На правах рукописи

Абрамов Николай Александрович

ПРОГРАММНАЯ СИСТЕМА ВЫЯВЛЕНИЯ НЕЛЕГИТИМНОЙ АКТИВНОСТИ НА ПРОМЫШЛЕННЫХ ПЛОЩАДКАХ

Специальность 05.13.11 – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей.

Автореферат диссертации на соискание ученой степени кандидата технических наук

Москва

2013

Работа выполнена в Федеральном государственном бюджетном Учреждении науки Вычислительный центр им. А.А. Дородницына Российской Академии Наук

Научный руководитель: доктор технических наук, профессор
Дулин Сергей Константинович

Официальные оппоненты:
Цурков Владимир Иванович доктор физико-математических наук, профессор, заведующий отделом ВЦ РАН

Бецков Александр Викторович доктор технических наук, доцент кафедры управления деятельностью служб обеспечения общественного порядка Академии управления МВД России

Ведущая организация: Федеральное государственное бюджетное Учреждение науки Институт системного анализа Российской Академии Наук

Защита состоится 12 декабря 2013 года в 13 часов на заседании диссертационного совета Д 002.017.02 при Федеральном государственном бюджетном Учреждении науки Вычислительный центр им. А.А. Дородницына Российской Академии Наук по адресу: 119333, Москва, ул. Вавилова, дом 40, конференц-зал.

С диссертацией можно ознакомиться в библиотеке ВЦ РАН.

Автореферат разослан 10 ноября 2013 года

Ученый секретарь диссертационного совета
Д 002.017.02, д.ф.-м.н., профессор В.В. Рязанов

Общая характеристика работы

Актуальность темы. В настоящее время появились новые информационные технологии, связанные с бихевиористическим анализом деятельности различных объектов. Эти технологии требуют сложного математического и программного обеспечения. Одной из областей бихевиористического анализа является разработка математических моделей и соответствующего программного обеспечения для компьютеризированных комплексов, предназначенных для автоматизированного анализа поведения различных объектов в заданных средах.

Исключительно большой практический интерес в бихевиористическом анализе имеет задача обнаружения нелегитимной активности на промышленных объектах. В настоящее время уровень математического и программного обеспечения такого анализа недостаточен для успешного решения практических задач в данной области. Фактически, системы, используемых для выявления нелегитимной активности, не имеют математической и программной составляющих, и сводятся к неавтоматизированному видеоконтролю территорий предприятий. В связи с этим разработка математического и программного обеспечения указанных бихевиористических технологий является чрезвычайно актуальной.

Цель данного диссертационного исследования – разработка математического и программного обеспечения информационной технологии для бихевиористического анализа потоков транспорта на предприятии в режиме реального времени.

Для достижения поставленной цели диссертационного исследования были решены следующие задачи:

- Разработана математическая модель классификации автотранспортных средств, позволяющая на основе анализа потока видеоданных выявлять транспортные средства с признаками нелегитимной активности;
- Разработана архитектура программной системы бихевиористического анализа;
- Разработан программный модуль, реализующий интеграцию критериев в информационную систему;

- На базе разработанного математического обеспечения создан программно-аппаратный комплекс, реализующий систему анализа бихевиористического поведения.

Методы исследования. Теоретические и практические исследования базируются на методах распознавания образов (алгоритмы классификации, распознавание текстовых меток), системного программирования, методах построения вычислительных систем и математических методах моделирования нелегитимной активности.

Научная новизна. В диссертационной работе разработано новое математическое и программное обеспечение системы компьютеризированного выявления нелегитимной активности. Данная система основана на анализе признаков нелегитимной активности математическими методами.

Автором получены следующие результаты:

- Разработана математическая и программная модель обнаружения нелегитимной активности автотранспорта;
- На их основе разработано прикладное программное обеспечение, которое в комбинации с программным модулем распознавания автомобильных номеров без участия оператора с высокой вероятностью выявляет нелегитимную активность в режиме реального времени;
- Разработана архитектура программного комплекса для бихевиористического анализа данных видеонаблюдения;
- На основе анализа программных средств выявления нелегитимной активности разработаны требования к техническим параметрам проектируемой системы;
- Создано программно-прикладное средство управления обработкой данных: программный блок, реализующий признаки классификации в системе выявления нелегитимной активности;
- Разработана программная система, объединяющая распознавание и классификацию объектов бихевиористического анализа.

Практическая значимость работы заключается в создании промышленного образца системы безопасности и контроля транспортных потоков на производственных территориях. Продемонстрирована на практике эффективность

методики выделения нелегитимной активности, исходя из шаблонов поведения. Функционирование системы «Цербер» привело к резкому снижению нелегитимной активности на Ижорской промышленной площадке. Разработанная программная система, обладая достаточной масштабируемостью, может быть использована и для контроля в крупных транспортных узлах.

На защиту выносятся следующие результаты:

- Разработанная математическая и программная модель выявления нелегитимной активности;
- Разработанная архитектура распределенной программной системы выявления нелегитимной активности на промышленных площадках;
- Разработанные программные блоки классификации нелегитимных событий;
- Программно-аппаратный комплекс «Цербер», реализующий разработанные математические модели и программные средства выявления нелегитимной активности в режиме реального времени.

Апробация. Основные положения диссертации докладывались на семинарах ИПИ РАН, а также ВЦ РАН в период с 2011 по 2013 годы. Результаты, полученные в ходе выполнения данной работы, вошли в ежегодные отчеты по проекту Российского фонда фундаментальных исследований № 11-07-00225 «Интеллектуализация методов описания геоинформационных объектов и технологических процессов». Также, результаты данной работы докладывались на IV Международной научной конференции «Фундаментальные проблемы системной безопасности и устойчивости», а также на Третьей Всероссийской научной конференции «Методы и средства обработки информации».

Публикации. По теме диссертационной работы опубликованы пять печатных работ [1-5], из них четыре в изданиях по перечню ВАК.

Структура диссертации. Диссертация состоит из введения, трех глав, заключения, списка литературы (наименований) и трех приложений. Работа изложена на 113 страницах, включающих 14 рисунков и 4 таблицы. Список использованной литературы включает в себя 79 наименований.

Содержание работы

Во введении обоснована актуальность работы, сформулирована цель и задачи диссертационного исследования, изложена его научная новизна, раскрыты теоретическое значение и практическая ценность полученных результатов, кратко излагается содержание диссертационной работы.

Первая глава посвящена обзору математических и программных средств современных бихевиористического контроля. По результатам данного обзора делаются следующие выводы:

1. Большинство современных систем бихевиористического анализа основаны на видеонаблюдении. В целях повышения стабильности такие системы являются распределенными и модульными. Также их можно назвать «облачными» ввиду наличия у многих из них единого координирующего центра, хранящего большую часть данных системы.
2. Большинство систем реализуется на стандартном и легко заменяемом оборудовании, что повышает их надежность и стабильность работы, что крайне важно для системы обеспечения безопасности, которая должна работать в режиме реального времени.
3. Системы бихевиористического анализа могут работать в Real-time режиме, не требуя значительных вычислительных мощностей – достаточно относительно современных настольных компьютеров со стандартным системным программным обеспечением.
4. Возможности алгоритмов распознавания изображений и машинного зрения в современных системах используются недостаточно эффективно.
5. Существующие системы выявления нелегитимной активности на основе видеонаблюдения используют большое число камер, громоздкие алгоритмы и показывают низкую эффективность.

Вторая глава посвящена разработке математического и программного обеспечения анализа бихевиористического поведения с целью выявления нелегитимной активности.

Приведем краткое описание разработанной математической модели.

Пусть $x_1, \dots, x_n$ – различные признаки, каждый из которых принимает конечные дискретные значения. Пусть S – множество событий. При этом каждому событию

$s \in S$ можно однозначно сопоставить вектор $\bar{x} = (x_1, \dots, x_n)$, компонентами которого являются значения признаков.

Множество S состоит из двух подмножеств: легитимные события L и нелегитимные N. При этом:

$$S = L \cup N, L \cap N = \emptyset.$$

Сопоставим легитимным событиям -1, а нелегитимным -0. Также определим обучающую выборку $S' \in S$ и $L' \in L$. В таком случае, по обучающей выборке S'и L' необходимо построить отображение α, которое:

$$\forall \bar{x} = (x_1, \dots, x_n) \in S,\ \alpha: \bar{x} \rightarrow \{0,1\}.$$

Таким образом, система должна решать задачу классификации, то есть по некоторой конечной обучающей выборке относить событие по ряду признаков или к легитимным или к нелегитимным событиям.

Исходя из формальной постановки задачи, целевая функция α имеет вид:

$$\forall \bar{\mathrm{x}} = (\mathrm{x}_1, \dots, \mathrm{x}_n) \in \mathrm{S},\ \alpha: \bar{\mathrm{x}} \rightarrow \{0,1\}.$$

Она относит событие с признаками $(x_1, \dots, x_n)$ к классу легитимных – 1 или нелегитимных – 0 событий.

Входными данными в решаемой задаче являются дискретные признаки $\mathrm{x}_1, \dots, \mathrm{x}_n$. Фактическое значение n = 8. Перечислим все используемые признаки с их кратким описанием.

Первый признак: «Мониторинг времени прохождения осмотра автомобилем».

Интерпретация признака: если автомобиль на выезде стоит на осмотре дольше некоторого критического времени, то вероятность того, что произошло нарушение, растет пропорционально времени осмотра.

Единица измерения: секунды

Эталонные значения: вычисляется как среднее арифметическое всех времен досмотра тестовой выборки, т.н. $t_{\text{среднее}}$ – среднее время осмотра.

Формула подсчета $x_1 = mod\left(t_{\text{среднее}} - t_{\text{фактическое}}\right)$где: $t_{\text{фактическое}}$ – фактическое время осмотра автомобиля в секундах.

Принимаемые значения:$x_1 \in \overline{1, \dots, 7200} \in N$

Второй признак: «Мониторинг времени движения между различными точками контроля».

Интерпретация признака: измерение времени движения автомобиля из одной точки маршрута в другую. Если автомобиль двигался нетипично долго – есть подозрение, что с грузом были совершены противоправные действия.

Единица измерения: секунды

Эталонные значения: $t_{\text{среднее}}$ – среднее время движения автомобиля по его маршруту движения, Δ – допустимое отклонение:

Формула подсчета: $x_2 = mod\ (\Delta - t_{\text{фактическое}} - t_{\text{среднее}})$, где: $t_{\text{фактическое}}$ – фактическое время движения по маршруту в секундах.

Принимаемые значения:$x_2 \in \overline{0, \dots, 4000} \in N$

Третий признак: «Контроль проезда машины в непредусмотренное время»

Интерпретация признака: каждая машина однозначно идентифицируется по номеру, а также по заявке на въезд, в которой написано ее предполагаемое время прибытия на территорию объекта. Если машина въезжает на территорию в неустановленное время существует вероятность того, что данный въезд не является легитимным.

Единица измерения: секунды

Эталонные значения: $t_{\text{предполагаемого проезда}}$ – предполагаемое время проезда, $t_{\text{фактического проезда}}$ – фактическое время проезда. Оба измеряются в секундах от 0:00:00 дня предполагаемого въезда.

Формула подсчета: $x_3 = mod\ (t_{\text{предполагаемого проезда}} - t_{\text{фактического проезда}})$

Принимаемые значения: $x_3 \in \overline{0, \dots, 180000} \in N$.

Четвертый признак: «Принятие решения на возможность проезда непредусмотренным лицом»

Интерпретация признака: у каждого охранника есть своя именная карточка, путем поднесения которой к считывателю дается разрешение машины на въезд и производится открытие шлагбаума. На объекте ведется график смен охранников. Поэтому если открытие было произведено охранником не из текущей смены, это также является признаком нелегитимного действия. Система может сама выдавать значения – 1 в случае охранника текущей смены и 0 – в обратном случае.

Принимаемые значения: $x_4 = \{0,1\} \in N$

Пятый признак: «Изменение картины воздействия на датчики магнитного поля»

Интерпретация признака: На каждом направлении движения на КПП установлены датчики, измеряющие степень возмущения магнитного поля земли (это позволяет определить факт подъезда машины к шлагбауму). Для каждой машины всегда известны две характеристики: предполагаемое возмущение (исходя из типа машины и груза), а также фактическое возмущение. Оба этих значения измеряются в микротеслах.

Единица измерения: микротеслы.

Формула подсчета: $x_5 = mod\ (T_{\text{предполагаемое}} - T_{\text{фактическое}})$

Принимаемые значения:$x_5 \in \overline{0, \dots, 1000} \in N$

Шестой признак: «Контроль проезда дополнительных точек»

Интерпретация признака: при въезде автомобиля на территорию объекта водителю в пропуск выдается специальная метка дальнего действия. На территории объекта во всех местах вероятной погрузки и выгрузки установлены считыватели данных меток. Это позволяет определить места остановки водителя. С учетом того, что системе известен маршрут движения автомобиля, возможно оценить в каком количестве неустановленных мест останавливался водитель.

Принимаемые значения: $x_6 \in \overline{0, \dots, 100} \in$

Седьмой признак: «Контроль непрохождения обязательных точек»

Интерпретация признака: аналогично предыдущему признаку, только в данном случае оценивается какое количество предполагаемых точек движения не проехал автомобиль.

Принимаемые значения: $x_7 \in \overline{0, \dots, 100} \in N$

Восьмой признак: «Фиксация критической разности качества распознавания номера».

Интерпретация признака: водитель может попытаться обмануть систему путем скрытия номера средства: замазывания грязью и т.д. Система распознавания номеров может оценить, насколько хорошо был распознан номер по шкале от одного до десяти. Разность значений, измеренная на двух разных КПП, может являться одним из критериев выявления нелегитимных действий.

Формула подсчета: $x_8 = mod\ (Q_{\text{въезда}} - Q_{\text{выезда}})$, где $Q_{\text{въезда}}$ – качество распознавания, измеренное на въезде, где $Q_{\text{выезда}}$ – качество распознавания, измеренное на выезде.

Принимаемые значения: $x_8 \in \overline{0,\dots,10} \in N$

Описание на псевдокоде:

recognitionQuality(); функция, возвращающая качество распознавания,

event(CarInSight);

temp:=recognitionQuality();

event(CarExit);

x_8:=mod(recognitionQuality() – temp);

Далее в работе проводится анализ алгоритмов классификации, а также оценка вероятности правильного распознавания текстовых меток на кадре в зависимости от времени анализа.

Пусть анализируются m кадров из N, где $m \le N-k$. При $m > N-k$ вероятность идентификации равна 1, так как хотя бы один кадр, на котором метка распознана, попадает в состав выборки.

Обозначим событие R = «метка объекта распознана хотя бы на одном кадре из m». Оно является противоположным событию $\overline{R}$ = «метка объекта не распознана ни на одном кадре из m».

При $m = 1$ вероятность события $\overline{R}$ равна:

$$P(\overline{R}) = \frac{N-k}{N}.$$

При $m = 2$:

$$P(\overline{R}) = \frac{N-k}{N} \cdot \frac{N-k-1}{N-1},$$

так как вероятность $\overline{R}$ складывается из совместной вероятности событий «метка не распознана на первом выбранном кадре» и «метка не распознана на втором выбранном кадре». Поскольку выбор кадра осуществляется без возврата, то количество благоприятных исходов с каждым выбором уменьшается на единицу и общее количество исходов уменьшается на единицу. Отсюда следует:

$$P(R \mid m) = 1 - \prod_{i=1}^{m} \frac{N-k-i+1}{N-i+1},$$

где $m \le N-k$, $N \ge k$.

В общем виде вероятность распознавания метки на m кадрах, случайным образом выбранных из N, при заданном k:

$$P(R \mid m, N, k) = 1 - \prod_{i=0}^{m-1} \frac{N-k-i}{N-i}.$$

С увеличением m вероятность распознавания метки растет и при $i = N - k \Rightarrow m = N - k + 1$ становится равной единице.

Пусть m – количество кадров, которое система распознавания способна анализировать из общего числа N, где $m \le N$. Данное количество определяется временем анализа отдельного кадра, то есть $m = t_O / T$.

При ограничении количества анализируемых кадров до m ($m < N$) вероятность $P(K_k)$ снижается пропорционально вероятности события $P(R \mid m, N, k)$:

$$P(K_k \mid m, N, k) = P(K_k) P(R \mid m, N, k).$$

Тогда вероятность идентификации объекта контроля рассчитывается:

$$P(Id \mid m, N) = \sum_{k=1}^{N} P(K_k) P(R \mid m, N, k).$$

В итоге связь между показателями T и $P(Id)$ определяется следующим образом:

$$P(Id) \approx \sum_{k=1}^{N} \left(P(K_k) \cdot \left(1 - \prod_{i=0}^{\lfloor t_O / T \rfloor - 1} \frac{N-k-i}{N-i} \right) \right).$$

Эта формула демонстрирует, что с уменьшением времени анализа кадра системой распознавания вероятность идентификации объектов контроля повышается.

Третья глава посвящена описанию программной реализации и результатам тестирования системы. На основании общих требований была разработана архитектура распределенной системы, которая обеспечивает высокую защищенность системы и низкую нагрузку на каналы передачи данных.

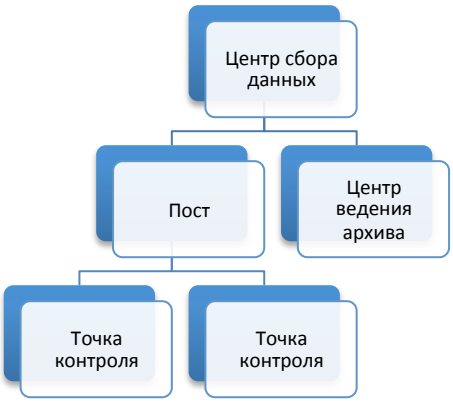

Рисунок 1. Архитектура системы

Особенности архитектуры:

- Высокая защищенность, за счет реализации кеширования данных в точках контроля. В том случае, если центр сбора данных выходит из строя, данные записываются локально до тех пор, пока не появится связь с выделенным сервером
- За счет реализации веб-интерфейса центра сбора данных в рамках стандарта HTML 5.0 интерфейс системы доступен с любого современного устройства, что позволяет отслеживать события с различных мобильных устройств
- Центр ведения архива вынесен в отдельную сущность с целью обезопасить данные от возможной кражи и компрометации. В нем применяется трехкратное дублирование данных для повышения надежности.
- Обмен данными производится при помощи защищенного протокола SSL

Разработанное программное обеспечение представляет собой программный модуль, написанный на платформе Microsoft.Net. Данный модуль скомпилирован в виде исполняемого файла для платформы Wintel, который работает 24/7 и обеспечивает выполнение вышеперечисленных функций. Ниже приводится схема ключевых блоков данного модуля:

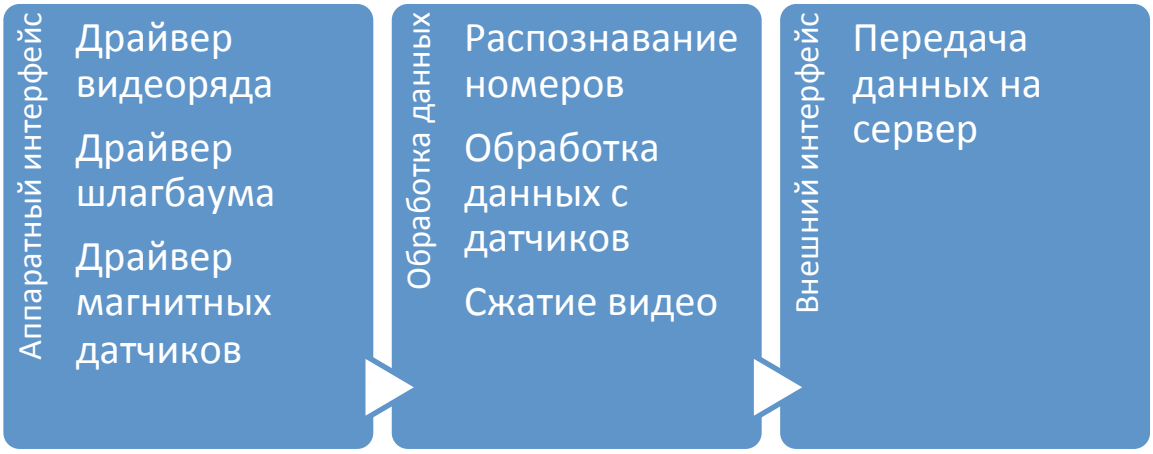

Рисунок 2. Схема работы программного модуля на посту.

Отдельно опишем, какими информационными потоками обмениваются различные части системы. На рисунке ниже приведена схема потоков данных внутри разработанной системы.

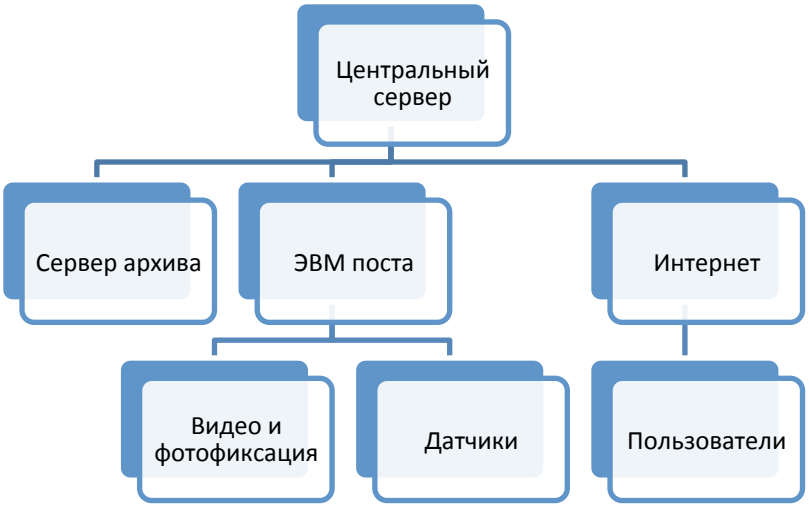

Рисунок 3. Потоки данных внутри системы

Основная решенная задача – предварительная обработка и сжатие фото и видео на ЭВМ поста, позволившая снизить нагрузку на каналы связи на 80%.

Разработанная система была внедрена в ОАО «Ижорские заводы» при следующих условиях:

- ОАО «Ижорские заводы» является одним из крупнейших заводов в России, занимающимся тяжелым машиностроением
- 94% процента потока, проезжающего через КПП заводы – большегрузный транспорт

- Количество выездов машин в месяц – не менее 300 000
- Количество краж в месяц с использованием крупных автомобилей – не менее 30, около 25 из них выявляется постфактум
- До внедрения системы «Цербер» объем краж составлял около 105 миллионов рублей в год
- Система была установлена на трех КПП (на всех, через которые проходит большегрузный транспорт)
- Срок эксплуатации – 5 месяцев

Система была установлена в октябре 2012 года. С октября 2012 года в течение пяти месяцев собиралась статистика, а каждые 24 часа оператор корректировал массив легитимных выездов в системе.

За это время система позволила выявить 100% случаев нелегальных вывозов, из тех, пропажи которых были обнаружены.

Таблица 1. Результаты эксплуатации системы

Параметр	**Значение**	**Описание**
Количество выехавших машин	1 754 874	Количество машин прошедших хотя бы через одно КПП на территории завода с начала эксплуатации системы
Количество выявленных попыток нелегитимных действий	145	Всего выявлено – 145, из них 7 – попытка вывоза продукции, 138 – несоблюдение регламента и/или установленного порядка
Количество невыявленных попыток нелегитимных действий	0	Успешных краж с использованием автотранспорта на территории промышленной площадки с момента установки

		системы не зафиксировано

В заключении диссертации приведены основные результаты работы.

Список публикаций по теме диссертации

1. Абрамов, Н.А. Компьютеризированная система контроля трафика на крупных предприятиях (система «Цербер») [Текст] Труды ИСА РАН, Том 62, выпуск 3, 2012, с. 3-10.
2. Абрамов, Н.А. Кошелев И.И., Применение стереоскопического эффекта для расчета динамических характеристик движущихся автотранспортных средств и схема распознавания номеров в системе «Цербер» [Текст] Международный технико-экономический журнал, №2, 2013, с. 79-84.
3. Абрамов Н.А. Две задачи оптимального управления технологическим процессом [Текст] Труды Института системного анализа РАН, Том 53, вып. 1, 2010, с. 124-131.
4. Абрамов Н.А. Об одном критерии в задаче выбора оптимального маршрута [Текст] Труды Института системного анализа РАН 2010. Т. 32 вып. 2, 2010, с. 316-321.
5. Абрамов Н.А., Качалин А.И. Выбор моделей распространения ВПО при разработке модели глобальной сети [Текст] Труды третьей Всероссийской научной конференции «Методы и средства обработки информации» /Под ред. Л.Н. Королева. – М.: Издательский отдел Факультета ВМиК МГУ им. М.В. Ломоносова, 2009, с. 433-438.