ

В данном разделе проводится численный анализ влияния системных параметров, представленных в таблице, на рассматриваемые характеристики.

На рис. 1 представлен график зависимости вероятности успешного завершения сессии от числа обслуживающих БС. Ожидаемо, что увеличение числа БС, которые использует ПУ для мультисвязности, влечет рост вероятности того, что сессия не будет потеряна до своего окончания. Стоит отметить, что наиболее эффективно характеристика откликается на рост числа БС вплоть до пяти, после чего эффект от увеличения их числа резко снижается.

Требования к качеству предоставления услуг диктуют для приложений различные длительности допустимого интервала отсутствия связи, т. е. такого прерывания соединения, которое не приводит к обрыву всей сессии в целом.

На рис. 2 показано влияние полной длительности сессии на вероятность ее завершения при различных допустимых интервалах прерывания и одной обслуживающей БС. Здесь $T_N = 5$ мс и 5 с соответствуют приложениям, которые

Системные параметры

Обозначение	Описание	Значения по умолчанию
λ_A	Плотность ТГц-БС	10–3 шт$/$м^2
λ_B	Плотность блокаторов	0,3 шт$/$м^2
r_B	Радиус блокатора	0,4 м
h_B	Высота блокатора	1,7 м
h_U	Высота ПУ	1,5 м
h_T	Высота ТГц-БС	4 м
v	Скорость блокатора	1 м$/$с
P_T	Излучаемая мощность на БС	2 Вт
ζ_T	Коэффициент затухания	2,1
K	Коэффициент абсорбции	0,2
σ	Время переключения массива антенной решетки	2 мкс
T_C	Полная длительность сессии	30 с
T_N	Допустимый интервал прерывания	1 с
N_T	Число конфигураций антенны БС	64×64
N_U	Число конфигураций антенны ПУ	4×4
N	Число ТГц-БС	1–10
N_O	Шум	-84 дБ
$\Delta\theta, \Delta\phi$	Средняя скорость смещения вокруг осей	0,1 град$/$с
$\Delta X, \Delta Y$	Средняя скорость смещения от осей	3 см$/$с

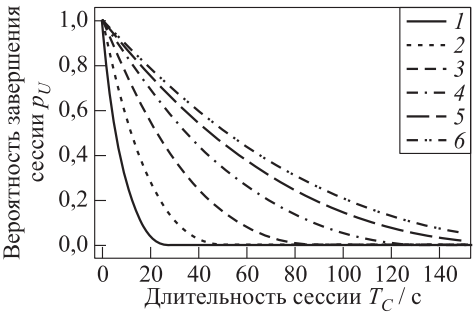


Рис. 1 Вероятность успешного завершения сессии для разных степеней мультисвязности: *1* — $N = 1$; *2* — 2; *3* — 3; *4* — 4; *5* — 5; *6* — $N = 6$

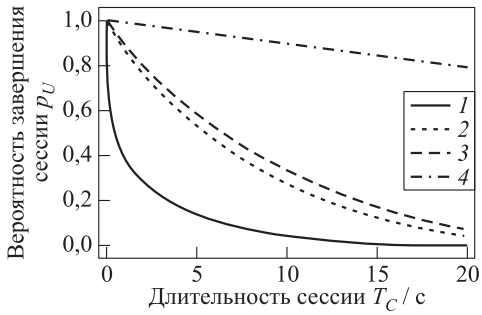


Рис. 2 Вероятность успешного завершения сессии при различных допустимых интервалах отсутствия связи: *1* — $T_N =$ $= 5$ мс; *2* — 50 мс; *3* — 0,5 с; *4* — $T_N = 5$ с

крайне чувствительны и невосприимчивы к прерываниям соответственно. В первом случае и блокировка LoS, и микромобильность приводят к сбросу сессии, так как оба эффекта всегда длятся более 5 мс. Также можно заметить, что, хотя в случае $T_N = 0{,}5$ с допустимый интервал на порядок больше, чем при $T_N = 5$ мс, разница в вероятности успешного завершения сессии не столь суще-

ственна. Это объясняется тем, что оба значения позволяют с успехом справляться с микромобильностью, но не устойчивы к блокировкам LoS.

5 Заключение

В работе исследованы временны́е характеристики процесса обслуживания пользовательской сессии в сетях ТГц-диапазона частот в условиях динамической микромобильности пользователей и блокировки прямой видимости подвижными блокаторами. Для этой цели использована полумарковская структура, характеризующая альтернирующий процесс наличия и отсутствия связи с помощью матрицы интервальных переходных вероятностей. Проведен численный анализ вероятности успешного завершения сессии в зависимости от разных степеней мультисвязности и допустимых интервалов отсутствия связи. В частности, показано, что положительный эффект мультисвязности наблюдается при одновременном использовании до 5 каналов связи и сильно зависит от продолжительности сессии. Вероятность успешного завершения сессии для приложений, чувствительных к длительности допустимого интервала отсутствия связи, в основном зависит от микромобильности, в результате которой необходимо выполнять процедуру поиска луча. Задача дальнейших исследований — анализ влияния микромобильности на энергоэффективность систем ТГц-диапазона.

Литература

1. *Polese M., Jornet J. M., Melodia T., Zorzi M.* Toward end-to-end, full-stack 6G terahertz networks // IEEE Commun. Mag., 2020. Vol. 58. No. 11. P. 48–54. doi: 10.1109/MCOM.001.2000224.
2. *Akyildiz I. F., Jornet J. M.* Realizing ultra-massive MIMO (1024 × 1024) communication in the (0.06–10) Terahertz band // Nano Commun. Netw., 2016. Vol. 8. P. 46–54. doi: 10.1016/j.nancom.2016.02.001.
3. *Petrov V., Moltchanov D., Koucheryavy Y., Jornet J. M.* The effect of small-scale mobility on terahertz band communications // 5th ACM Conference (International) on Nanoscale Computing and Communication Proceedings. — New York, NY, USA: Association for Computing Machinery, 2018. Art. 40. 2 p. doi: 10.1145/3233188.3242902.
4. *Petrov V., Moltchanov D., Koucheryavy Y., Jornet J. M.* Capacity and outage of terahertz communications with user micro-mobility and beam misalignment // IEEE T. Veh. Technol., 2020. Vol. 69. No. 6. P. 6822–6827. doi: 10.1109/TVT.2020.2988600.
5. 3GPP. Technical Specification 37.340; NR; Multi-connectivity; Overall description (Release 16), June 2021. https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3198.
6. *Gapeyenko M., Petrov V., Moltchanov D., Akdeniz M. R., Andreev S., Himayat N., Koucheryavy Y.* On the degree of multi-connectivity in 5G millimeter-wave cellular urban deployments // IEEE T. Veh. Technol., 2019. Vol. 68. No. 2. P. 1973–1978. doi: 10.1109/TVT.2018.2887343.

7. *Begishev V., Sopin E., Moltchanov D., Pirmagomedov R., Samuylov A., Andreev S., Koucherayvy Y., Samouylov K.* Performance analysis of multi-band microwave and millimeter-wave operation in 5G NR systems // IEEE T. Wirel. Commun., 2021. Vol. 20. No. 6. P. 3475–3490. doi: 10.1109/TWC.2021.3051027.
8. *Shafie N. Y., Han C.* Multi-connectivity for indoor terahertz communication with self and dynamic blockage // IEEE Conference (International) on Communications Proceedings. — Piscataway, NJ, USA: IEEE, 2020. Art. 9148716. 7 p. doi: 10.1109/ICC40277.2020.9148716.
9. *Moltchanov D.* Distance distributions in random networks // Elsevier Ad Hoc Networks, 2012. Vol. 10. P. 1146–1166. doi: 10.1016/j.adhoc.2012.02.005.
10. *Begishev V., Moltchanov D., Sopin E., Samuylov A., Andreev S., Koucheryavy Y., Samouylov K.* Quantifying the impact of guard capacity on session continuity in 3GPP new radio systems // IEEE T. Veh. Technol., 2019. Vol. 68. No. 12. P. 12345–12359. doi: 10.1109/TVT.2019.2948702.
11. *Boronin P., Petrov V., Moltchanov D., Koucheryavy Y., Jornet J. M.* Capacity and throughput analysis of nanoscale machine communication through transparency windows in the terahertz band // Nano Commun. Netw., 2014. Vol. 5. No. 3. P. 72–82. doi: 10.1016/j.nancom.2014.06.001.
12. *Cinlar E.* Markov renewal theory // Adv. Appl. Probab., 1969. Vol. 1. No. 2. P. 123–187. doi: 10.2307/1426216.
13. *Gerasimenko M., Moltchanov D., Gapeyenko M., Andreev S., Koucheryavy Y.* Capacity of multiconnectivity mmWave systems with dynamic blockage and directional antennas // IEEE T. Veh. Technol., 2019. Vol. 68. No. 4. P. 3534–3549. doi: 10.1109/TVT.2019.2896565.
14. *Gapeyenko M., Samuylov A., Gerasimenko M., Moltchanov D., Singh S., Akdeniz M. R., Aryafar E., Himayat N., Andreev S., Koucheryavy Y.* On the temporal effects of mobile blockers in urban millimeter-wave cellular scenarios // IEEE T. Veh. Technol., 2017. Vol. 66. No. 11. P. 10124–10138. doi: 10.1109/TVT.2017.2754543.

Поступила в редакцию 12.10.21

UNINTERRUPTED CONNECTIVITY TIME PERFORMANCE ANALYSIS IN TERAHERTZ SYSTEMS

***V. A. Beschastnyi*[1], *D. Yu. Ostrikova*[1], *V. S. Shorgin*[2], *D. A. Moltchanov*[3], *and Yu. V. Gaidamaka*[1,2]**

[1]Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation
[2]Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation
[3]Tampere University, 7 Korkeakoulunkatu, Tampere 33720, Finland

Abstract: Terahertz (THz) band is considered as the main candidate for new radio access technology in sixth-generation (6G) cellular systems. Requiring large antenna arrays at base station (BS) and user equipment (UE) sides to compensate for extreme path losses, THz systems will utilize extremely directional antenna radiation patterns. As a result, the performance of these systems will be severely affected by not only blockage but also UE micromobility in hands of a user. The negative effects of these phenomena can be alleviated by utilizing the multiconnectivity functionality that allows UE to maintain two or more links to nearby BSs and use them when the currently active link is lost. By accounting for THz specific propagation, antenna and beam searching design, the density of THz BS deployment, and multiconnectivity operation, the successful session completion probability under both types of impairments has been investigated. The present results indicate that the gains of multiconnectivity are observed up to 5 simultaneously supported links and heavily depend on the application outage tolerance time and are mostly affected by micromobility. To improve it, one needs to ensure that the application may tolerate outage caused by beam searching time which is of the order of milliseconds.

Keywords: terahertz communications; micromobility; outage; multiconnectivity; human body blockage; beam searching

DOI: 10.14357/08696527210412

Acknowledgments

The publication has been prepared with the support of RUDN University Strategic Academic Leadership Program and funded by the Russian Foundation for Basic Research according to the research projects No. 19-07-00933 and No. 20-07-01064.

References

1. Polese, M., J. M. Jornet, T. Melodia, and M. Zorzi. 2020. Toward end-to-end, full-stack 6G terahertz networks. *IEEE Commun. Mag.* 58(11):48–54. doi: 10.1109/MCOM.001.2000224.

2. Akyildiz, I. F., and J. M. Jornet. 2016. Realizing ultra-massive mimo (1024 × 1024) communication in the (0.06–10) terahertz band. *Nano Commun. Netw.* 8:46–54. doi: 10.1016/j.nancom.2016.02.001.
3. Petrov, V., D. Moltchanov, Y. Koucheryavy, and J. M. Jornet. 2018. The effect of small-scale mobility on terahertz band communications. *5th ACM Conference (International) on Nanoscale Computing and Communication Proceedings*. New York, NY: ACM. 40. 2 p. doi: 10.1145/3233188.3242902.
4. Petrov, V., D. Moltchanov, Y. Koucheryavy, and J. M. Jornet. 2020. Capacity and outage of terahertz communications with user micro-mobility and beam misalignment. *IEEE T. Veh. Technol.* 69(6):6822–6827. doi: 10.1109/ TVT.2020.2988600.
5. 3GPP. June 2021. NR; Multi-connectivity; Overall description (Release 16). Technical Specification 37.340. Available at: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3198 (accessed October 12, 2021).
6. Gapeyenko, M., V. Petrov, D. Moltchanov, M. R. Akdeniz, S. Andreev, N. Himayat, and Y. Koucheryavy. 2019. On the degree of multi-connectivityin 5G millimeter-wave cellular urban deployments. *IEEE T. Veh. Technol.* 68(2):1973–1978. doi: 10.1109/TVT.2018.2887343.
7. Begishev, V., E. Sopin, D. Moltchanov, R. Pirmagomedov, A. Samuylov, S. Andreev, Y. Koucherayvy, and K. Samouylov. 2021. Performance analysis of multi-band microwave and millimeter-wave operation in 5G NR systems. *IEEE T. Wirel. Commun.* 20(6):3475–3490. doi: 10.1109/TWC.2021.3051027.
8. Shafie, N. Y., and C. Han. 2020. Multi-connectivity for indoor terahertz communication with self and dynamic blockage. *IEEE Conference (International) on Communications Proceedings*. Piscataway, NJ: IEEE. Art. 9148716. 7 p. doi: 10.1109/ICC40277.2020.9148716.
9. Moltchanov, D. 2012. Distance distributions in random networks. *Elsevier Ad Hoc Networks* 10:1146–1166. doi: 10.1016/j.adhoc.2012.02.005.
10. Begishev, V., D. Moltchanov, E. Sopin, A. Samuylov, S. Andreev, Y. Koucheryavy, and K. Samouylov. 2019. Quantifying the impact of guard capacity on session continuity in 3GPP new radio systems. *IEEE T. Veh. Technol.* 68(12):12345–12359. doi: 10.1109/TVT.2019.2948702.
11. Boronin, P., V. Petrov, D. Moltchanov, Y. Koucheryavy, and J. M. Jornet. 2014. Capacity and throughput analysis of nanoscale machine communication through transparency windows in the terahertz band. *Nano Commun. Netw.* 5(3):72–82. doi: 10.1016/j.nancom.2014.06.001.
12. Cinlar, E. 1969. Markov renewal theory. *Adv. Appl. Probab.* 1(2):123–187. doi: 10.2307/1426216.
13. Gerasimenko, M., D. Moltchanov, M. Gapeyenko, S. Andreev, and Y. Koucheryavy. 2019. Capacity of multiconnectivity mmWave systems with dynamic blockage and directional antennas. *IEEE T. Veh. Technol.* 68(4):3534–3549. doi: 10.1109/TVT.2019.2896565.
14. Gapeyenko, M., A. Samuylov, M. Gerasimenko, D. Moltchanov, S. Singh, M. R. Akdeniz, E. Aryafar, N. Himayat, S. Andreev, and Y. Koucheryavy. 2017. On the temporal effects of mobile blockers in urban millimeter-wave cellular scenarios. *IEEE T. Veh. Technol.* 66(11):10124–10138. doi: 10.1109/TVT.2017.2754543.

Received October 12, 2021

Contributors

Beschastnyi Vitalii A. (b. 1992) — Candidate of Science (PhD) in physics and mathematics, assistant professor, Department of Applied Probability and Informatics, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; beschastnyy-va@rudn.ru

Ostrikova Daria Yu. (b. 1988) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Applied Probability and Informatics, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; ostrikova-dyu@rudn.ru

Shorgin Vsevolod S. (b. 1978) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vshorgin@ipiran.ru

Moltchanov Dmitri A. (b. 1978) — Doctor of Science in technology, associate professor, Department of Electronics and Communications Engineering, Tampere University, 7 Korkeakoulunkatu, Tampere 33720, Finland; dmitri.moltchanov@tuni.fi

Gaidamaka Yuliya V. (b. 1971) — Doctor of Science in physics and mathematics, professor, Department of Applied Probability and Informatics, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; senior scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; gaydamaka-yuv@rudn.ru

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ, 2021. Т. 31. № 4. С. 157–167

ИСПОЛЬЗОВАНИЕ ВЕБ-КРАУЛЕРОВ В ТЕХНОЛОГИИ ПОДДЕРЖКИ КОНКРЕТНО-ИСТОРИЧЕСКИХ ИССЛЕДОВАНИЙ

И. М. Адамович[1]***, О. И. Волков***[2]

Аннотация: Статья посвящена дальнейшему развитию распределенной технологии поддержки конкретно-исторических исследований (ПКИИ), основанной на принципах краудсорсинга и ориентированной на широкий круг не относящихся к профессиональным историкам и биографам пользователей. Развитие осуществляется за счет автоматизации одного из основных видов интернет-поиска (косвенный интернет-поиск (КИП)), используемого при проведении биографических исследований. Проанализированы возможные подходы к автоматизации интернет-поиска с учетом специфики конкретно-исторического исследования. Обосновано использование веб-краулеров и сформулированы требования к ним, вытекающие из особенностей рассматриваемой технологии. Оценена возможность использования готовых решений. Описаны необходимые изменения объектной модели технологии и модификации ее алгоритмов, относящихся к КИП. В качестве дополнительной меры сокращения трудозатрат при проведении КИП предложен и подробно описан механизм автоматизации взаимодействия пользователей технологии, ведущих свои исследования в близких направлениях.

Ключевые слова: конкретно-историческое исследование; распределенная технология; веб-краулер; модель данных; интернет-поиск

DOI: 10.14357/08696527210413

1 Введение

Поддержка конкретно-исторических исследований стала одной из актуальных задач современности, что обусловлено вовлечением в исследовательский процесс не только членов профессионального исторического сообщества, но и самых широких слоев непрофессионалов в связи со все возрастающим интересом к частной, семейной истории [1].

Специфика биографического исследования состоит в том, что в центре внимания исследователя находится конкретная личность и все без исключения стороны (социальные, экономические, политические, этнические, художественные и т. п.)

[1] Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Adam@amsd.com

[2] Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Volkov@amsd.com

ее реальной жизни [2]. Многообразие изучаемых аспектов жизни индивидуума, стоящего в центре исследования, приводит к огромному числу направлений поиска. Следствием этого становится необходимость использования абсолютно всех доступных источников информации. Среди них интернет, наряду с архивами и специальной литературой, оказывается одним из наиболее значимых. При этом, как было показано в [3], биограф, как правило, не может рассчитывать, что прямые вопросы, стоящие перед ним на данном этапе исследования, могут быть непосредственно преобразованы в некоторые интернет-запросы, на которые поисковые машины (ПМ) выдадут информацию, содержащую столь же прямые ответы на них. Поэтому поиск информации при биографическом исследовании идет, как правило, в двух направлениях:

(1) «косвенный» интернет-поиск, подробно рассмотренный в [3] и характеризующийся тем, что исследователь, опираясь на уже известную ему, как правило фрагментарную, противоречивую и недостоверную информацию по теме исследования, придумывает косвенные вопросы и формулирует соответствующие запросы к ПМ, постепенно заполняя информационные лакуны;

(2) «свободный» поиск (СП), заключающийся в просмотре текстов самой разной направленности, потенциально имеющих отношение к теме исследования, с целью обнаружения полезной для исследования информации, характер которой исследователь не может заранее предугадать, и, следовательно, не может отыскать ее с помощью каких-либо запросов.

В статье [4] обосновано использование систем автоматического извлечения фактов с целью упрощения задачи исследователя при проведении СП за счет возможности быстро ознакомиться с краткой «выжимкой» и по ней уже принять решение о целесообразности чтения всего документа.

В [5, 6] описана разработанная в ФИЦ ИУ РАН распределенная технология ПКИИ, основанная на принципах краудсорсинга (мобилизации ресурсов широкого круга добровольцев посредством информационных технологий) и поддерживающая СП. Данные в этой технологии организованы в форме семантической сети. Узлы сети представляют собой именованные универсальные классы объектов. Факты задаются значениями экземпляров классов и связями между ними. Связи наследуются из сети классов [5].

В основу технологии положена система Т-парсер [4], осуществляющая автоматическое, адаптированное к специфике биографического поиска извлечение фактов из текстов историко-биографической направленности и решающая задачи фактографического индексирования источников и их семантической разметки (вычленения значимых фрагментов и оснащения их метаданными).

При этом КИП в технологии ПКИИ осуществляется на базе набора правил использования стандартных ПМ и средств операционной системы без применения специальных программных средств поддержки.

2 Специфика проведения косвенного интернет-поиска в технологии поддержки конкретно-исторических исследований

Как показано в [3], при проведении КИП исследователю приходится многократно переформулировать запрос, пока результаты поиска его не удовлетворят, т. е. осуществлять итерационный поиск. Анализ результатов, найденных на предыдущем шаге итерации, уточняет направление поиска на последующем шаге. Таких уточнений на одном шаге может быть несколько, поэтому КИП представляет собой не только итерационный, но и ветвящийся процесс. На каждом шаге итерации интересующий исследователя результат с большой вероятностью будет находиться отнюдь не в первых строчках выдачи. Это объясняется тем, что современные алгоритмы ранжирования результатов выдачи учитывают авторитетность страницы, которую они вычисляют по числу и качеству ссылок на нее с других сайтов. Интерес же для исследователя часто представляет малоизвестная информация. Также отличительной чертой КИП является то, что исследователь часто во время поиска не может быть абсолютно уверен, содержится ли в полной мере интересующая его информация в уже найденных документах или еще нет, поскольку ответ на этот вопрос требует тщательного изучения этих документов. Поэтому КИП, как правило, проводится до тех пор, пока новые документы на интересующую тематику не перестанут появляться в выдаче или пока найденных документов не наберется достаточно много, т. е. поиск новых начнет приводить к неоправданному увеличению их числа до объемов, затруднительных для изучения в разумное время.

Таким образом, специфика КИП состоит в том, что он

- проводится, как правило, итерационно с ветвлениями;
- требует на каждой итерации просмотра значительного объема результатов выдачи;
- осуществляется «до упора», а не до нахождения первой удовлетворяющей исследователя ссылки.

Из этого следует очень высокая трудоемкость КИП. Но ситуация дополнительно осложняется высокой динамичностью сети Интернет, при этом динамика характеризуется устойчивым ростом с нарастанием темпа [7]: по статистике, объем цифровой информации удваивается каждые 18 месяцев. Поскольку продолжительность конкретно-исторического исследования может измеряться годами, за это время может появиться новая информация по интересующей теме. Это означает, что время и трудозатраты исследователя на проведение КИП с учетом его вынужденного многократного повторения неоправданно велики и технологию ПКИИ целесообразно дополнить средствами автоматизации этой деятельности.

3 Подход к автоматизации косвенного интернет-поиска в технологии поддержки конкретно-исторических исследований

Поскольку КИП представляет собой итерационный и ветвящийся процесс, направления и области поиска уточняются исследователем на каждом шаге и в полной мере автоматизировать его невозможно. Но по мере выявления сайтов, содержащих или могущих содержать полезную для исследования информацию, повторный поиск по ним вполне может быть автоматизирован.

Как показано в [8], весь поток данных для мониторинга можно условно разделить на три части по категориям источников поиска информации в сети Интернет:

(1) информация, получаемая через ПМ и их API (Application Programming Interface — программный интерфейс приложения), например Google, Яндекс, Yahoo и т. п.;

(2) информация, получаемая через API и RSS (Rich Site Summary — краткое описание новой информации, появившейся на сайте) — ленты площадок (ресурсов), которые предоставляют доступ к внутреннему поиску, например Twitter, Facebook, Vk и т. п.;

(3) информация, получаемая поисковыми роботами самих мониторинговых систем.

Использование RSS в КИП не представляется целесообразным, поскольку RSS служит для ознакомления с информацией и принятия решения о переходе по ссылке на полную версию сайта. Детальная информация, необходимая исследователю, в файле RSS не содержится.

Популярные ПМ (Яндекс, Google и др.) дают наиболее полную картину по поисковым запросам, т. е. полнота выдачи у них самая высокая. Однако наблюдается различие в выдаче по API, предоставляемых этими сервисами, и выдаче на самом сайте поисковой системы. Иногда в результатах выдачи по поисковому запросу через API можно вообще не увидеть до 90% выдачи, которую можно видеть в браузерной версии [8], а также существующие ПМ не всегда релевантно удовлетворяют пользовательские запросы [9]. Отсюда следует, что автоматизация КИП через обращение к ПМ через их API нецелесообразна и требуется альтернативное решение. Таким решением может стать использование поискового робота (краулера). Главное преимущество краулеров — возможность собирать любую находящуюся в открытом доступе информацию на ресурсе, в том числе и ту, что ресурс может не отдавать через API.

Дополнительным направлением автоматизации КИП может стать обмен списками выявленных за время исследования сайтов, содержащих конкретно-историческую информацию, между пользователями технологии ПКИИ, ведущими свои исследования в близких направлениях.

4 Требования к веб-краулеру для поддержки косвенного интернет-поиска

Общий список требований, которым должен отвечать инструмент сбора данных, состоит из следующих составляющих [10]:

(1) масштабируемость. Краулер должен иметь возможность увеличения производительности за счет подключения дополнительного оборудования — серверов или виртуальных машин;

(2) распределенность. Краулер должен иметь возможность работать в распределенном режиме на нескольких серверах (работа в кластере) или виртуальных машинах;

(3) модульность. Краулер должен предоставлять возможность добавления новой функциональности путем подключения дополнительных модулей, например для анализа новых форматов данных, протоколов и т. д.;

(4) производительность и эффективность. Краулер должен обеспечивать эффективное использование системных ресурсов, включая процессор, память и полосу пропускания сети;

(5) надежность. Краулер не должен переходить в бесконечный цикл или ожидать загрузки на недоступном сайте;

(6) механизм противодействия антикраулингу. На некоторых сайтах есть фильтры для защиты от сканирования, чтобы препятствовать сбору данных. Программное обеспечение должно быть в состоянии обойти эти механизмы, чтобы собирать соответствующую информацию;

(7) прозрачность. Исходный код краулера должен быть открыт, в том числе с лицензионной точки зрения. Это обеспечит безопасность и возможность доработки программного обеспечения;

(8) качество данных. Парсер должен уметь предоставлять собранную неструктурированную информацию в необходимом формате, например JSON (текстовый формат обмена данными, основанный на JavaScript), а также отделять полезный контент от спама;

(9) поддержка. Сообщество или документация должны быть сильно развиты, чтобы можно было найти нужную информацию при модификации инструмента;

(10) вежливость. Должны соблюдаться политики веб-сайтов, описанных в файле robots.txt, регулирующие частоту, с которой краулер может парсить страницы;

(11) актуальность. Краулер должен поддерживать обновление собранных данных.

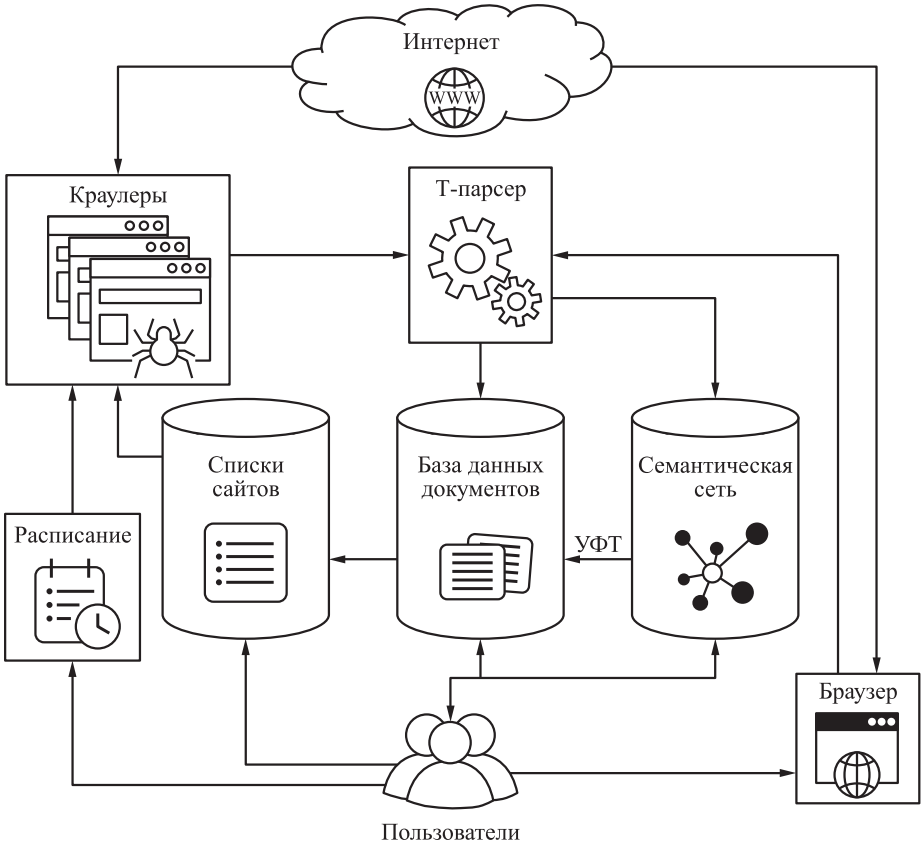
Все вышеперечисленные требования актуальны для задач КИП в рамках технологии ПКИИ. Разработка оригинального веб-краулера, учитывающего все эти требования, весьма трудоемка. Но, как показано в [11], существуют готовые решения с поддержкой основных требований и представляющие собой как краулеры с открытым кодом, так и облачные краулеры, представленные в виде сервисов, которые могут быть отнесены к следующим категориям.

1. Сфокусированный краулер (FocusedWeb Crawler) — это краулер, задача которого заключается в том, чтобы загрузить связанные с друг другом страницы по конкретной теме. Еще такой вид краулеров называют тематическим (Topic Crawler). Принцип его работы основан на том, что каждый раз, переходя к новому документу, данный краулер проверяет, насколько он релевантен обозначенной теме, переход осуществляется только на документы, соответствующие теме.
2. Инкриментный краулер (Incremental Crawler) — традиционный вид краулера, который периодически обновляет собранные в своем хранилище документы. Помимо замены старых версий документов на новые он может обновлять ранг документов, сортируя их на менее важные и более важные.
3. Распределенный краулер (Distributed Crawler) — это тип краулера, базирующегося на распределенных вычислениях. Как правило, он включает несколько вычислительных узлов, один из которых назначается мастером, а другие — дочерними узлами. Этот тип краулеров использует алгоритм Page Rank для улучшения релевантности поиска.
4. Параллельный краулер (Parallel Crawler) — такой краулер состоит из нескольких краулер-процессов, каждый из которых работает над своим выбранным множеством данных.
5. Кроссплатформенный краулер (Cross-Platform Crawler) — такой краулер должен одинаково устанавливаться и настраиваться на машинах с разными операционными системами.

Для задач КИП в рамках технологии ПКИИ необходим инкрементный параллельный краулер. Сфокусированность и распределенность желательны, но не обязательны на данном этапе развития технологии ПКИИ. Как показано в [11], существует целый ряд готовых решений, удовлетворяющих этим условиям: Nutch, Open Search Server, Norconex HTTP Collector, Bixo и др.

5 Модификация технологии поддержки конкретно-исторических исследований

Автоматизированный КИП с использованием краулеров требует внесения ряда изменений в технологию ПКИИ. Изменения должны коснуться как алгоритмов функционирования технологии, так и ее объектной модели. На рисунке



Косвенный интернет-поиск в модифицированной технологии ПКИИ

представлена схема проведения КИП как в ручном, так и в автоматическом режиме.

Объектная модель технологии [6] требует следующих модификаций.

1. Добавляется сущность «список сайтов». Атрибутами сущности служат:
   - унифицированный указатель ресурса (URL — uniform resource locator);
   - ID пользователя.
2. Добавляется сущность «расписание». Атрибутами сущности служат:
   - дата и время последней проверки;
   - межпроверочный интервал, задаваемый пользователем;
   - ID пользователя.

3. В список атрибутов сущности «документ» добавляется текстовый атрибут «источник». Для документов, найденных в интернете, значением данного атрибута будет URL соответствующего сайта.

Алгоритм поиска следующий.

1. Пользователь осуществляет ручной КИП в соответствии с [3]. Найденные документы обрабатываются системой автоматического извлечения фактов из текстов — Т-парсером [4]. В случае выявления фактов в документе он помещается в базу данных (БД) документов. Выявленные факты Т-парсер автоматически помещает в семантическую сеть с указанием их связи с документом, помещенным в БД, посредством указателя фрагмента текста (УФТ) [5]. Автоматически выявленные факты проходят процедуру ручной валидации пользователем.
2. Указатель URL значимых сайтов, содержащих документы, имеющие отношение к теме исследования, выявленных в процессе ручного КИП, сохраняются пользователем в индивидуальном списке сайтов.
3. Для пользователей, для которых в процессе их профессиональной коммуникации, поддерживаемой технологией ПКИИ, выявлено сходство тем исследований (связанные пользователи), осуществляется автоматическая рассылка изменений в индивидуальных списках сайтов в формате:
   - URL выявленного сайта;
   - перечень ссылок на выявленные документы с этого сайта, размещенные в БД документов.
4. Для каждого полученного автоматического сообщения об изменениях в списках сайтов от связанных пользователей пользователь после анализа принимает решение о добавлении того или иного URL в свой список сайтов.
5. В соответствии с расписанием, заданным пользователем, краулер автоматически осуществляет просмотр сайтов, зафиксированных в списке сайтов пользователя. Выявленные новые документы обрабатываются Т-парсером аналогично ручному режиму проведения КИП.

6 Выводы

Предложенный механизм автоматизации проведения такой важной составляющей конкретно-исторического исследования, как КИП, существенно дополняет и развивает технологию ПКИИ, ориентированную на широкий круг не относящихся к профессиональным историкам и биографам пользователей, что очень актуально в связи со все возрастающим общественным интересом к частной, семейной истории.

Обновленный алгоритм проведения КИП в полной мере отражает специфику конкретно-исторического исследования и позволяет существенно снизить время и трудозатраты на проведение поиска информации в интернете.

Описанная модификация технологии ПКИИ предполагает усовершенствование ее объектной модели за счет введения в нее новых типов списков, служащих для автоматизации работы веб-краулеров и взаимодействия пользователей при проведении исследований.

Литература

1. *Помникова А. Ю.* Семейная история в дискурсивном пространстве // Вестник Мининского университета, 2019. Т. 7. № 1(26). Ст. 9. 22 с.
2. *Иконникова С. Н.* Биографика как часть исторической культурологии // Вестник СПбГУКИ, 2012. № 2(11). С. 6–10.
3. *Адамович И. М., Волков О. И.* Средства поддержки интернет-поиска при проведении биографических исследований // Системы и средства информатики, 2014. Т. 24. № 2. С. 178–192.
4. *Адамович И. М., Волков О. И.* Система извлечения биографических фактов из текстов исторической направленности // Системы и средства информатики, 2015. Т. 25. № 3. С. 235–250.
5. *Адамович И. М., Волков О. И.* Технология распределенного автоматизированного анализа исторических текстов // Системы и средства информатики, 2016. Т. 26. № 3. С. 148–161.
6. *Адамович И. М., Волков О. И.* Единая технология поддержки конкретно-исторических исследований // Системы и средства информатики, 2019. Т. 29. № 1. С. 194–205.
7. *Минаков В. Ф.* Знания в цифровом обществе // Nauka-rastudent.ru, 2016. № 11. Ст. 21. 10 с.
8. *Кирюшин К.* Мониторинг социальных медиа — возможности и реальность // Cossa, 2013. https://www.cossa.ru/trends/43220.
9. *Андреева К. А., Шайдуров Р. С.* Разработка персональной документальной информационно-поисковой системы для сети Интернет // Решетневские чтения, 2014. Т. 2. С. 223–224.
10. *Гудков К. В., Тонкушин М. В.* Анализ автоматизированных систем сбора информации в сети Интернет // Современные информационные технологии, 2018. № 28. С. 27–31.
11. *Пудикова Е. М.* Обзор веб-краулеров для решения задачи сбора данных о представительских сайтах заданной предметной области // Системный анализ, 2016. Т. 20. С. 1–16.

Поступила в редакцию 18.05.21

THE USE OF WEB-CRAWLERS IN TECHNOLOGY OF CONCRETE HISTORICAL INVESTIGATION SUPPORT

I. M. Adamovich and O. I. Volkov

Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The article is devoted to the further development of the distributed technology of concrete historical investigation support based on the principles of crowdsourcing and focused on a wide range of users which are nonprofessional historians and biographers. Development is carried out through the automation of one of the main types of Internet searches (indirect Internet search) used in biographical research. The article analyzes the possible approaches to the automation of Internet search taking into account the specifics of concrete historical investigation. The use of web-crawlers is substantiated and the requirements for them arising from the distinctive of this technology are formulated. The possibility of using ready-made solutions is estimated. The necessary changes in the object model of the technology and the modifications of its algorithms related to indirect Internet search are described. As an additional measure to reduce the difficulty of indirect Internet search, the new mechanism for automating of the interaction of the technology users which execute their investigations in similar directions is proposed and described in detail.

Keywords: concrete historical investigation; distributed technology; web-crawler; data model; Internet search

DOI: 10.14357/08696527210413

References

1. Pomnikova, A. Yu. 2019. Semeynaya istoriya v diskursivnom prostranstve [Family stories in different types of discourse]. *Vestnik of Minin University* 7(1):9. 22 p.
2. Ikonnikova, S. N. 2012. Biografika kak chast' istoricheskoy kul'turologii [Biografical studies as part of the historical cultural studies]. *Vestnik SPbGUKI* [Bull. Saint Petersburg State University of Culture and Art] 2(11):6–10.
3. Adamovich, I. M., and O. I. Volkov. 2014. Sredstva podderzhki internet-poiska pri provedenii biograficheskikh issledovaniy [The technology of Internet search as a part of biographic investigation]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 24(2):178–192.
4. Adamovich, I. M., and O. I. Volkov. 2015. Sistema izvlecheniya biograficheskikh faktov iz tekstov istoricheskoy napravlennosti [The system of facts extraction from historical texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(3):235–250.
5. Adamovich, I. M., and O. I. Volkov. 2016. Tekhnologiya raspredelennogo avtomatizirovannogo analiza istoricheskikh tekstov [The distributed automated technology of

historical texts analysis]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 26(3):148–161.

6. Adamovich, I. M., and O. I. Volkov. 2019. Edinaya tekhnologiya podderzhki konkretno-istoricheskikh issledovaniy [Unified technology of concrete historical research support]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(1):194–205.
7. Minakov, V. F. 2016. Znaniya v tsifrovom obshchestve [Knowledge in digital society]. *Nauka-rastudent.ru* 11:21. 10 p.
8. Kiryushin, K. 2013. Monitoring sotsial'nykh media — vozmozhnosti i real'nost' [Social media monitoring — opportunities and reality]. *Cossa*. Available at: https://www.cossa.ru/trends/43220 (accessed September 12, 2021).
9. Andreeva, K. A., and R. S. Shaydurov. 2014. Razrabotka personal'noy dokumental'noy informatsionno-poiskovoy sistemy dlya seti Internet [Development of personalized documentary information retrieval system for the Internet]. *Reshetnevskie chteniya* [Reshetnev Readings] 2:223–224.
10. Gudkov, K. V., and M. V. Tonkushin. 2018. Analiz avtomatizirovannykh sistem sbora informatsii v seti Internet [Analysis of automated systems for collecting information on the Internet]. *Sovremennye informatsionnye tekhnologii* [Contemporary Information Technologies] 28:27–31.
11. Pudikova, E. M. 2016. Obzor veb-kraulerov dlya resheniya zadachi sbora dannykh o predstavitel'skikh saytakh zadannoy predmetnoy oblasti [A review of web crawlers for solving the problem of collecting data on representative sites of a given subject area]. *Sistemnyy analiz* [System Analysis] 20:1–16.

Received May 18, 2021

Contributors

Adamovich Igor M. (b. 1934) — Candidate of Science (PhD) in technology, leading scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; Adam@amsd.com

Volkov Oleg I. (b. 1964) — leading programmer, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; Volkov@amsd.com

ВЛАСТНО-КООРДИНАЦИОННЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

В. Д. Ильин[1]

Аннотация: Определены властно-координационные статусные системы и обоснована зависимость их эффективности от совершенства применяемых информационных технологий. Обоснована необходимость планового изменения устройства таких систем по мере совершенствования методов и средств реализации информационных технологий. Приведены примеры, демонстрирующие связь уровня эффективности управления с уровнем совершенства программно-аппаратно реализованного организационного и информационного обеспечения системы государственного управления. Особое внимание уделено информационной обратной связи и оповещению населения о государственных законах и их реализации.

Ключевые слова: властно-координационные системы; отношения подчинения и координации; информационные технологии; механизм государственного управления; цифровые платформы

DOI: 10.14357/08696527210414

1 Введение

В наши дни цифровизация различных видов деятельности (хозяйственной, образовательной и пр.), алгоритмизирующая имущественные, административные и другие виды отношений, позволяет совершенствовать комплексирование организационных систем, взаимодействие (внутри систем и с их окружением), разделение труда, информированность, скорость и качество решения целевых задач. Современная цифровизация, как правило, осуществляется на основе цифровых платформ (*англ.* Digital Platforms) [1, 2].

Цифровая платформа (*цп*) в общем случае представляет собой совокупность программно-аппаратно реализованных методов и средств онлайн-взаимодействия ее пользователей (цп-пользователей), решения целевых задач (информационного обеспечения, планирования, управления и др.) и онлайн-документирования результатов деятельности цп-пользователей. Освоение и применение инструментального цп-арсенала (датчиков состояний, цифровых двойников, облачных сервисов, интернета вещей и др.) становится привычной необходимостью для конкурентоспособных компаний.

Среди крупномасштабных организационных систем наиболее важные (экономические, государственного управления и др.) относятся к *властно-координационным системам статусного соперничества* (*ст-системам*) [3]. Задачи

[1]Федеральный исследовательский центр «Информатика и управление» Российской академии наук, vdilyin@yandex.ru

формирования, удостоверения и реализации информационных управляющих воздействий в ст-системах постоянно актуальны.

Запись формул и выделение фрагментов текста. Для выделения определений, замечаний и примеров используются средства языка TSM-комплекса (TSM: textual symbolic modeling), разработанного для формализованного описания текстовых s-моделей сообщений[1].

В статье применены следующие средства выделения фрагментов текста:

□ ⟨фрагмент описания⟩ □ ≈ утверждение (определение, аксиома и др.) (здесь и далее символ ≈ заменяет слово «означает»);

◊ ⟨фрагмент описания⟩ ◊ ≈ замечание.

◯ ⟨фрагмент описания⟩ ◯ ≈ пример.

Курсивом выделены первые вхождения названий понятий и фрагменты описания, к которым автор хочет привлечь внимание.

Обсуждаемые результаты. В статье представлена часть результатов *методологического обеспечения технологий цифровизации организационных систем*. Результаты получены при выполнении научно-исследовательской работы «Моделирование социальных, экономических и экологических процессов» (№ 0063-2016-0005), выполняемой в соответствии с государственным заданием ФАНО России для Федерального исследовательского центра «Информатика и управление» РАН.

◊ Материал статьи адресован исследователям, занимающимся методологическим обеспечением разработок *информационных технологий цифровой экономики* [4–8]. ◊

2 Властно-координационные статусные системы

□ *Статусное соперничество* (статусная конкуренция) — отношение между отдельными людьми (физическими лицами) или организациями (юридическими лицами), существующее везде, где стремление одних сохранить или повысить свой статус сталкивается с противодействием других. □

Статус — положение физического или юридического лица в некоторой системе экономического, межгосударственного или другого соперничества.

◊ Чаще всего статус рассматривается как положение, определяющее властные, координационные и ресурсные возможности. Обычно с повышением статуса расширяется набор располагаемых воздействий на участников системы, включая *информационные воздействия*. ◊

□ *Система статусного соперничества* (*ст-система*) — совокупность участников статусного соперничества (ст-участников) и средств, необходимых для решения целевых задач ст-системы и реализации установленных правил *подчинения* и *координации* [3]. □

[1] *Ильин В Д.* Символьное моделирование // Большая российская энциклопедия — электронная версия. http://dev.bigenc.ru/technology_and_technique/text/4010980.

Ст-системы имеют разные масштабы и цели, в них по-разному истолковывается понятие статуса, действуют разные *обязательные* и *ориентирующие правила* статусного соперничества (ст-правила), механизмы их разработки, согласования и реализации. Один и тот же человек или организация обычно бывает участником нескольких ст-систем, в которых тип конкурентных отношений определяется типом статусов, являющихся предметом соперничества.

◯ Физические лица чаще всего стремятся повысить имущественный и квалификационный статусы. Эти же типы статусов занимают первые места и в устремлениях юридических лиц. ◯

Наиболее распространенный тип конкурентных отношений реализуется в экономической деятельности. Участники экономической деятельности из разных стран входят в *сложную многоуровневую глобальную систему экономического статусного соперничества*.

◊ Статус, дающий право определять ст-правила, формировать механизмы управления ст-системами, считается одним из наиболее привлекательных. К нему стремятся наиболее активные участники ст-систем. ◊

◯ Определять правила статусного экономического соперничества на глобальном уровне пытаются посредством ВТО, МВФ и других учреждений. На уровне стран этим занимаются парламенты и правительства. ◯

□ *Властно-координационной* будем называть ст-систему, в которой на множестве ст-участников заданы отношения *подчинения* и *координации*. □

Потребность в приращении конкурентоспособности заставляет соперничающие ст-системы совершенствовать средства осуществления отношений подчинения и координации.

Документирование статусных состояний ст-участников — одна из важных функций *цифровых платформ* [1, 2] ст-систем.

◊ Среди различных статусных документов (удостоверяющих гражданство, образовательный статус и др.) важную позицию занимают документы, удостоверяющие имущественный статус, а среди них — удостоверяющие денежный имущественный статус. ◊

3 Информационные технологии в ст-системах

Издавна информационные воздействия служат важным средством управления в ст-системах различного назначения (включая систему государственного управления). В наши дни вчерашние средства информационных управляющих воздействий вытесняются электронной почтой, мессенджерами, видеоконференц-связью, интернет-порталами и др., ставшими привычно необходимыми в цифровых платформах современных ст-систем.

Растущее внимание уделяется обратной связи с объектом.

◯ Власть и в прежние времена интересовалась «откликом объекта» на государственные управляющие воздействия. К примеру, российский император

Павел I повелел в одном из окон Зимнего дворца сделать специальное окошко, в которое бы его подданные могли свободно опускать свои письменные сообщения. ◯

Информационные управляющие воздействия — важнейшее средство построения и работы *механизма государственного управления* (важнейшей жизнеустроительной ст-системы).

Однако переход в эру *глобального информационного пространства* с очень большим запаздыванием влияет на устройство механизма государственного управления, хотя принципы его устройства должны пересматриваться по мере совершенствования научно-технических средств информационного взаимодействия, утвердившихся своей практической целесообразностью и эффективностью.

◊ Напомним, что важнейшие системы жизнеобеспечения, представленные в [3] *коммуникационным* и *информационным госпространством*, целесообразно изменяли принципы своего устройства по мере совершенствования опорных для них научно-технических средств. ◊

Почему же механизм государственного управления так отстает в совершенствовании от своих опорных научно-технических средств? Как жизнеустроительная ст-система государственный механизм имеет много общего с коммуникационным и информационным механизмами жизнеобеспечения. Как и они, механизм государственного управления имеет отношение к жизни всего населения страны. И, что наиболее важно, его повседневное функционирование в основном опирается на средства *информационного* и *коммуникационного госпространств* [3].

◯ Чтобы представить, насколько работа государственного механизма зависит от совершенства информационных средств осуществления его функций, достаточно вспомнить пример из недавней истории. Отсутствие средств электронного голосования в зале заседаний I Съезда народных депутатов СССР практически блокировало возможность нормальной работы (на первом заседании). Избирали специальных «счетчиков», и те считали поднятые руки в тысячном зале. ◯

Осуществление в государственном механизме регулирующих процессов типа *координации*, опирающейся на технологические принципы телеконференций, перестало быть чем-то редким. В наши дни несложно реализуется композиционная технология, включающая средства дистанционной идентификации ст-участников, телеконференции, телеголосования, документирования и архивирования результатов работы.

Не исключено, что в ближайшее время ряд привычных «постоянно действующих заседаний» (типа парламентских) перейдут в щадящий режим выборочных заседаний. Основную же часть деятельности станут выполнять в дистанционном режиме.

В свое время радиовещание и телевидение принципиально изменили арсенал средств информационного взаимодействия с населением страны. Однако до сих пор нет радио- и телепрограмм, предназначенных для формирования *правового пространства страны*. В рамках таких программ целесообразно пояснять

тексты законов и важных государственных нормативных актов. Не менее важно сообщать о механизмах их реализации.

Судебная составляющая государственного механизма существенно укрепилась бы в своем влиянии на правовое сознание населения, если бы работа судов показывалась по специальным региональным телеканалам.

Таким путем для составляющих государственного механизма сформировалась бы система взаимодействия с населением. И действовала бы она без посредников-толкователей, нередко искажающих своими произвольными добавками государственные сообщения, адресованные населению.

◊ Ведь речь идет о государственном информационном взаимодействии. Значит, посредники должны играть только техническую роль. Если посредством массовых средств информационного взаимодействия передан текст государственного закона, объяснено его значение и определен механизм реализации, создано правовое основание требовать выполнения закона.

Реализовано ли такое основание? «Незнание закона не освобождает от ответственности за его нарушение» — строгое и нередко повторяемое высказывание. А достаточно ли сделано для того, чтобы осуществилась возможность незнание закона заменить его знанием? ◊

На смену былым средствам обратной связи приходят современные (в частности, предоставляемые региональными и федеральными информационными порталами). Пока в их нынешнем виде они не везде и не всегда пригодны для применения в качестве важной информационной составляющей государственного механизма. Нужны разные типы обратной связи: по срочности, конфиденциальности, уровню защиты, документируемости и архивируемости сообщений.

Особого внимания заслуживают информационные средства осуществления процессов государственного управления в неординарных ситуациях. Оперативность, защищенность, высокая достоверность и документируемость передаваемых сообщений (содержанием которых служат принятые государственные решения и сообщения о ходе их исполнения) — все это специфика работы государственной власти в неординарных условиях.

◊ Немало случаев, когда утечка сообщений о готовящемся секретном решении происходила через так называемых «технических работников» (тех, кто печатал текст или доставлял проект документа из одного места в другое или обязан был обеспечить надежное хранение доставленного проекта документа). А ведь сегодня существуют технологические возможности резко уменьшить вероятность таких утечек. ◊

Сказано об этом для того, чтобы свести воедино две вещи, которые одна без другой теряют в полезности для усовершенствования существующего государственного механизма. Первая из них — научно обоснованная властно-координационная структура государственного механизма со спецификациями развертки управляющих воздействий и свертки информационных сообщений об отклике объекта. Вторая — методология построения и развития *цифровой платформы механизма государственного управления* (с утвержденным организаци-

онным и правовым обеспечением, правилами информационного взаимодействия и др. [3]).

4 Заключение

Предложены определения *властно-координационных систем статусного соперничества*, наиболее важными из которых являются системы государственного управления и экономические системы.

Обоснована необходимость изменения устройства таких систем (на основе *цифровых платформ*) по мере совершенствования методов и средств информационного взаимодействия и реализации информационных управляющих воздействий.

Литература

1. *Tiwana A., Konsynski B., Bush A.* Platform evolution: Coevolution of architecture, governance, and environmental dynamics // Inform. Syst. Res., 2010. Vol. 21. No. 4. P. 675–687.
2. *Parker G. G., Van Alstyne M. W., Choudary S. P.* Platform revolution: How networked markets are transforming the economy and how to make them work for you. — New York, NY, USA: W. W. Norton & Co., 2016. 352 p.
3. *Ильин В. Д.* Основания ситуационной информатизации. — М.: Наука, Физматлит, 1996. 180 с.
4. *Tapscott D.* The digital economy: Promise and peril in the age of networked intelligence. — New York, NY, USA: McGraw-Hill, 1996. 342 p.
5. *Christensen C. M.* The innovator's dilemma: When new technologies cause great firms to fail. — Boston, MA, USA: Harvard Business School Press, 1997. 288 p.
6. The new digital economy: How it will transform business. — Oxford Economics, 2015. 34 p.
7. G20 digital economy development and cooperation initiative // G20 Summit, 2016. http://en.kremlin.ru/supplement/5111.
8. Цифровая экономика Российской Федерации: Программа, утвержденная Распоряжением Правительства Российской Федерации от 28 июля 2017 г. № 1632-р. http://d-russia.ru/wpcontent/uploads/2017/07/programma-tsifrov-econ.pdf

Поступила в редакцию 18.05.21

POWER-COORDINATION SYSTEMS AND INFORMATION TECHNOLOGIES

V. D. Ilyin

Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: Power-coordination status systems have been defined and the dependence of their effectiveness on the excellence of the applied information technologies has been substantiated. The necessity of a planned change in the structure of such systems is substantiated as the methods and means of implementing information technologies are improved. Examples are given that demonstrate the relationship between the level of management efficiency and the level of perfection of the software and hardware implementing organizational and information support of the public administration system. Special attention is paid to informational feedback and notification of the population about state laws and their implementation.

Keywords: power-coordination systems; relations of subordination and coordination; information technology; mechanism of state administration; digital platforms

DOI: 10.14357/08696527210414

References

1. Tiwana, A., B. Konsynski, and A. Bush. 2010. Platform evolution: Coevolution of architecture, governance, and environmental dynamics. *Inform. Syst. Res.* 21(4):675–687.
2. Parker, G. G., M. W. Van Alstyne, and S. P Choudary. 2016. *Platform revolution: How networked markets are transforming the economy and how to make them work for you.* New York, NY: W. W. Norton & Co. 352 p.
3. Ilyin, V. D. 1996. *Osnovaniya situatsionnoy informatizatsii* [Fundamentals of situational informatization]. Moscow: Nauka, Fizmatlit. 180 p.
4. Tapscott, D. 1996. *The digital economy: Promise and peril in the age of networked intelligence.* New York, NY: McGraw-Hill. 342 p.
5. Christensen, C. M. 1997. *The innovator's dilemma: When new technologies cause great firms to fail.* Boston, MA: Harvard Business School Press. 288 p.
6. The new digital economy: How it will transform business. 2015. Oxford Economics. 34 p. Available at: http://www.pwc.com/mt/en/publications/assets/thenewdigital economy.pdf (accessed October 1, 2021).
7. G20 digital economy development and cooperation initiative. 2016. G20 Summit. Available at: http://en.kremlin.ru/supplement/5111 (accessed October 1, 2021).
8. Tsifrovaya ekonomika Rossiyskoy Federatsii: Programma, utverzhdennaya Rasporyazheniem Pravitel'stva Rossiyskoy Federatsii [Digital economy of the Russian Federation:

Program Approved by Order No. 1632-r dated July 28, 2017 of the Government of the Russian Federation]. Available at: http://d-russia.ru/wpcontent/uploads/2017/07/programma-tsifrov-econ.pdf (accessed October 1, 2021).

Received May 18, 2021

Contributor

Ilyin Vladimir D. (b. 1937) — Doctor of Science in technology, professor, leading scientist, A. A. Dorodnicyn Computing Center, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 40 Vavilov Str., Moscow119333, Russian Federation; vdilyin@yandex.ru

ОБ АВТОРАХ

Адамович Игорь Михайлович (р. 1934) — кандидат технических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Базилевский Михаил Павлович (р. 1987) — кандидат технических наук, доцент Иркутского государственного университета путей сообщения

Бесчастный Виталий Александрович (р. 1992) — кандидат физико-математических наук, ассистент кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

Борисов Андрей Владимирович (р. 1965) — доктор физико-математических наук, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Босов Алексей Вячеславович (р. 1969) — доктор технических наук, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Волков Олег Игоревич (р. 1964) — ведущий программист Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Гайдамака Юлия Васильевна (р. 1971) — доктор физико-математических наук, профессор кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Грушо Александр Александрович (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Дьяченко Юрий Георгиевич (р. 1958) — кандидат технических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Жуков Денис Владимирович (р. 1979) — главный специалист Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зацаринный Александр Алексеевич (р. 1951) — доктор технических наук, профессор, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Иванов Алексей Владимирович (р. 1976) — кандидат технических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Ильин Владимир Дмитриевич (р. 1937) — доктор технических наук, профессор, ведущий научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

Ионенков Юрий Сергеевич (р. 1956) — кандидат технических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кириков Игорь Александрович (р. 1955) — кандидат технических наук, директор Калининградского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

Косолапов Юрий Владимирович (р. 1982) — кандидат технических наук, доцент Института математики, механики и компьютерных наук им. И. И. Воровича Южного федерального университета

Кривенко Михаил Петрович (р. 1946) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Молдовян Александр Андрееевич (р. 1951) — доктор технических наук, профессор, главный научный сотрудник лаборатории кибербезопасности и постквантовых криптосистем Санкт-Петербургского института информатики и автоматизации РАН Санкт-Петербургского Федерального исследовательского центра РАН

Молдовян Дмитрий Николаевич (р. 1986) — кандидат технических наук, научный сотрудник лаборатории кибербезопасности и постквантовых криптосистем Санкт-Петербургского института информатики и автоматизации РАН Санкт-Петербургского Федерального исследовательского центра РАН

Молдовян Николай Андрееевич (р. 1953) — доктор технических наук, профессор, главный научный сотрудник лаборатории кибербезопасности и постквантовых криптосистем Санкт-Петербургского института информатики и автоматизации РАН Санкт-Петербургского Федерального исследовательского центра РАН

Молчанов Дмитрий Александрович (р. 1978) — доктор физико-математических наук, доцент Департамента электроники и телекоммуникаций Университета Тампере, Тампере, Финляндия

Орлов Георгий Александрович (р. 1994) — инженер-исследователь Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Острикова Дарья Юрьевна (р. 1988) — кандидат физико-математических наук, доцент кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

Румовская София Борисовна (р. 1985) — кандидат технических наук, научный сотрудник Калининградского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

Смирнов Дмитрий Владимирович (р. 1984) — бизнес-партнер по ИТ Департамента безопасности ПАО «Сбербанк России»

Степченков Юрий Афанасьевич (р. 1951) — кандидат технических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Тимонина Елена Евгеньевна (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Хилько Дмитрий Владимирович (р. 1987) — старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шикунов Юрий Игоревич (р. 1995) — инженер-исследователь Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шоргин Всеволод Сергеевич (р. 1978) — кандидат технических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шоргин Сергей Яковлевич (р. 1952) — доктор физико-математических наук, профессор, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

АВТОРСКИЙ УКАЗАТЕЛЬ ЗА 2021 г.

2021 AUTHOR INDEX

Правила подготовки рукописей статей для публикации в журнале «Системы и средства информатики»

Журнал «Системы и средства информатики» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информационных технологий.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- информационно-телекоммуникационные системы и средства их построения;
- архитектура и программное обеспечение вычислительных машин, комплексов и сетей;
- методы и средства защиты информации.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

   Публикация предоставленной автором(ами) рукописи не должна нарушать положений глав 69, 70 раздела VII части IV Гражданского кодекса, которые определяют права на результаты интеллектуальной деятельности и средства индивидуализации, в том числе авторские права, в РФ.

   Ответственность за нарушение авторских прав, в случае предъявления претензий к редакции журнала, несут авторы статей.

   Направляя рукопись в редакцию, авторы сохраняют свои права на данную рукопись и при этом передают учредителям и редколлегии журнала неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на перевод ее на английский язык, а также на ее распространение в России и за рубежом. Каждый автор должен представить в редакцию подписанный с его стороны «Лицензионный договор о передаче неисключительных прав на использование произведения», текст которого размещен по адресу http://www.ipiran.ru/publications/licence.doc. Этот договор может быть представлен в бумажном (в 2-х экз.) или в электронном виде (отсканированная копия заполненного и подписанного документа).

   Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации представленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.

3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

   Возвращение рукописи на доработку не означает ее принятия к печати.

   Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам.

   Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.

5. Редактура статей высылается авторам для просмотра. Замечания к редактуре должны быть присланы авторами в кратчайшие сроки.
6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или LaTeX (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.
7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

   Параметры страницы: формат — A4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху и снизу — 2, от края до нижнего колонтитула — 1,3.

   Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

   Рекомендуемый объем рукописи — не свыше 10 страниц указанного формата. При превышении указанного объема редколлегия вправе потребовать от автора сокращения объема рукописи.

   Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

   Все страницы рукописи нумеруются.

   Шаблоны примеров оформления представлены в Интернете:

   http://www.ipiran.ru/publications/collected/template.doc
8. Статья должна содержать следующую информацию на ***русском и английском языках***:

   – название статьи;
   – Ф.И.О. авторов, на английском можно только имя и фамилию;
   – место работы, с указанием города и страны и электронного адреса каждого автора;
   – сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:
     http://www.ipiran.ru/journal/collected/2019_29_03_rus/authors.asp и
     http://www.ipiran.ru/journal/collected/2019_29_03_eng/authors.asp;
   – аннотация (не менее 100 слов на каждом из языков). Аннотация — это краткое резюме работы, которое может публиковаться отдельно. Она является основным источником информации в информационных системах и базах данных. Английская аннотация должна быть оригинальной, может не быть дословным переводом русского текста и должна быть написана хорошим английским языком. В аннотации не должно быть ссылок на литературу и, по возможности, формул;
   – ключевые слова — желательно из принятых в мировой научно-технической литературе тематических тезаурусов. Предложения не могут быть ключевыми словами.
   – источники финансирования работы (ссылкм на гранты, проекты, поддерживающие организации и т. п.

9. Требования к спискам литературы.
   Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в каждом из списков литературы в порядке первых упоминаний.

Списки литературы представляются в двух вариантах:

(1) **Список литературы к русскоязычной части**. Русские и английские работы — на языке и в алфавите оригинала.

(2) **References**. Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Необходимо для составления списка "References" пользоваться размещенной на сайте http://www.translit.net/ru/bgn/ бесплатной программой транслитерации русского текста в латиницу.

Список литературы "References" приводится полностью отдельным блоком, повторяя все позиции из списка литературы к русскоязычной части, независимо от того, имеются или нет в нем иностранные источники. Если в списке литературы к русскоязычной части есть ссылки на иностранные публикации, набранные латиницей, они полностью повторяются в списке "References".

Примеры ссылок на различные виды публикаций в списке "References":

Описание статьи из журнала:

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Russ. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Описание статьи из электронного журнала:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: http://www.ascusc.org/jcmc/vol5/issue2/ (accessed April 28, 2011).

Описание материалов конференций:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursosberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Описание книги (монографии, сборники):

Lindorf, L. S., and L. G. Mamikoniants, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Описание переводной книги (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:
   Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (*Timoshenko S. P., Young D. H., Weaver W.* Vibration problems in engineering. — 4th ed. — New York, NY, USA: Wiley, 1974. 521 p.)

2. В англоязычной части:
   Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. New York, NY: Wiley. 521 p.

Описание неопубликованного документа:

Latypov, A. R., M. M. Khasanov, and V. A. Baikov. 2004. Geology and production (NGT GiD). Certificate on official registration of the computer program No. 2004611198. (In Russian, unpubl.)

Описание интернет-ресурса:

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: http://www.scribd.com/doc/1034528/ (accessed February 7, 2011).

Описание диссертации или автореферата диссертации:

Semenov, V. I. 2003. Matematicheskoe modelirovanie plazmy v sisteme kompaktnyy tor [Mathematical modeling of the plasma in the compact torus]. Moscow. D.Sc. Diss. 272 p.

Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

Описание ГОСТа:

GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publs. 10 p.

Описание патента:

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.
11. При отправке файлов по электронной почте просим придерживаться следующих правил:
    - указывать в поле subject (тема) название журнала и фамилию автора;
    - использовать attach (присоединение);
    - в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.
12. Журнал «Системы и средства информатики» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Системы и средства информатики»:

Москва 119333, ул. Вавилова, д. 44, корп. 2, ФИЦ ИУ РАН
Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05
e-mail: ssi@frccsc.ru (Стригина Светлана Николаевна)
http://www.ipiran.ru/journal/collected

Requirements for manuscripts submitted to Journal "Systems and Means of Informatics"

Journal "Systems and Means of Informatics" publishes theoretical, review, and discussion articles on the research and development in the field of information technology.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

Topics covered include the following areas:

- information and communication systems and tools of their design;
- architecture and software of computational complexes and networks; and
- methods and tools of information protection.

1. The Journal publishes original articles which have not been published before and are not intended for simultaneous publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

   Agreement on the transfer of rights to publish:

   "We, the undersigned authors of the manuscript ". . . ," pass to the Founder and the Editorial Board of the Journal "Systems and Means of Informatics" the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.

   Author(s) signature(s): (name(s), address(es), date)."

   This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

   The Editorial Board has the right to request from the authors an official expert conclusion that the submitted article has no classified data prohibited for publication.
2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If authors receive their article for correction after reviewing, it does not mean that the article is approved to be published. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or LaTeX (.tex) formats, and additionally in the .pdf format. All documents

may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.

7. The recommended typesetting instructions for manuscript.

   Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

   Text: font —Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

   The recommended manuscript size: not more than 10 pages of the specified format. If the specified size exceeded, the editorial board is entitled to require the author to reduce the manuscript.

   Use only standard abbreviations. Avoid abbreviations in the title and abstract. The full term for which an abbreviation stands should precede its first use in the text unless it is a standard unit of measurement.

   All pages of the manuscript should be numbered.

   The templates for the manuscript typesetting are presented on site:

   http://www.ipiran.ru/publication/collected/template.doc

8. Articles should enclose data both in **Russian and English**:

   - title;
   - author's name and surname;
   - affiliation — organization, its address with ZIP code, city, country, and official e-mail address;
   - data on authors according to the format (see site):
     http://www.ipiran.ru/journal/collected/2019_29_03_rus/authors.asp and
     http://www.ipiran.ru/journal/collected/2019_29_03_eng/authors.asp;
   - abstract (not less than 100 words) both in Russian and in English. Abstract is a short summary of the article that can be published separately. The abstract is the main source of information on the article and it could be included in leading information systems and data bases. The abstract in English has to be an original text and should not be an exact translation of the Russian one. Good English is required. In abstracts, avoid references and formulae.
   - Indexing is performed on the basis of keywords. The use of keywords from the internationally accepted thematic Thesauri is recommended.

     Important! Keywords must not be sentences.
   - Acknowledgments.

9. References. Russian references have to be presented both in English translation and in Latin transliteration (refer http://www.translit.net/ru/bgn/).

   Please take into account the following examples of Russian references appearance:

   Article in journal:

   Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Russ. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

   Journal article in electronic format:

   Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic

exchange. *JCMC* 5(2). Available at: http://www.ascusc.org/jcmc/vol5/issue2/ (accessed April 28, 2011).

Conference proceedings:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursosberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Books and other monographs:

Lindorf, L. S., and L. G. Mamikoniants, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Dissertation and Thesis:

Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. Moscow: IPI RAN. PhD Thesis. 23 p.

State standards and patents:

GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform Publs. 10 p.

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

References in Latin transcription are presented in the original language.

References in the text are numbered according to the order of their first appearance; the number is placed in square brackets. All items from the reference list should be cited.

10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.
11. Submissions of files by e-mail must include:
    - the journal title and author's name in the "Subject" field;
    - an article and additional materials have to be attached using the "attach" function;
    - an electronic version of the article should contain the file with the text and a separate file with figures.
12. "System and Means of Informatics" journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

Editorial Board address:

FRC CSC RAS, 44, block 2, Vavilov Str., Moscow 119333, Russia

Ph.: +7 (499) 135 86 92, Fax: +7 (495) 930 45 05

e-mail: ssi@frccsc.ru (to Svetlana Strigina)

http://www.ipiran.ru/english/journal_systems.asp

SYSTEMS AND MEANS OF INFORMATICS (SISTEMY I SREDSTVA INFORMATIKI)

SCIENTIFIC JOURNAL

Volume 31 No.4 Year 2021

I N T H I S I S S U E: