

Министерство науки и высшего образования Российской Федерации  
«МИРЭА – Российский технологический университет»  
РТУ МИРЭА

На правах рукописи

Ермакова Алла Юрьевна

**РАЗРАБОТКА МЕТОДОВ ПРОГНОЗИРОВАНИЯ НЕГАТИВНЫХ  
СОБЫТИЙ ПРИ КОМПЬЮТЕРНОЙ ОБРАБОТКЕ ДАННЫХ**

Диссертация

на соискание ученой степени кандидата технических наук

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

Научный руководитель –  
доктор технических наук,  
профессор  
А.М. Чеповский

Москва – 2021

## ОГЛАВЛЕНИЕ

<b>Введение .....</b>	<b>5</b>
<b>Глава 1. Анализ методов оценки рисков в результате возникновения негативных компьютерных событий в вычислительных машинах, комплексах и компьютерных сетях .....</b>	<b>14</b>
§ 1.1. Постановка задачи и направления исследований .....	14
§ 1.2. Подходы к проблеме обеспечения эффективности процессов обработки информации в вычислительных машинах, комплексах и компьютерных сетях на основе оценки рисков.....	18
§ 1.3. Модели компьютерных событий на основе подбора вероятностных распределений.....	27
§ 1.4. Методы прогнозирования .....	32
1.4.1. Методы дисперсионного анализа .....	32
1.4.2. Метод статистической регрессии .....	36
1.4.3. Модели экспоненциального сглаживания .....	37
1.4.4. Модели авторегрессии .....	39
1.4.5. Комбинированные модели прогнозирования .....	41
Выводы по 1 главе.....	44
<b>Глава 2. Теоретико-вероятностные модели компьютерных событий, возникающих в вычислительных машинах, комплексах и компьютерных сетях, и применение их к анализу эффективности процессов обработки данных.....</b>	<b>46</b>
§ 2.1. Модели появления активных событий в результате взаимодействия программных систем и анализ эффективности процесса обработки данных.....	46
§ 2.2. Модель управления рисками, учитывающая нарастающую сумму ущерба .....	53
§ 2.3. Модель активного компьютерного события в условии ограниченных возможностей по обеспечению эффективности обработки данных и модель попыток доступа к элементам программного обеспечения вычислительных машин, комплексов и компьютерных сетей .....	57
Выводы по 2 главе .....	63

<b>Глава 3. Экспериментальные исследования методов прогнозирования интенсивности компьютерных событий .....</b>	<b>65</b>
§ 3.1. Прогнозирование интенсивности активных событий.....	65
§ 3.2. Прогнозирование негативных событий, связанных с несанкционированными операциями со счетами юридических лиц с использованием платежных карт .....	69
§ 3.3. Прогнозирование динамики появления веб-угроз .....	77
§ 3.4. Прогнозирование появления уязвимостей программных систем.....	82
§ 3.5. Прогнозирование сетевых активных событий .....	86
§ 3.6. Прогнозирование интенсивности компьютерных активных событий с целью майнинга криптовалют. ....	90
§ 3.7. Прогнозирование негативных событий, связанных с ущербом от хищений со счетов юридических лиц с использованием платежных карт....	94
Выводы по 3 главе .....	97
<b>Глава 4. Разработка методов прогнозирования компьютерных событий на основе построения непрерывных аппроксимирующих функций.....</b>	<b>99</b>
§ 4.1. Выбор математического аппарата.....	99
§ 4.2. Построение аппроксимирующих функций на основе модифицированного метода наименьших квадратов.....	105
§ 4.3. Теоретическая оценка точности прогнозирования модифицированного МНК.....	108
§ 4.4. Разработка инструментария для построения аппроксимирующих функций .....	112
Выводы по 4 главе .....	119
<b>Глава 5. Исследование точности применяемого метода прогнозирования компьютерных событий .....</b>	<b>120</b>
§ 5.1. Экспериментальные исследования по построению прогнозных моделей .....	121
5.1.1. Построение прогнозных моделей динамики изменения валютных курсов.....	121
5.1.2. Прогнозирование курсов драгоценных металлов.....	129

5.1.3. Прогнозирование динамики изменения параметров средств вычислительной техники.....	135
§ 5.2. Экспериментальные исследования точности прогнозирования на примерах прогнозирования курсов валют и драгоценных металлов.....	139
§ 5.3. Сравнение точности прогнозирования модифицированным методом МНК и методом полиномов Лагранжа.....	144
§ 5.4. Методика оценки эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях.....	148
Выводы по 5 главе .....	149
<b>Заключение .....</b>	<b>151</b>
<b>Глоссарий .....</b>	<b>156</b>
<b>Литература.....</b>	<b>158</b>
Приложение 1. Свидетельство о государственной регистрации программы на ЭВМ .....	173
Приложение 2. Акты внедрения .....	174

## ВВЕДЕНИЕ

Настоящая диссертационная работа посвящена актуальной задаче – разработке методов исследования компьютерных событий, возникающих в процессе обработки, хранения и передаче данных, которые далее применяются для оценки эффективности указанных процессов. При этом под эффективностью процесса обработки (передачи, хранения) данных понимается свойство указанного процесса, при котором обеспечивается целостность информации и возможность доступа к ней. Под событием (компьютерным событием) понимается любое явление, выходящее за рамки штатной работы вычислительных машин, комплексов и компьютерных сетей, прямо, косвенно или потенциально, ведущее к остановке процессов системы или негативно отражающееся на качестве ее функционирования. Под активным событием понимается взаимодействие программных систем, при котором одна из них осуществляет направленное воздействие на другую.

Вопросы оценки эффективности процессов обработки данных являются центральными при функционировании вычислительных машин, комплексов и компьютерных сетей. В условиях, возникновения событий, вызываемых различными факторами, недостаточное внимание к вопросам обеспечения эффективности информационных процессов может привести к серьезным последствиям, в частности, к значительному ущербу: потере активов, постоянных клиентов, репутации и многому другому. С другой стороны, избыточные меры в этом вопросе приводят не только к необоснованным расходам на приобретение, установку и эксплуатацию дорогостоящего оборудования, но и к значительным затруднениям в работе вычислительных комплексов.

Применяемый в настоящее время большинством специалистов подход к оценке эффективности процессов обработки данных [1-6], основан на оценке рисков, возникающих при появлении компьютерных событий, включающих, в частности, активные события. Указанные риски строятся путем оценки вероятности появления компьютерных событий и оценки ущерба при их реали-

зации. Вывод об эффективности процессов обработки данных в вычислительных машинах, комплексах и компьютерных сетях, как правило, делается на основе сравнения полученных оценок рисков с максимальным допустимым ущербом.

Нетрудно видеть, что в результате применения указанного подхода может быть получен ответ на вопрос об эффективности процессов обработки данных только на момент проведения исследования, а вопрос о времени, в течение которого возможна эффективная реализация процессов обработки данных, очевидно, остается открытым. При этом следует заметить, что данный вопрос более всего интересует руководителей вычислительных комплексов и информационных систем.

#### **Степень разработанности темы.**

Проблемам исследования эффективности обработки данных и управления информационными ресурсами посвящены работы известных российских ученых и специалистов, как А.М. Астахов, В.С. Артамонов, Ю.М. Батулин, А.П. Баранов, В.А. Герасименко, Г.В. Емельянов, В.И. Завгородний, А.А. Зацаринный, П.Д. Зегжда, А.М. Ивашко, А.С. Кузьмин, А.И. Костогрызлов, А.П. Курило, А.Б. Лось, В.Н. Лопатин, А.А. Малюк, В.Г. Матюхин, В.А. Минаев, А.А. Молдовян, А.Г. Остапенко, С.А. Петренко, А.А. Петров, С.П. Расторгуев, С.В. Симонов, С.В. Скрыль, А.В. Старовойтов, А.А. Стрельцов, Д.С. Черешкин и др.

Большой вклад в развитие исследования эффективности обработки данных внесли также и зарубежные исследователи: Р. Андерсон, С. Бармен, Л. Маккарти, М. Мур, Д. Пикфорд и др.

Различные методологические аспекты блокирования доступа к данным вычислительных машин, комплексов и компьютерных сетей отражены в работах специалистов: О.С. Авсентьев, О.И. Бокова, П.Н. Девянин, О.В. Казарин, В.А. Конявский, И.В. Котенко, А.С. Марков, В.А. Минаев, А.Г. Остапенко, С.В. Скрыль, Н.С. Хохлов, А.А. Шелупанов и др. Следует отметить

также ряд работ по отдельным вопросам исследования эффективности процессов обработки данных, в частности доступа к этим данным. К ним относятся труды В.К. Джогана, А.В. Душкина, и др., а также диссертации В.В. Александрова, П.Ф. Сушкова, Н.С. Шимона.

В работе Калашникова А.О. предложена методология временного риск-анализа, которая рассматривается в приложении к переменным состояниям автоматизированных систем управления технологическими процессами в критически важных объектах.

Проблемам прогнозирования и моделирования динамики изменения различных технических и экономических показателей, таких, как курсы мировых валют и мировых цен на продукцию промышленного производства посвящено значительное количество работ как отечественных, так и зарубежных исследователей.

Большое количество научных статей по данной тематике посвящено разработке методов прогнозирования биржевых курсов иностранных валют и котировок ценных бумаг. Из опубликованных работ заслуживают внимание исследования следующих авторов: П.А. Крюков, В.В. Рычков, А.Н. Зинин, Д.С. Литинский, С.В. Смирнов, О.А. Гуляева, М.А. Панилов, Н.В. Рычкова, О.Б. Веретенникова, Д.В. Мамин, А.А. Качур, В.М. Дегтярев, В. Бирюков, Л.Р. Болотова, Т.В. Струченкова, А.Ю. Кузьмин, П.В. Кратович, Е.В. Соколов, Д.В. Бородин, Э.М. Меликов, Л.А. Сафонова, Н.Е. Егорова, К.А. Торжевский и др.

Следует отметить, что все известные подходы к проблеме моделирования различны с точки зрения определения понятия прогнозирования, однако большинство работ рассматривает процедуру построения прогнозных моделей, как определение направления соответствующего тренда, а не как вычисление будущего значения анализируемого параметра.

Анализ известных результатов построению различных прогнозных моделей показывает, что достаточно точный прогноз предлагаемыми методами

может быть дан только на краткосрочный период. Попытки прогнозирования абсолютных значений исследуемых показателей на достаточно длительные временные промежутки дают значительные погрешности. В частности, при построении прогнозов более чем на месяц, ошибки составляют не менее 10%.

Проведенный анализ научных и практических работ показал, что в настоящее время большинством специалистов в области информационных технологий применяется подход к оценке эффективности обработки данных в вычислительных комплексах, основанный на оценке рисков, возникающих при появлении различных событий. Указанные риски строятся путем оценки вероятности появления негативных событий и оценки возможного ущерба при их реализации. Вывод об эффективности процессов обработки данных делается на основе сравнения полученных оценок рисков с максимальным допустимым ущербом.

В результате применения указанного подхода может быть получен ответ на вопрос об эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях только на момент проведения исследования, а вопрос о времени, в течение которого возможно обеспечение эффективности указанного процесса остается открытым. При этом следует заметить, что данный вопрос более всего интересует руководителей вычислительных комплексов и информационных систем, поскольку их работа, как правило, строится в соответствии с перспективным планом.

Решением данной проблемы может стать разработка методов прогнозирования компьютерных событий и применение их для оценки эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях, чему посвящен ряд результатов настоящей работы.

Таким образом, актуальность темы диссертационного исследования заключается в необходимости разработки методов анализа компьютерных событий и алгоритмов прогнозирования информационных рисков с целью разработки инструментальных средств, обеспечивающих анализ эффективности



функционирования вычислительных машин, комплексов и компьютерных сетей.

Актуальность диссертационных исследований подтверждается обсуждением рассматриваемых вопросов на многочисленных форумах, посвященных проблемам эффективности обработки данных. В частности, на проходившей 12 – 13 декабря 2019 года в Москве «Межведомственной сессии по вопросам безопасности критических информационных инфраструктур (КИИ)», представителем подразделения ФСБ России, отвечающем за государственную систему СОПКА, отмечена, в качестве основной задачи ближайшего будущего, разработка методов прогнозирования появления компьютерных событий, поскольку их предотвращение значительно менее болезненная процедура, чем ликвидация последствий.

**Целью диссертационной работы** является разработка и анализ методик, направленных на повышение точности методов прогнозирования событий и определения на этой основе параметров, характеризующих эффективность процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях.

**Основные задачи диссертационного исследования:**

1. Разработать вероятностные модели появления компьютерных событий, возникающих при обработке (передачи, хранения) данных в вычислительных машинах, комплексах и компьютерных сетях, методы их исследования, в частности методика оценки их возможной реализации, а так же методику оценки времени эффективности процесса обработки данных.

2. Разработать алгоритмы и программное обеспечение процедур прогнозирования актуальных угроз, в частности, угроз возникновения негативных компьютерных событий в вычислительных машинах, комплексах и компьютерных сетях и провести экспериментальные исследования эффективности указанных процедур.

3. Разработать методiku прогнозирования динамических процессов обработки данных в вычислительных машинах, комплексах и компьютерных сетях и методiku исследования точности прогнозирования.

**Методы исследования.** Для реализации намеченной цели и решения поставленных задач применяются методы теории вероятностей и математической статистики, математической логики, теории алгоритмов, дискретной математики и теории множеств, численные методы.

**Объект исследования.** Объектом исследования являются процессы обработки (передачи, хранения) данных в вычислительных машинах, комплексах и компьютерных сетях и события, возникающие при их функционировании.

**Предмет исследования.** Предмет исследования определен областью исследования п. 3 «Модели, методы, алгоритмы, языки и программные инструменты для организации взаимодействия программ и программных систем» и п. 10. «Оценка качества, стандартизация и сопровождение программных систем» паспорта специальности 05.13.11 «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей» и перечнем решаемых в диссертации задач.

#### **Научная новизна.**

1. Разработаны вероятностные модели и методы исследования компьютерных событий, возникающих в процессе обработки данных в вычислительных машинах, комплексах и компьютерных сетях, и получены оценки вероятности их реализации. Специфика разработанных методов позволяет рассчитать введенный в данной работе важный параметр эффективности процесса обработки данных в вычислительных комплексах – время, в течение которого обеспечивается целостность информации и возможность доступа к ней.

2. Разработан метод прогнозирования состояний динамических процессов в вычислительных машинах, комплексах и компьютерных сетях, осно-

ванный на построении непрерывных аппроксимирующих функций и предполагающий вычисление конкретного значения на прогнозируемом интервале, а также позволяющий делать обоснованные прогнозы на более длительные промежутки времени и снижающий ошибки прогнозирования более чем в 2 раза по сравнению с известными методами. Указанный метод является универсальным и, в частности, не зависит от природы входных данных, что повышает эффективность обработки данных в вычислительных машинах, комплексах и компьютерных сетях.

3. Проведены теоретические и экспериментальные исследования точности прогнозирования интенсивности появления активных событий в вычислительных машинах, комплексах и компьютерных сетях.

4. Разработана методика оценки эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях, основанная на построении непрерывной временной функции рисков.

**Достоверность и обоснованность** полученных выводов подтверждается вычислительными экспериментами в части исследования динамики поведения актуальных событий, возникающих в процессе обработки данных и корректным использованием математического аппарата прикладной математики.

#### **Теоретическая значимость.**

Теоретическая значимость научных результатов диссертационного исследования заключается в построении методики описания и исследования событий, приводящих к нарушению процессов обработки и передачи данных в вычислительных машинах, комплексах и компьютерных сетях, а также методики оценки эффективности указанных процессов.

**Практическая значимость.** Научные и практические результаты диссертации использованы при проведении исследований эффективности процесса обработки и передачи данных в вычислительных машинах, комплексах и компьютерных сетях в рамках проведения ряда научно-исследовательских

и опытно-конструкторских работ, о чем свидетельствуют: акты о внедрении результатов в работу компании ООО «Смарт-ком»; акт о внедрении результатов в научно-исследовательскую работу кафедры компьютерной безопасности НИУ ВШЭ «Применение риск-ориентированного подхода к оценке стойкости криптографических алгоритмов», акты внедрения результатов диссертации в учебный процесс МТУСИ и РТУ МИРЭА.

Практическая значимость подтверждена свидетельством о государственной регистрации программы для ЭВМ в Федеральной службе по интеллектуальной собственности «Программа построения прогнозных моделей динамических систем».

**Основные положения, выносимые на защиту.**

1. Вероятностные модели и методы исследования компьютерных событий, описывающих взаимодействие программных систем.
2. Метод прогнозирования состояния динамического процесса, теоретические и экспериментальные исследования точности прогнозирования.
3. Методика оценки эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях, основанная на теоретических и экспериментальных методах анализа негативных компьютерных событий и предполагающая прогнозирование их интенсивности.

**Апробация работы.** Результаты диссертации докладывались на Международной научно-практической конференции «Математическое и компьютерное моделирование в экономике, страховании и управлении рисками» в г. Саратове в 2020 г., на Международной конференции «Управление информационной безопасностью», проводимой в НИУ ВШЭ в 2018 году, Международной конференции по современным информационным технологиям в г. Саратове в 2018 году, Международной научно-практической конференции «Традиционная и инновационная наука: история, современное состояние, перспективы» в г. Пермь в 2018 г., Международной научно-практической конференции «Актуальные проблемы социально-экономических исследова-

ний» в 2013 г., на научных семинарах кафедры компьютерной безопасности МИЭМ НИУ ВШЭ и кафедры «Прикладные информационные технологии» РТУ МИРЭА.

**Публикации.** Основные результаты диссертационного исследования опубликованы в 12 работах, из них 8 в изданиях, включённых в перечень рецензируемых журналов, рекомендованных ВАК.

**Личный вклад соискателя в получение результатов, изложенных в диссертации.**

Все результаты, выносимые на защиту, получены автором лично.

**Структура и объем диссертации** Диссертационное исследование состоит из введения, списка основных терминов, 5 глав и списка использованной литературы. Общий объем работы составляет 177 страницы, в работе присутствует 37 рисунков, 41 таблица, библиографический список из 146 наименований и приложения на 5 страницах.

## **ГЛАВА 1. АНАЛИЗ МЕТОДОВ ОЦЕНКИ РИСКОВ В РЕЗУЛЬТАТЕ ВОЗНИКНОВЕНИЯ НЕГАТИВНЫХ КОМПЬЮТЕРНЫХ СОБЫТИЙ В ВЫЧИСЛИТЕЛЬНЫХ МАШИНАХ, КОМПЛЕКСАХ И КОМПЬЮТЕРНЫХ СЕТЯХ**

В настоящей главе рассмотрены основные задачи, решаемые в исследовании, а именно: вопросы оценки эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях, включая методы оценки рисков и их прогнозирование с целью определения возможных сроков – времени, в течение которого обеспечивается эффективность процесса обработки данных.

Рассмотрен традиционный подход к данной проблеме, основанный на оценке рисков и предложено его дальнейшее развитие с учетом динамики изменения существующих угроз и возможных потерь при их реализации. Показано, что применение данного подхода основано на построении прогнозных временных моделей угроз и возможных потерь, что, в свою очередь, требует развития методов прогнозирования состояний динамических процессов.

Рассмотрены известные методы и способы прогнозирования динамики изменения состояний динамических процессов. Дается сравнительный анализ применяемых подходов с точки зрения обоснованности, удобства применения и точности прогнозирования.

### **§ 1.1. Постановка задачи и направления исследований.**

Традиционный подход к оценке эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях, основан на определении актуальных угроз, оценке вероятности их успешной реализации и подсчете возникающих при этом вероятных потерь. Итогом данного исследования является вычисление функции рисков:

$$R = \sum_{i=1}^n p(y_i) \cdot u_i, \quad (1.1.1)$$

где  $p(y_i)$  – вероятность реализации злоумышленником угрозы  $y_i$ ,  $u_i$  – величина ущерба от успешного осуществления данной угрозы,  $i = 1, 2, \dots, n$ .

Далее определяется граница допустимых потерь  $R_0$  и, в случае выполнения условия:

$$R \leq R_0 \quad (1.1.2)$$

процесс обработки данных в вычислительных машинах, комплексах и компьютерных сетях, считается эффективным.

Существенным недостатком такого подхода является то обстоятельство, что вывод об эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях, может быть сделан только на момент проведения исследований, и вопрос о сроке возможной ее эффективной эксплуатации при этом остается открытым. Данная ситуация является весьма неудобной для руководителей организации, специалистов, управляющих разработкой и эксплуатацией программного обеспечения вычислительных машин, комплексов и компьютерных сетей. Их работа осуществляется в соответствии с перспективным планом и, им необходимо знать временной интервал, в течение которого возникновение проблем с программным обеспечением вычислительных машин, комплексов и компьютерных сетей маловероятно.

Решением данной проблемы может служить подход к оценке эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях, предложенный в [71], суть которого состоит во введении зависимости величин от времени вероятности реализации угрозы возникновения событий  $p(y_i)$  и величины ущерба  $u_i$ , характеризующих указанную эффективность процессов:

$$p(y_i) = p_{y_i}(t), \quad u_i = u_i(t).$$

В этом случае функция рисков  $R_0$  также становится функцией от времени  $t$ :

$$R(t) = \sum_{i=1}^n p_{y_i}(t) \cdot u_i(t).$$

Учитывая, что, как правило, функции вероятности реализации угроз  $p_{y_i}(t)$  и функция потерь  $u_i(t)$  являются неубывающими функциями времени  $t$ , можно заметить, что функция рисков  $R(t)$  также является неубывающей функцией  $t$ . В этом случае можно составить уравнение относительно неизвестного переменного  $t$ :

$$R(t) = \sum_{i=1}^n p_{y_i}(t) \cdot u_i(t) = R_0, \quad (1.1.3)$$

где  $R_0$  - максимальное допустимое значения ущерба.

Обозначим через  $T_0$  - решение уравнения (1.1.3).

Учитывая сказанное выше, нетрудно видеть, что в данном методе величина  $T_0$  является временным интервалом, в течение которого потери эффективности процесса обработки данных в вычислительных машин, комплексов и компьютерных сетей достигнут предельно допустимого значения и процесс обработки данных перестанет быть эффективным.

Таким образом, имеет смысл рассматривать величину  $T_0$ , как объективную характеристику эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях. При таком подходе главной задачей становится разработка методов построения непрерывных функций  $p_{y_i}(t)$  и  $u_i(t)$ , в том числе и функций прогнозирования.

Поскольку, как правило, в отношении функционирования вычислительных машин, комплексов и компьютерных сетей, находящихся в эксплуатации, ведется статистика событий, приводящих к нарушению эффективности процесса обработки данных, и существуют методы оценки соответствующих потерь, то эксперты, оценивающие уровень указанной эффективности, располагают данными о динамике изменения указанных параметров.



В этом случае, возможным подходом к построению функций  $p_{y_i}(t)$  и  $u_i(t)$ , может служить метод прогнозирования динамического процесса обработки данных, основанный на построении, так называемых, «аппроксимирующих» функций, наиболее близко расположенных от имеющихся значений параметров состояний данного процесса.

В данной главе приведены результаты аналитического обзора известных подходов к оценке эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях, в том числе, на основе оценки рисков, методов описания процессов возникновения компьютерных событий, а также известных на данный момент методов прогнозирования состояний динамических процессов.

В главе 2 настоящей работы изложены результаты исследований по разработке методов оценки последствий активных компьютерных событий, в том числе, по прогнозированию их интенсивности и применению полученных результатов к построению оценок эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях.

В главе 3 настоящей работы изложены результаты исследований по моделированию компьютерных событий и применению полученных результатов к построению оценок эффективности процесса обработки данных в информационной системе. В ходе экспериментальных исследований рассмотрены вопросы прогнозирования компьютерных событий в части построения непрерывных функций прогнозирования, зависящих от времени на основе данных об интенсивности событий.

Проведенные исследования показали, что наиболее эффективным из рассмотренных методов является метод МНК. В связи с этим, в главах 4 и 5 проведены подробные исследования данного метода, разработан инструментарий для его реализации и проведены исследования его эффективности, в части точности прогнозирования, путем построения статистических критериев.

## **§ 1.2. Подходы к проблеме обеспечения эффективности процессов обработки информации в вычислительных машинах, комплексах и компьютерных сетях на основе оценки рисков.**

Риск нарушения эффективности обработки данных - потенциальная возможность появления различных компьютерных событий, нарушающих штатный режим вычислительных машин, комплексов и компьютерных сетей в части нарушения эффективности обработки данных и доступа к ним, что в свою очередь может привести к причинению ущерба организации [1-3].

В настоящее время управление информационными рисками представляет собой одно из наиболее актуальных и динамично развивающихся направлений менеджмента в области информационных технологий. Его основная задача — объективно идентифицировать и оценить наиболее значимые информационные риски компании, а также адекватность используемых средств контроля рисков для увеличения эффективности и рентабельности экономической деятельности компании.

В 2019 году Комитетом по стандартизации РФ был принят федеральный стандарт «ГОСТ Р 58771-2019. Национальный стандарт Российской Федерации. Менеджмент риска. Технологии оценки риска» [7].

В настоящем стандарте содержатся рекомендации по выбору и применению различных технологий, которые могут быть использованы для совершенствования понимания неопределенности и риска.

### *Управление рисками*

Идеи управления рисками во многом восходят к модели безопасности с полным перекрытием, разработанной в 70-х годах.

Модель системы безопасности с полным перекрытием строится исходя из постулата, что система безопасности должна иметь, по крайней мере, одно средство для обеспечения безопасности на каждом возможном пути воздействия на компьютерную систему.

В модели точно определяется, как обеспечивается эффективность процесса обработки данных, оцениваются средства обеспечения безопасности с точки зрения их эффективности и их вклад в ее обеспечение во всей вычислительной системе.

### *Этапы оценки рисков*

Согласно стандарту [6], оценка рисков включает в себя два последовательных этапа: анализ рисков и оценивание рисков.

Анализ рисков включает в себя:

- идентификацию активов;
- категорирование активов с учетом требований организации и требований законодательства, а также последствий нарушения их конфиденциальности, целостности и доступности;
- определение значимых угроз и уязвимостей идентифицированных активов;
- оценку вероятности реализации угроз и величины уязвимостей.

Оценивание рисков заключается в определении их количественных и качественных значений, ранжирование рисков.

При оценивании рисков рекомендуется рассматривать следующие аспекты:

- шкалы и критерии, по которым можно измерять риски;
- оценку вероятностей угроз;
- методы оценки рисков.

### *Идентификация активов*

Ключевым элементом риска является актив, подверженный этому риску. Риски обеспечения эффективности процесса обработки данных обусловлены наличием у организации информационных активов. К информационным активам относится любая информация, представляющая ценность для организации. Они включают в себя информацию, напечатанную или записанную на бумаге, пересылаемую по почте или демонстрируемую в видеозаписях, пере-

даваемую устно, информацию, хранимую в электронном виде на серверах, веб-сайтах, мобильных устройствах, магнитных и оптических носителях и т.п., информацию, обрабатываемую в корпоративных информационных системах и передаваемую по каналам связи, а также программное обеспечение: операционные системы, приложения, утилиты, программную документацию и т.п.

Информационный актив – это любая информация, независимо от вида её представления, имеющая ценность для организации и находящаяся в её распоряжении. У каждой организации свой набор активов, относящихся к тому или иному типу, например,

- персональные данные;
- стратегическая информация, необходимая для достижения бизнес-целей;
- ноу-хау;
- коммерческая тайна;
- служебная тайна и т.д.

Для каждого актива необходимо определить владельца, который несет за него ответственность.

Стандарт ISO 17799-2005 [6], подробно описывающий процедуры системы управления ИБ, выделяет следующие виды активов:

- информационные ресурсы (базы и файлы данных, контракты и соглашения, системная документация, научно-исследовательская информация, документация, обучающие материалы и пр.);
- программное обеспечение;
- материальные активы (компьютерное оборудование, средства телекоммуникаций и пр.);
- другие активы организации.

При определении рисков следует определить, проблема какими активами может нанести ущерб компании. В этом случае актив будет считаться

ценным, и его необходимо будет учитывать при анализе информационных рисков.

Инвентаризация заключается в составлении перечня ценных активов компании. Как правило, данный процесс выполняют владельцы активов.

Помимо информации организация располагает и другими видами материальных и нематериальных активов, используемые для достижения своих целей. Это имущество организации, имущественные и неимущественные права, интеллектуальная собственность, кадровые ресурсы, а также имидж и репутация организации. Современные международные стандарты также определяют еще одну категорию активов – это процессы, а также информационные и неинформационные сервисы. Это агрегированные типы активов, которые оперируют другими активами для достижения целей.

К видам информационных активов организации, в частности, относятся:

- имущественные и неимущественные права;
- интеллектуальная собственность;
- ноу-хау;
- процессы и сервисы;

Все активы определенным образом взаимосвязаны. Реализация угроз в отношении одних активов, например, помещений или оборудования, может приводить к нарушению эффективности обработки информации, хранимой в этих помещениях или обрабатываемой на данном оборудовании. В свою очередь, нарушение эффективности процесса обработки информации, например, ее достоверности или доступности, может обуславливать финансовые или политические риски. Сбой сервера влияет на доступность хранящихся на нем информации и приложений, а его ремонт отвлекает людские ресурсы, создавая их дефицит на определенном участке работ и вызывая дезорганизацию бизнес-процессов, при этом временная недоступность клиентских сервисов может негативно повлиять на имидж компании.

### *Определение стоимости активов/*

Определение стоимости активов действительно очень сложный процесс [8]. Однако для оценки информационных рисков достаточно оценить стоимость активов приблизительно, т.е. упорядочить ресурсы относительно друг друга. Кроме этого, сотрудникам, оценивающим стоимость активов, следует подробно описать, что от них требуется. Присвоение категорий эффективности процессам обработки информации и компьютерным системам производится на основе оценки ущерба, который может быть нанесен нарушениями процесса эффективности обработки данных в результате наступлений компьютерных событий. Подобные компьютерные события могут помешать организации в выполнении возложенной на нее миссии, скомпрометировать активы, поставить компанию в положение нарушителя действующего законодательства, создать угрозу повседневной деятельности, подвергнуть опасности персонал. Категории эффективности используются совместно с данными об уязвимостях и угрозах в процессе анализа рисков, которым подвержена организация.

Потенциальный ущерб для организации оценивается как низкий, средний или высокий в зависимости от последствий, вызываемых компьютерными событиями, в случае, если в результате компания теряет способность выполнять все или некоторые из своих основных функций.

Категорировать необходимо и пользовательскую, и системную информацию, представленную как в электронной форме, так и в бумажном виде.

При категорировании компьютерной системы принимаются во внимание категории хранимой, обрабатываемой и передаваемой ее средствами информации, а также ценность активов самой КС, то есть берется максимум категорий по всем видам информации и активов. Для получения интегральной оценки следует взять максимум категорий по основным аспектам нарушения эффективности процесса обработки данных.

Организацию процесса оценки стоимости информационных ресурсов обычно обеспечивает руководство компании. Руководству следует сначала издать приказ об официальном утверждении процесса оценки стоимости информационных ресурсов компании, а затем назначить экспертную комиссию, а также сотрудника, ответственного за выполнение оценки стоимости информационных ресурсов.

В соответствии с международным стандартом по информационной безопасности ISO 27001/ISO 17799 [6] оценку стоимости информационного ресурса выполняет его владелец. Владельцем является лицо или сторона, которые имеют утвержденные руководством компании обязанности по управлению созданием, разработкой, поддержанием, использованием и защитой информационных ресурсов. Понятие «владелец» не означает, что какое-либо лицо действительно имеет имущественные права на ресурсы. Таким образом, владельцем информационного ресурса, в большинстве случаев, является сотрудник, создавший и обрабатывающий информационный ресурс. В некоторых случаях сотрудник, обрабатывающий и создающий информационный ресурс, может определить его стоимость (самостоятельно или с помощью руководителей подразделения, сотрудников финансового отдела). Но часто владельцу информационного ресурса выполнить такую оценку сложно. Поэтому стоимость информации оценивает специалист, который имеет достаточные для этого знания и квалификацию (например, руководитель подразделения).

При оценке эффективности процесса обработки данных, в качестве критериев можно взять: стоимость упущенной выгоды (невозможность предоставления услуги или неадекватного функционирования информационной системы); расходы на восстановление информационного ресурса; расходы на создание информационного ресурса (затраты на производство).

### *Идентификация угроз*

Угроза обладает потенциалом причинения вреда активам. Угрозы могут быть природного или человеческого происхождения, они могут быть случайными или умышленными. Должны быть идентифицированы и случайные и умышленные источники угроз и оценена вероятность их возникновения. Важно, чтобы ни одна угроза не была упущена.

Входные данные для оценки угроз должны быть получены от владельцев или пользователей активов, персонала отдела кадров, руководства учреждения и специалистов в сфере информационно-коммуникационных технологий, а также от лиц, отвечающих за обеспечение безопасности организации. Возможно использование статистики угроз из других организаций, таких как юридические организации и правительственные учреждения. При использовании реестров угроз и статистики угроз или результатов проводившихся ранее оценок угроз нужно сознавать, что угрозы постоянно меняются, особенно если меняется деловая среда или информационно-коммуникационные технологии.

После идентификации источника угрозы (кто и что вызывает угрозу) и объекта угрозы (какие элементы системы могут быть затронуты угрозой) необходимо оценить вероятность угроз. При этом следует принимать в расчет:

- частоту угрозы - насколько часто она может происходить в соответствии с опытом, применимой статистикой;
- для умышленных источников угрозы: мотивацию возможных активных компьютерных событий, их возможности, восприятие привлекательности и уязвимости активов компьютерной системы;
- для случайных источников угрозы: географические факторы, такие как близость, к химическим или нефтяным заводам, возможность экстремальных погодных условий и факторы, которые могут влиять на человеческие ошибки и неправильное функционирование оборудования.



В зависимости от потребности в точности, может возникнуть необходимость разбиения активов на компоненты и определения связи угроз с компонентами.

По завершении оценки угроз составляется список идентифицированных угроз, затрагиваемых ими активов или групп активов и меры вероятности того, что угроза произойдет, например, по шкале, такой как высокая, средняя или низкая.

*Количественное определение величины риска* [Астахов ] Количественно величину риска, связанного с осуществлением конкретной угрозы безопасности в отношении конкретного актива, можно выразить при помощи следующей простой математической формулы:

$$РИСК = P_{\text{события}} \times \text{ЦЕНА ПОТЕРИ},$$

где

$$P_{\text{события}} = P_{\text{угрозы}} \times P_{\text{уязвимости}}$$

Соответственно, риск рассчитывается следующим образом:

$$РИСК = P_{\text{угрозы}} \times P_{\text{уязвимости}} \times \text{ЦЕНА ПОТЕРИ}$$

$P_{\text{события}}$  – вероятность успешной реализации угрозы в отношении актива с использованием уязвимости и причинения ущерба организации;

$P_{\text{угрозы}}$  – вероятность того, что угроза в отношении актива будет реализовываться (успешность либо не успешность реализации угрозы определяется величиной уязвимости). На практике для вычисления риска используется не математическая вероятность угрозы, а ожидаемое количество попыток реализации угрозы за определенный период времени.

$P_{\text{уязвимости}}$  – вероятность того, что в случае реализации угрозы в отношении актива эта угроза будет реализована успешно с использованием данной уязвимости, то есть безопасность актива в результате будет нарушена и организация понесет определенный ущерб.

Для определения величины риска используются оценочные количественные значения, полученные путем экспертных оценок, прогнозирования,

а также на основании статистических данных. Размер ущерба, как правило, выражается в денежных единицах, величина уязвимости принимает значения в диапазоне от 0 до 1, а вероятность угрозы является целым положительным числом, определяющим ожидаемое количество попыток реализации угрозы за определенный период времени.

На практике оценка рисков всегда проводится на определенном уровне детализации. Все составляющие риска могут быть разложены на более мелкие составные элементы, что позволяет получить более точные и детализированные оценки рисков, и, наоборот, факторы риска могут быть сгруппированы для получения более общих оценок.

В зависимости от уровня детализации для одной организации могут рассматриваться от нескольких десятков до нескольких сотен и даже тысяч рисков нарушения эффективности процесса обработки данных. Всегда следует начинать с высокоуровневой оценки рисков, которой соответствует наиболее низкий уровень детализации, повышая детализацию по мере необходимости. Поводом для проведения более низкоуровневой (более детализированной) оценки рисков может служить то, что проведенная высокоуровневая оценка не предоставляет достаточной информации для принятия руководством организации обоснованных решений по обработке рисков.

*Качественное определение величины риска.* Конечно, точно определить вероятность угрозы, величину уязвимости либо размер ущерба на практике обычно не представляется возможным, поэтому речь может идти только о числовых оценках в некотором диапазоне величин. Так, ущерб может лежать в диапазоне от 100 тыс. до 1 млн. рублей, либо в диапазоне от 100 млн. рублей до 500 млн. рублей, либо превышать 1 млрд. рублей. Каждому количественному диапазону можно сопоставить определенный качественный уровень риска.

Вероятность сценария инцидента в свою очередь может быть представлена вероятностью осуществления угрозы и уровнем уязвимости. Комбини-

руя угрозу и уязвимости, которые используются для ее реализации, мы получаем сценарий инцидента.

После определения величины рисков, их необходимо упорядочить с целью определения приоритетов обработки этих рисков.

На практике количественный подход к оценке риска, как правило, превращается в качественный подход и, наоборот, в связи с чем, противопоставление качественных и количественных методов оценки рисков является, как минимум нецелесообразным.

### **§ 1.3. Модели компьютерных событий на основе подбора вероятностных распределений**

В работе [82] рассматриваются риск-модели компьютерных событий в распределенных платежных системах на основе банковских карт.

С помощью сетей Петри моделируется простейшее событие, например, заказ товаров и услуг посредством теле маркетинга, имеющая следующий вид:

$S_1$ - событие, состоящее в обращении в Интернет-магазин для заказа товаров;

$S_2$ - событие, состоящее в выборе товара или услуги в Интернет-магазине на определённую сумму, заполнение web-формы,

$S_3$ - событие, состоящее в доставке товара/услуги;

$S_4$ - процессинговый центр проверил введенные данные о карте;

$S_5$ - процессинговый центр заблокировал данный карточный счет и занес карту в стоп-лист;

$t_1$ - время события, состоящее в выборе товара или услуги в Интернет-магазине на определённую сумму;

$t_2$ - время события, состоящее в заполнении в web-форме следующие поля: номер карты (PAN), срок окончания действия карты, проверочный код CVV2;

$t_3$ - время, в течение которого процессинговый центр идентифицирует компьютерное событие.

Построенная сеть Петри для МО «заказ товаров и услуг посредством теле маркетинга» представлена рисунке 1.3.1.

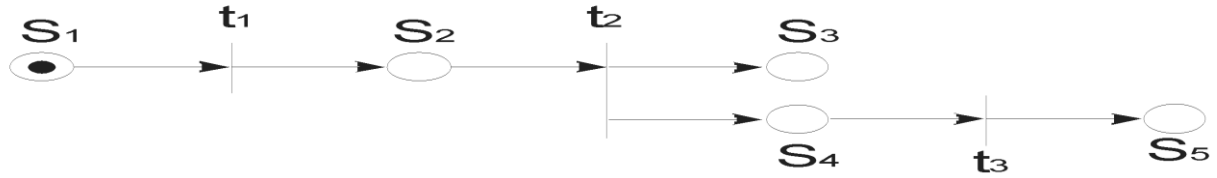


Рис. 1.3.1. Сеть Петри для МО «заказ товаров и услуг посредством теле маркетинга»

Элементы матрицы, определяющие логические функции срабатывания сети, записаны (без учета направленности дуг графа) следующим образом:

$V_{s_1 t_1} =$		$t_1$	$t_2$	$t_3$
	$s_1$	1	0	0
	$s_2$	1	$t_2 s_3 \cap t_2 s_4$	0
	$s_3$	0	1	0
	$s_4$	0	1	1
	$s_5$	0	0	1

Для данной сети Петри-Маркова имеет место следующая система интегро-дифференциальных уравнений:

$$F_{s_1 t_1}(t) = p_{11} \int_0^t f_{s_1 t_1}(\tau) d\tau; \quad F_{s_2 t_2}(t) = p_{22} \int_0^t f_{s_2 t_2}(\tau) F_{s_1 t_1}(t-\tau) d\tau;$$

$$F_{s_4 t_3}(t) = p_{43} \int_0^t f_{s_4 t_3}(\tau) \int_0^t f_{s_3 t_2}(t-\varphi) F_{s_4 t_2}(t-\varphi) + f_{s_4 t_2}(t-\varphi) F_{s_3 t_2}(t-\varphi) d\varphi d\tau.$$

Предполагается, что плотности распределения вероятностей являются экспоненциальными зависимостями и имеют вид:

$$f_{s_i t_j} = \lambda_{ij} \cdot e^{-\lambda_{ij} t}, \quad \text{где } \lambda_{ij} = \frac{1}{\tau_{ij}},$$

где  $\tau_{ij} (i=1...5, j=1...3)$  - средние времена вышеперечисленных действий соответственно.

В частности:

$\tau_{11}$  - среднее время выбора товара (услуги),

$\tau_{22}$  - среднее время заполнения web-формы,

$\tau_{23}$  - среднее время доставки товара

$\tau_{24}$  - среднее время обработки введенных в web-форму данных,

$\tau_{43}$  - среднее время выявления события с момента обработки данных.

Далее применяется пуассоновское приближение, для получения среднего время перемещения по сети Петри-Маркова из начальной позиции до конечного перехода и вероятности этого перемещения:

$$\tau_1 = \tau_{11} + \tau_{22}; \quad \tau_2 = \frac{\tau_{23}^2 + \tau_{23} \cdot \tau_{24} + \tau_{24}^2}{\tau_{23} + \tau_{24}}; \quad \tau = \tau_1 + \tau_{22} + \tau_{43}.$$

Вероятность успеха МО будет равна:

$$p(t) = 1 - \exp\left(-\frac{\tau_{23} + \tau_{24}}{(\tau_{11} + \tau_{22} + \tau_{43})(\tau_{23} + \tau_{24}) + \tau_{23}^2 + \tau_{23} \cdot \tau_{24} + \tau_{24}^2} t\right).$$

Методологические аспекты обеспечения эффективности процесса обработки данных, отражены в работах [109, 110]. В этих трудах разработаны методы обеспечения эффективности процесса обработки данных.

Заслуживает внимания предложенный в [110] подход на основе применения аппарата сетей Петри-Маркова для моделирования угроз информации в компьютерных системах и для построения моделей надежности электронного документооборота.

В частности, в результате рассмотрения основных законов распределения, используемых для модели анализа выживаемости, в [84-98] было установлено, что логлогистическое распределение является наиболее подходящим распределением для описания плотности вероятности отказа системы и её компонентов.

В работе [71] рассматривается распределенная платежная система (РПС) на основе банковских карт (БК) с магнитной полосой и модель несанкционированного доступа типа «кража информации во время ее хранения». В этом случае активное событие и его блокирование состоит из нескольких этапов: выбор хранилища для анализа (время реализации  $t_I$ ), обна-

ружение в выбранном хранилище области авторизованных запросов (время реализации  $t_2$ ), взлом области авторизованных запросов, похищение информации с действующих БК (время реализации  $t_3$ ), блокирование доступа (время реализации  $t_4$ ). В этом случае простейшая модель вероятности успешной реализации данного события имеет вид:

$$p_y(t) = (1 - e^{-\alpha t}) ,$$

где  $\alpha = \frac{1}{t_1 + t_2 + t_3 + t_4}$ .

Нетрудно видеть, что вероятность успешной реализации данного события уменьшается с увеличением общего времени действия, включая время срабатывания блокировки.

Пусть значение возможного ущерба  $u$  от успешной реализации рассматриваемого активного события является постоянной величиной и не зависит от времени. Тогда уравнение для определения параметра  $\tau_0$  – *времени безопасной работы информационной системы* имеет вид:

$$u \cdot (1 - e^{-\alpha t}) = R_0 ,$$

а решением данного уравнения является величина:

$$T_0 = \frac{1}{\alpha} \cdot \ln \frac{u}{u - R_0} .$$

В работе [71] также отмечается, что могут быть построены и другие модели попыток несанкционированного доступа к компьютерным системам, например, модель, предполагающая изменение информации на магнитной карте.

Рассмотрен еще один пример применения данного подхода на основе прогнозирования активных событий на программные системы. Пусть в информационной системе имеется сервер, на который другая информационная система осуществляет взаимодействие путем генерации потока запросов с разных адресов. Предполагается, что максимальное число запросов, обрабатываемых сервером в единицу времени равно  $m_0$ . При превышении данного

числа запросов в единицу времени сервер блокируется, перестает обслуживать запросы.

Для проведения численных расчетов предполагается, что вероятностное распределение числа запросов, поступающих на сервер в единицу времени является стационарным и подчиняется закону Пуассона с параметром  $\lambda$ . Вероятностные модели такого типа часто применяются в системах массового обслуживания [22]. При этом вероятность  $p(m)$  поступления на сервер ровно  $m$  запросов имеет вид:

$$p(m) = \frac{e^{-\lambda} \cdot \lambda^m}{m!}.$$

Из сделанных выше предположений следует, что штатный режим работы сервера обеспечивается в случае, когда число заявок поступающих на него в единицу времени не превышает величины  $m_0$ , вероятность данного события  $p_{um}$ , очевидно, имеет вид

$$p_{um} = \sum_{k=0}^{m_0} \frac{e^{-\lambda} \cdot \lambda^k}{k!}.$$

Вводится обозначение  $p_{am}(t)$  вероятности успешной реализации активного события в течение  $t$  единиц времени. Нетрудно видеть, что

$$p_{am}(t) = 1 - p_{um}^t = 1 - (1 - q)^t,$$

где  $q = 1 - p_{um}$ .

При достаточно малых значениях вероятности  $q$  имеет место соотношение  $(1 - q)^{1/q} \approx e^{-1}$ , поэтому для вероятности  $p_{am}(t)$  справедливо равенство

$$p_{am}(t) \approx 1 - e^{-q \cdot t}.$$

Далее, как и в предыдущем примере, значение возможного ущерба  $u$  от успешной реализации рассматриваемого активного события является постоянной величиной и не зависит от времени и среднее значение допустимого ущерба равно  $R_0$ . Тогда учитывая соотношение (1.1.1), для времени безопасной эксплуатации информационной системы  $T_0$  получаем:

$$T_0 = \frac{1}{q} \cdot \ln \frac{u}{u - R_0}.$$

В общем случае, когда блокировка сервера происходит при поступлении более  $m_0$  запросов в течение промежутка времени  $\tau$ , вероятность работы системы в штатном режиме в течение промежутка времени  $\tau$  имеет вид:

$$P_{um}(\tau) = \sum_{k=0}^{m_0} \frac{e^{-\lambda\tau} \cdot (\lambda\tau)^k}{k!}.$$

Вероятность  $p_{am}(t)$  успешной реализации активного события в течение времени  $t$  может быть оценена величиной:

$$P_{am}(t) = 1 - (P_{um}(\tau))^{t/\tau} = 1 - (1 - q(\tau))^{t/\tau} \approx 1 - e^{-q(\tau) \cdot t/\tau},$$

где  $q(\tau) = 1 - P_{um}(\tau)$ .

Аналогично (1.1.3), для времени безопасной эксплуатации информационной системы  $T_0$  получаем:

$$T_0 = \frac{\tau}{q(\tau)} \cdot \ln \frac{u}{u - R_0}.$$

## § 1.4. Методы прогнозирования

В данном параграфе проведено аналитическое исследование известных подходов к проблеме моделирования состояния динамических процессов. Вопросы прогнозирования далее рассматриваются в 3, 4 и 5 главе настоящей диссертации в связи с экспериментальными исследованиями по прогнозированию появления компьютерных событий, приводящих к нарушению эффективности обработки данных в вычислительных системах и сетях.

### 1.4.1. Методы дисперсионного анализа

Одним из основных методов обнаружения зависимости признака  $y$  от одного или нескольких факторов, является дисперсионный анализ, подробное описание представлено в [83].

Основная задача прогнозирования на основе дисперсионного анализа состоит в проверке гипотезы о равенстве нулю части дисперсии результатив-



ного признака, отражающей его межгрупповую вариацию или/и о равенстве генеральных средних отдельных групп комплекса.

Метод дисперсионного анализа, в частности, состоит в разложении вариации признака  $y$  на части, обусловленные воздействием факторов и их комбинаций, а также в статистическом его анализе. Данный метод был разработан в 1938 году английским статистиком Р. Фишером.

Различают однофакторный и двухфакторный дисперсионный анализ.

*1.4.1.1. Однофакторный дисперсионный анализ при фиксированных и случайных уровнях фактора:*

Суть однофакторного дисперсионного анализа при фиксированных уровнях фактора состоит в следующем [83]:

Для фактора  $A$  задается  $m$  уровней:  $A_1, A_2, \dots, A_m$  и  $y_{ij}$  - результат  $i$ -го наблюдения, выполненного при  $A_j$ -м уровне фактора,

где  $j=1, 2, \dots, m$ ;  $i=1, 2, \dots, n_j$ ;  $n_j$  - число наблюдений при  $A_j$ -м уровне фактора;

$N = \sum_{j=1}^m n_j$  - общее число наблюдений.

Тогда модель однофакторного дисперсионного анализа имеет вид:

$$y_{ij} = \mu + a_j + \varepsilon_{ij},$$

где  $\mu$  - генеральная средняя рассматриваемого комплекса;

$a_j$  - отклонение от генеральной средней  $\mu$ , обусловленное влиянием  $A_j$ -го уровня фактора.

Отклонения  $a_j$  - фиксированные (постоянные) величины, удовлетворяющие условию

$$\sum_{j=1}^m a_j n_j = 0;$$

$\varepsilon_{ij}$  - взаимно независимые, нормально распределенные случайные величины, удовлетворяющие условию:  $\varepsilon_{ij} \in N(0; \sigma)$ .

*Однофакторный дисперсионный анализ при случайных уровнях фактора:*

Формально модель имеет тот же вид что и при фиксированных уровнях фактора

$$y_{ij} = \mu + a_j + \varepsilon_{ij}$$

где  $j=1,2,\dots,m$ ;  $i=1,2,\dots,n_j$ ;

$y_{ij}$  - результат  $i$ -го наблюдения, выполненного при  $A_j$ -м уровне фактора  $A$ ;

$\mu$  - генеральная средняя комплекса;

Основное отличие однофакторного дисперсионного анализа при фиксированных уровнях фактора от рассматриваемого анализа состоит в вероятностной природе величины  $a_j$  - отклонения от генеральной средней  $\mu$ , обусловленное влиянием  $A_j$ -го уровня фактора. Для второго вида анализа величины  $a_j$  - взаимно независимые, нормально распределенные случайные величины с математическим ожиданием

$$M(a_j) = 0 \text{ и дисперсией } \sigma_a^2, \text{ то есть } a_j \in N(0; \sigma_a).$$

*1.4.1.2. Двухфакторный дисперсионный анализ при фиксированных и случайных уровнях факторов и когда уровни факторов имеют смешенную природу [83].*

При двухфакторном дисперсионном анализе исследуется влияние на признак  $Y$  со стороны двух факторов –  $A$  и  $B$ . При этом проверяются следующие гипотезы  $H_0$ :

- фактор  $A$  не влияет на признак  $Y$ ;
- фактор  $B$  не влияет на признак  $Y$ ;
- взаимодействие факторов  $AB$  не влияет на признак  $Y$ .

Все имеющиеся данные, как правило, представляются в виде таблицы, в которой по строкам располагаются уровни  $A_i, i = \overline{1, m}$  фактора  $A$ , по столбцам – уровни  $B_j, j = \overline{1, r}$  фактора  $B$ , а в соответствующих ячейках таблицы нахо-

дятся наблюдаемые значения результативного признака  $y_{ijk}$ , где  $k=1,2,\dots,n_{ij}$  – число наблюдений  $Y$  в ячейке, соответствующей  $i$ -му уровню фактора  $A$  и  $j$ -му уровню фактора  $B$ .

Двухфакторная дисперсионная модель имеет вид:

$$y_{ijk} = \mu + a_i + \beta_j + (a\beta)_{ij} + \varepsilon_{ijk}; \quad i = \overline{1, m}; \quad j = \overline{1, r}; \quad k = \overline{1, n_{ij}},$$

где  $y_{ijk}$  - наблюдаемое  $k$ -е значение результативного признака, полученного на  $i$ -м уровне фактора  $A$ ,  $j$ -м уровне фактора  $B$ , с  $k$ -м порядковым номером,

$\mu$  - общая генеральная средняя комплекса;

$\varepsilon_{ijk}$  - нормальное распределенные случайные величины (остатки), отражающие влияние на  $Y$  все неконтролируемые факторы, т.е. вариацией переменной внутри отдельной ячейки, удовлетворяющей следующим условиям:

$$M(\varepsilon_{ijk}) = 0; \quad M(\varepsilon_{i_1 j_1 k_1} \varepsilon_{i_2 j_2 k_2}) = 0 \text{ для } j_1 \neq j_2 \text{ или } i_1 \neq i_2 \text{ или } k_1 \neq k_2$$

$$M(\varepsilon_{ijk}^2) = \sigma^2 - \text{остаточная дисперсия:}$$

$a_i, \beta_j, (a\beta)_{ij}$  - отклонения от  $\mu$ , обусловленного влиянием  $i$ -го уровня фактора  $A$ ,

$j$ -го уровня фактора  $B$ , взаимодействием этих уровней соответственно.

При этом для модели, когда оба фактора имеют место условия:

$a_i, \beta_j, (a\beta)_{ij}$  - постоянные величины, удовлетворяющие условиям:

$$\sum_{i=1}^m a_i = 0, \quad \sum_{j=1}^r \beta_j = 0 \text{ и, как следствие } \sum_i (a\beta)_{ij} = \sum_j (a\beta)_{ij} = 0.$$

Для модели, когда оба фактора имеют случайные уровни:

$a_i, \beta_j, (a\beta)_{ij}$  - независимые между собой и  $\varepsilon_{ijk}$  - случайные величины, удовлетворяющие условиям:

$$a_i \in N(0; \sigma_a); \quad \beta_j \in N(0; \sigma_\beta); \quad (a\beta)_{ij} \in N(0; \sigma_{a\beta}).$$

Для смешанной модели фактор  $A$  имеет случайные уровни, а фактор  $B$  - фиксированные.

$\beta_j$  - постоянные величины,

$a_i, (\alpha\beta)_{ij}$  - независимые между собой и  $\varepsilon_{ijk}$  случайные величины, удовлетворяющие условиям:

$$a_i \in N(0; \sigma_a); (\alpha\beta)_{ij} \in N(0; \sigma_{\alpha\beta}), \text{ причем } \sum_i (\alpha\beta)_{ij} \neq 0, \sum_j (\alpha\beta)_{ij} = 0.$$

Основная идея метода заключается в разложении суммы квадратов отклонений значений  $y_{ijv}$  от общей средней  $y$  на компоненты, обусловленные влиянием факторов  $A$  и  $B$  и их взаимодействием  $AB$ .

#### 1.4.2. Метод статистической регрессии

Рассмотрим случайные величины  $Y$  и  $X = (X_1, \dots, X_p)$ , связанные некоторой статистической зависимостью, которую можно выразить их совместной функцией распределения  $F_{xy}(x_1, \dots, x_p, y)$ . Пусть случайная величина  $X$  доступна наблюдению, а значение  $Y$  измерить не представляется возможным. При этом возникает задача предсказания величины  $Y$  на основании информации о значениях величин  $X_1, \dots, X_p$ , которые называются предсказывающими переменными. Функция от предсказывающих переменных  $\varphi(X)$ , которую используют в качестве оценки для  $Y$ , называют предиктором величины  $Y$  по  $X$ . Задачей разработки методов оптимальных предикторов занимается теория статистической регрессии [119].

Допустим, совместное распределение  $\lambda(X, Y)$  известно. Тогда можно определить условное распределение  $\lambda(Y | X = x)$ . В частности, если исходное распределение абсолютно непрерывно и  $f_{XY}(x, y)$  - его плотность, то соответствующая условная плотность равна  $f_{X|Y}(y | x) = f_{XY}(x, y) / \int_{-\infty}^{\infty} f_{XY}(x, t) dt$ .

Для дискретных распределений в последней формуле интеграл заменяют соответствующей суммой. Это условное вероятностное распределение имеет среднее  $M(x) = E(Y | X = x)$ , которое зависит от  $x$  и называется функцией регрессии  $Y$  на  $X_1, \dots, X_p$ .

Для абсолютного непрерывного распределения среднее имеет вид:

$M(x) = \int y f_{X|Y}(y|x) dy$  и предполагается, что все соответствующие моменты существуют.

Пусть  $\varphi(X)$  - произвольный предиктор  $Y$  по  $X$ . Тогда среднеквадратической ошибкой этого предиктора является величина  $E[Y - \varphi(X)]^2$ . Предиктор вида  $\varphi^*(X)$  называют оптимальным, если

$$\Delta \equiv E[Y - \varphi^*(X)]^2 = \inf_{\varphi} E[Y - \varphi(X)]^2.$$

Оптимальный предиктор  $M(X)$  обладает следующим свойством: он имеет минимальную корреляцию с  $Y$  среди всех предикторов.

Различают прогнозирование: в случае линейной функции регрессии, линейное и стационарных процессов.

#### 1.4.3. Модели экспоненциального сглаживания.

При краткосрочном прогнозировании динамики экономических показателей, неплохие результаты дают адаптивные методы прогнозирования, на основе процедуры экспоненциального сглаживания. При применении данного метода возникают задачи, связанные с корректировкой временных рядов, выбором параметра адаптации и чрезмерным весом нулевого значения экспоненциальной средней, требуемой для начала работы алгоритма [40,76,121]. Процедура прогнозирования одномерных временных рядов с помощью адаптивных методов состоит в следующем.

##### *Модель экспоненциального сглаживания*

Предполагается, что временной ряд может быть представлен в виде:

$$y_t = a_1 + \varepsilon_t, \text{ где } a_1 = \text{const};$$

$\varepsilon_t$  - случайные не автокоррелированные отклонения с ненулевым математическим ожиданием и дисперсией  $\sigma^2$ . Модель экспоненциального сглаживания ряда описывается следующей рекуррентной формулой:

$$S_t = \alpha \cdot y_t + \beta S_{t-1}, \quad (1.4.3)$$

где  $S_t$  - значение экспоненциальной средней в момент  $t$ ;

$\alpha$  - параметр сглаживания,  $\alpha = const$ ,  $0 < \alpha < 1$ ;  $\beta = 1 - \alpha$ .

При последовательном применении соотношения (1.4.3) экспоненциальную среднюю  $S_t$  можно выразить через предшествующие значения уровней временного ряда. Таким образом, экспоненциальная средняя  $S_t$  имеет вид:

$$S_t = \alpha \sum_{i=0}^{n-1} \beta^i y_{t-i} + \beta^n S_0,$$

где  $n$  – длина временного ряда;

$S_0$  – начальное значение экспоненциальной средней.

Величина  $S_t$  является взвешенной суммой всех членов ряда, выражаемая соотношением:

$$S_t = \alpha \sum_{i=0}^{\infty} \beta^i y_{t-i}, \text{ при } n \rightarrow \infty, \beta^n \rightarrow 0.$$

#### 1.4.4. Модели авторегрессии.

Модели авторегрессии – интегральные значения скользящего среднего [АРПСС( $p, d, q$ )] (autoregressive integrated moving average, ARIMA-models). При прогнозировании нестационарных временных рядов экономических процессов они используются, как правило, со значениями параметров, не превышающих величину 2.

Параметры  $q$  и  $p$  определяют, соответственно, порядок скользящего среднего и порядок авто регрессионной составляющей, а параметр  $d$  – порядок разности (дискретной производной). Скользящие средние используются для исследования компонентов временных рядов, при выявлении характера тренда, при реализации процедур сезонной корректировки и в других исследованиях.

Для описания временных рядов модели ARIMA используются два свойства:

- ряд имеет (аддитивно) составляющую  $f(t)$ , имеющую вид полинома;

- ряд, получившийся после применения к нему процедур последовательных разностей – модель ARMA ( $p, q$ ).

Принцип построения модели ARIMA состоит в следующем:

Задача состоит в получении стационарного ряда. Для этого применяется визуальный анализ графика исследуемого временного ряда, исследование автокорреляционных функций и статистические критерии. Широкое применение получили статистические тесты на наличие единичного корня: тест (расширенный тест) Дики-Фуллера (DF), (Augmented Dickey-Fuller test, ADF), тест KPSS (Kwiatkowski, Phillips, Schmidt, Shin) [24,29]. Для перехода к стационарному ряду применяется процедура дискретного дифференцирования.

Исследуются характер поведения выборочной автокорреляционной функции (АКФ) и выборочной частной автокорреляционной функции (ЧАКФ) и выдвигаются гипотезы о значениях параметров  $p$  (порядок авторегрессии) и  $q$  (порядок скользящего среднего).

Оцениваются параметры моделей с помощью пакетов прикладных программ, в которых используются: метод наименьших квадратов, метод максимального правдоподобия. Однако, как отмечается в [12] эти оценки при больших объемах выборок асимптотически эквивалентны.

Модели проверяются на адекватность с помощью анализа ряда остатков. У адекватной модели распределение коэффициентов автокорреляции приближается к нормальному распределению с нулевым математическим ожиданием и дисперсией  $\frac{1}{n}$ , а именно к  $N(0; \frac{1}{n})$ .

Модель авторегрессии включает члены, описывающие авто регрессионные составляющие, а так же члены моделирующие остаток в виде процесса скользящих средних. Параметры  $p$  и  $q$  определяют соответственно порядок авто регрессионной составляющей и порядок скользящих средних.

Модель ARMA( $p, q$ ) имеет вид:

$$y_t = a_1 \cdot y_{t-1} + a_2 \cdot y_{t-2} + \dots + a_p \cdot y_{t-p} + \varepsilon_t - \theta_1 \cdot \varepsilon_{t-1} - \theta_2 \cdot \varepsilon_{t-2} - \dots - \theta_q \cdot \varepsilon_{t-q}.$$

и представляет собой линейную модель множественной регрессии, в которой объединяющими переменными являются прошлые значения самой зависимой переменной, а регрессионным остатком – скользящее среднее из элементов белого шума [9,12].

Сезонная модель Бокса-Дженкинса имеет вид:

$ARIMA(p,d,q)(P_s,D_s,Q_s)$ , где к параметрам модели  $p,d,q$  добавлены

$P_s$  - сезонный параметр авторегрессии;

$D_s$  - сезонный параметр скользящего среднего;

$Q_s$  - параметр, определяющий порядок сезонной разности (сезонной производной).

При наличии ярко выраженной сезонной компоненты целесообразно включение в модель операции сезонного дифференцирования.

Современные подходы к разработке методики построения моделей ARIMA включают вопросы выявления и учета наличия выбросов в данных.

Многообразие факторов приводит к наличию различных выбросов, отличающихся степенью удаленности, продолжительностью реакции на воздействие внешних факторов, скоростью ее затухания и так далее. Идентифицировать различные типы выбросов позволяют средства, встроенные в аналитические пакеты, в частности, в IBM SPSS STATISTIKA [13].

К недостаткам данного метода можно отнести следующие позиции:

- негативные последствия сглаживания временных рядов с помощью скользящих средних, связанных с эффектом Слущкого-Юла. Производные, сглаженные временные ряды могут содержать систематические колебания, вызванные лишь усреднением случайных составляющих.

- проблема применения скользящих средних связана с необходимостью восстановления потерянных уровней, причем интерес вызывает восстановление последних, а именно, краевых значений. Наиболее распространенным является прием, при котором для определения первых и последних значений  $p$  потерянных уровней используются расчетные значения, полученные с по-



мощью аппроксимирующих полиномов той же степени, что и для сглаживания остальных членов ряда. Неизвестные коэффициенты полиномов определяются соответственно по  $q$  первым и последним уровням временного ряда. Данный прием, описанный в литературе [9,29,56,122], не удовлетворяет потребностям практики, особенно при большой длине интервала.

#### *1.4.5. Построение комбинированных моделей прогнозирования*

Построение прогнозов, на основе комбинированных частных прогнозных оценок, полученных по отдельным моделям из базового набора, посвящены работы зарубежных ученых Дж. Бейтса, П. Ньюболда, С. Макридакиса, К. Гренджера, С. Уилрайта, а также ряда отечественных специалистов: Э.Б. Ершова, Ю.П. Лукашина, В.А. Половникова, А.А. Френкаля, и др. [25,40,76,105,123,125].

Укрупненный процесс построения модели прогнозирования с использованием частных прогнозных оценок может быть представлен в виде следующих этапов [76].

- постановка задачи и начальный анализ исходной статистической информации;
- формирование базового набора моделей с использованием формальных критериев и экономического анализа;
- получение прогнозных оценок на основе моделей, включенных в базовый набор;
- построение итогового прогноза с учетом результатов, полученных на основе частных прогнозных моделей.

На завершающем этапе моделирования проводится анализ частных прогнозных моделей и исследование возможности построения на их основе окончательного прогноза. Возможны два подхода: селективный и гибридный.

При селективном подходе на каждом шаге организуется выбор наилучшей модели по заданному критерию из базового набора. Важное значение имеет как процесс формирования базового набора моделей, так и обоснован-

ный выбор критерия выбора (селекции), позволяющего осуществлять переключение процедуры прогнозирования на более точную модель [76].

При гибридном подходе на каждом шаге прогноз формируется как средневзвешенное значение частных прогнозов, полученных по базовым моделям. Предлагаются различные подходы к определению весовых коэффициентов, характеризующих степень доверия к частным моделям из базового набора. Весовые коэффициенты предполагаются неотрицательными, а их сумма, очевидно, равной единице.

В общем виде задача построения комбинированной модели гибридного типа состоит в необходимости получения на каждом шаге расчетного значения (комбинированного прогноза) в виде взвешенной суммы частных прогнозных оценок:

$$y_{\tau}(t) = \sum_{i=1}^m c_i(t) y_{\tau}^i(t), \quad (1.4.4)$$

где  $y_{\tau}(t)$  – прогноз уровня временного ряда из точки  $t$  на  $\tau$  шагов вперед ;  
 $y_{\tau}^i(t)$  – частный прогноз, полученный по  $i$ -й модели, включенной в базовый набор,  $i = 1 \div m$ ;

$c_i(t)$  – вес, придаваемый  $i$ -му частному прогнозу.

На коэффициенты комбинирования  $c_i(t)$  налагаются следующие ограничения:

$$\sum_{i=1}^m c_i(t) = 1 \quad \text{и} \quad 0 \leq c_i(t) \leq 1.$$

Основная проблема при построении обобщенного прогноза состоит в определении весовых коэффициентов.

Основополагающий метод построения комбинированных прогнозов предложен в работе К. Гренджера и Дж. Бейтса, согласно которому весовые коэффициенты в (1.4.4.1) определяются из минимума дисперсии ошибок обобщающего прогноза.

Весовой коэффициент  $c_1$  в фиксированный момент времени  $t$  определяется следующим образом:

$$c_1 = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2},$$

где  $\sigma_j^2$  – оценки дисперсии ошибок, полученных по  $j$ -й модели ( $j=1,2$ );

$\rho_{12}$  – оценка коэффициента корреляции между ошибками, полученными по частным моделям; весовой коэффициент  $c_2 = 1 - n_1$ . Если между ошибками частных моделей корреляция не выявлена, то выражение упрощается:

$$c_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

Оценки весовых коэффициентов можно считать оценками максимального правдоподобия, так как они находятся из условия максимальной плотности нормального распределения для ошибок частных прогнозов. Распространение полученных результатов на случай объединения двух прогнозов подробно рассмотрено в [105].

Рассмотренные методы имеют следующие недостатки.

Для метода экспоненциального сглаживания основным и наиболее трудным моментом является выбор параметра сглаживания  $\alpha$ , начальных условий и степени прогнозирующего полинома. Кроме того, для определения начальных параметров модели остаются актуальными перечисленные выше проблемы автокорреляций.

Метод вероятностного моделирования. Недостатком этого метода является требование большого количества наблюдений и незнание начального распределения, что может привести к неправильным оценкам.

Метод адаптивного сглаживания. При наличии достаточной информации, в частности, длинных рядах, можно получить надежный прогноз на интервал больший, чем при обычном экспоненциальном сглаживании. Тем не менее, для данного метода нет строгой процедуры оценки необходимой или достаточной длины исходной информации, для конечных рядов нет конкрет-

ных условий оценки точности прогноза. В связи с этим, для конечных рядов существует риск получить весьма приблизительный прогноз.

Модель авторегрессии - метод скользящего среднего. Проблемы метода связаны, прежде всего, с неоднородностью временных рядов и возможности практической реализации данного метода из-за своей сложности.

На основании проведенных исследований можно сделать вывод, что в настоящее время не разработан простой в реализации и эффективный метод прогнозирования динамики изменения состояний динамических процессов на достаточно длительный промежуток времени. Большинство предлагаемых методов весьма громоздки, сложны в реализации и обладают значительными погрешностями.

Помимо перечисленных выше проблем, в ходе проведенного анализа установлено, что для применения всех рассмотренных выше методов прогнозирования необходима обработка большого количества статистических данных. Кроме того, ни один из рассмотренных методов не предполагает построения непрерывной прогнозной функции, которая необходима для прогнозирования возможных сроков, при которых обеспечивается эффективность процессов обработки данных в вычислительных системах и компьютерных сетях. И, что весьма существенно, при построении прогнозных оценок более, чем на 15 временных интервалов (дней, недель, месяцев), ошибки прогнозирования составляют более 10%.

### **Выводы по 1 главе.**

В главе 1 настоящей диссертационной работы приведены результаты аналитического обзора известных подходов к прогнозированию появления компьютерных событий и методов оценки на их основе параметров эффективности процессов обработки данных в вычислительных системах, комплексах и сетях, в том числе, на основе оценки рисков и разработки сценариев появления компьютерных событий.

Рассмотрен подход на основе применения аппарата сетей Петри–Маркова для моделирования компьютерных событий, применяемый ранее для оценки надежности различных систем.

Значительное внимание в 1 главе уделено вопросам прогнозирования, поскольку данные методы позволяют строить экспериментальные модели компьютерных событий и применять их для оценки параметров эффективности процессов обработки данных в вычислительных системах, комплексах и сетях.

## **ГЛАВА 2. ТЕОРЕТИКО-ВЕРОЯТНОСТНЫЕ МОДЕЛИ КОМПЬЮТЕРНЫХ СОБЫТИЙ, ВОЗНИКАЮЩИХ В ВЫЧИСЛИТЕЛЬНЫХ МАШИНАХ, КОМПЛЕКСАХ И КОМПЬЮТЕРНЫХ СЕТЯХ, И ПРИМЕНЕНИЕ ИХ К АНАЛИЗУ ЭФФЕКТИВНОСТИ ПРОЦЕССОВ ОБРАБОТКИ ДАННЫХ.**

Данная глава посвящена вопросам построения теоретико-вероятностных моделей компьютерных событий и применением полученных результатов к построению оценок эффективности процесса обработки данных. С точки зрения теории, разработаны вероятностные модели активных событий в результате взаимодействия программных систем, модель управления рисками учитывающую нарастающую сумму ущерба, модель попыток доступа к элементам программного обеспечения вычислительных машин, комплексов и компьютерных сетей, а также модели активного компьютерного события в условиях ограниченных возможностей противодействия.

Построение теоретико-вероятностных моделей активных компьютерных событий и расчет вероятностей их реализации относится к вопросам оценки информационных рисков. Однако, как будет показано далее в работе, даже в наиболее простых ситуациях численные расчеты искомых параметров, в частности вероятности успешной реализации активного компьютерного события, чрезвычайно сложны и громоздки для получения практических оценок информационных рисков. В связи с этим, в последующих разделах рассмотрен подход к оценке исследуемых параметров на основе применения экспериментальных методов.

### **§ 2.1. Модели появления активных событий в результате взаимодействия программных систем и анализ эффективности процесса обработки данных.**

В данном параграфе построены две вероятностные модели появления активных событий в результате взаимодействия программных систем и исследован вопрос обеспечения эффективности процесса обработки данных [30].

Рассмотрим удаленную сеть (интернет), состоящую из сервера и  $N$  компьютеров. В сети наблюдается процесс событий, среди которых могут осуществляться активные события в отношении процесса обработки данных в ее компонентах. Активные события можно классифицировать как: активные события на уровне приложений; активные события для перегрузки сетевого канала или программы. В любом случае цель состоит в том, чтобы привести к перегрузке ресурсов сервера или вывести веб-ресурс из строя. Причем активные события могут происходить на каждом из семи уровнях модели OSI, которые охватывают все среды коммуникации: начиная с физической среды (1-й уровень) и заканчивая уровнем приложений (7-й уровень), на котором «общаются» между собой программы.

Рассмотрим подробнее реализацию данного процесса на различных уровнях.

1. Процесс возникновения активного события на уровне приложений. Для осуществления активного события с  $N$  компьютеров генерируются запросы на сервер или веб-сайт для доступа к содержащейся там информации. Суть активного события состоит в целенаправленной генерации большого количества запросов в ограниченный промежуток времени, обработать которые сервер или веб-сайт в штатном режиме не имеют возможности по причине ограниченности вычислительных ресурсов. Следствием реализации указанного активного события является снижение производительности или блокировка системы, то есть, потеря эффективности процесса обработки данных.

2. Процесс возникновения активного события на транспортном уровне.

В варианте протокольных активных событий с  $N$  компьютеров постоянно отправляются пакеты данных на сервер или веб-ресурс для последующей сборки. В процессе обработки пакетов данных, серверу требуются подтверждения от IP-адресов, с которых направлены данные и, которые он не может получить из-за загруженности канала. Но, тем не менее, он продолжает далее получать пакеты для распаковки, пока по причине отсутствия необ-

ходимых вычислительных ресурсов, не происходит блокировка системы. Данное активное событие приводит к снижению производительности или блокировке системы, то есть, к частичной, или полной потере эффективности процесса обработки данных.

3. Процесс возникновения активного события на канальном уровне. Для осуществления активного события с  $N$  компьютеров генерируются запросы с целью исчерпания полосы пропускания сетевого канала. Суть активного события состоит в целенаправленной генерации большого количество запросов с целью заблокировать доступ к внешней сети серверу или веб-сайту и существенно снизить эффективность процесса обработки данных в штатном режиме по причине ограниченности полосы пропускания сетевого канала.

Модель появления активного компьютерного события, направленного на затруднении работы программного обеспечения или блокирования доступа к нему, определяется следующим образом. В случайные дискретные моменты времени  $t, t = 1, 2, \dots, T, \dots$  на сервер поступает поток запросов. При этом в фиксированный момент времени поступает только один запрос или не поступает ни одного. Запросы поступают независимо, вероятность появления запроса равна  $p$ .

Определим модель появления активных событий в результате взаимодействия программных систем следующим образом.

В дискретные моменты времени  $t, t = 1, 2, \dots, T, \dots$  на сервер поступают запросы. При этом в фиксированный момент времени поступает только один запрос или не поступает ни одного. Запросы поступают независимо, вероятность появления запроса равна  $p$  [30].

Определим величины  $m_0$  и  $t_0$  следующим образом. Будим считать, что сервер в течение  $t_0$  единиц времени может обработать не более  $m_0$  запросов,  $t_0 > m_0$ .



При поступлении в этот интервал времени  $m_0 + 1$  и более запросов сервер блокируется и активное компьютерное событие считается успешно реализованным.

Обозначим  $p_{am}(T)$  - вероятность успешной реализации активного компьютерного события за время  $T$ .

Для вычисления величины  $p_{am}(T)$  определим случайные величины:

$$\xi_t(t_0) = \mu_t + \mu_{t+1} + \dots + \mu_{t+t_0-1}, \quad t = 1, 2, \dots$$

где  $\mu_t$  - индикатор появления запроса в момент времени  $t$ :

$$\mu_t = \begin{cases} 1, & \text{если запрос наступил} \\ 0, & \text{в противном случае} \end{cases}.$$

Последовательность  $\xi_t(t_0)$ , образованная из независимых случайных величин  $\mu_t$  в научной литературе называется процессом скользящего суммирования или частичными суммами Эрдеша - Реньи [70].

Введем далее индикатор события, состоящего в первом выходе процесса  $\xi_t(t_0)$  на заданную границу  $m_0 + 1$ :

$$v_t(m_0+1) = \begin{cases} 1, & \text{если } \xi_{t-1}(t_0) = m_0, \xi_t(t_0) = m_0 + 1 \\ 0, & \text{в противном случае} \end{cases},$$

а также случайную величину

$\tau(m_0 + 1)$  – время первого выхода процесса  $\xi_t$  на границу  $m_0 + 1$ :

$$\tau(m_0 + 1) = \min \{t \mid v_t(m_0 + 1) = 1\}.$$

Тогда, для вероятности  $p_{am}(T)$  успешной реализации активного компьютерного события за время  $T$  имеем соотношение:

$$p_{am}(T) = p \{ \tau(m_0 + 1) \leq T - t_0 + 1 \}.$$

В соответствии с введенным выше определением нетрудно видеть, что

$$p_{am}(T) = p \{ \tau(m_0 + 1) \leq T \} = p \left\{ \max_{1 \leq t \leq T - t_0 + 1} \xi_t(t_0) \geq m_0 + 1 \right\}. \quad (2.1.1)$$

Изучению предельного распределения статистики  $\max_{1 \leq t \leq T-t_0+1} \xi_t(t_0)$  посвящено довольно много работ в научной литературе. В частности, в работе [70] доказаны следующие утверждения, позволяющие оценить вероятность (2.1.1), и, следовательно, вероятность  $p_{am}(T)$  успешной реализации активного компьютерного события.

Пусть далее  $X_1, X_2, \dots, X_n, \dots$  - независимые, одинаково распределенные случайные величины. Пусть также  $n, N$  - некоторые натуральные числа. Для  $k = 0, 1, 2, \dots$ , положим:

$$\xi_k(n) = X_{k+1} + X_{k+2} + \dots + X_{k+n},$$

$$\eta(n, N) = \max_{0 \leq k \leq N-n} \xi_k(n),$$

$\tau(n, m) = \min(N | \eta(n, N) > 0)$  - время первого достижения заданного уровня  $m$  процессом  $\xi_k(n)$ .

$R^m = R \times R \times \dots \times R$  -  $m$  - мерное Евклидово пространство.

$$\varphi_\gamma(x) = (2\pi)^{-\frac{1}{2}} \cdot \exp\left\{\frac{-x^2}{2\gamma}\right\}.$$

Теорема 1. Пусть  $N, n \rightarrow \infty$  так, что  $\frac{N}{n} \rightarrow m, m > 0$  - целое. Тогда, при любом  $a \in R$  имеет место равенство:

$$\lim_{N, n \rightarrow \infty} p\{\eta(n, N) < a \cdot N^{1/2}\} = \int_{D_1} \dots \int \det \|\varphi_\gamma(y_i - y_{i+1} + a)\|_{i,j=0, \dots, m-1} dy_1 \dots dy_m, \quad (2.1.2)$$

где интегрирование ведется по области  $D_1 \subset R^m$ , задаваемой соотношениями:

$$D_1 = \{0 = y_0 < y_1 < \dots < y_m\}, \gamma = m^{-1}.$$

Теорема 2. Пусть  $N, n \rightarrow \infty$  так, что  $\frac{N}{n} \rightarrow m + \theta, m > 0$  - целое,  $0 < \theta < 1$ .

Тогда, при любом  $a \in R$  имеет место равенство:

$$\begin{aligned}
& \lim_{N, n \rightarrow \infty} p\{\eta(n, N) < a \cdot N^{1/2}\} = \\
& = \int_{D_2} \dots \int \det \|\varphi_{\theta\delta}(x_i - y_j)\|_{i, j=0, \dots, m-1} \det \|\varphi_{(1-\theta)\delta}(y_i - x_{j+1} + a)\|_{i, j=0, \dots, m-1} \times \\
& \times dx_0 \dots dx_m dy_0 \dots dy_m,
\end{aligned} \tag{2.1.3}$$

где интегрирование ведется по области  $D_2$ , задаваемой соотношениями:

$$D_2 = \{0 = x_0 < x_1 < \dots < x_m, y_0 < y_1 < \dots < y_m\}, \delta = (m + \theta)^{-1}.$$

Следствием данных теорем являются следующие утверждения.

**Теорема 3.** Пусть  $n, m \rightarrow \infty$  так, что  $n - m = k \geq 0$ ,  $k = \text{const}$ .

Тогда,

$$p\{\tau(n, m) \cdot n^k \cdot p^n \leq x\} \rightarrow 1 - \exp\left\{-\frac{x \cdot q^{k+1}}{\gamma! p^k}\right\}. \tag{2.1.4}$$

**Теорема 4.** Пусть при  $n, m \rightarrow \infty$  величина  $p \rightarrow 0$  так, что

$$n \cdot p \rightarrow \delta \geq 0, \quad \delta = \text{const},$$

$$\frac{m}{n} \rightarrow \gamma > 0, \quad \gamma = \text{const}.$$

Тогда,

$$p\{\tau(n, m) \cdot C_{N-1}^{m-1} \cdot p^m < x\} \rightarrow 1 - \exp\{-x \cdot e^{-\delta(1-\gamma)}\}, \tag{2.1.5}$$

$$\text{где } C_{N-1}^{m-1} = \frac{(N-1)!}{(N-m+1)! (m-1)!}.$$

Доказательство теорем 3 и 4 следует из теоремы 1 работы [70].

Заметим, что в наших условиях  $n = t_0$ ,  $m = m_0 + 1$ ,  $N = T$ .

В модели, основанной на теории массового обслуживания, задача вычисления вероятности успешной реализации активного компьютерного события в течение времени  $T$  при условии, что блокировка сервера происходит в случае поступления более  $m_0$  запросов в течение промежутка времени  $t_0$ , состоит в следующем.

Рассмотрим простейший поток событий (поступления заявок на сервер), в котором промежутки времени  $X_1, X_2, \dots, X_N, \dots$  между последователь-

ными событиями являются независимыми, одинаково распределенными случайными величинами с функцией распределения  $p\{X_i < t\} = 1 - e^{-\lambda t}$ , где  $\lambda$ -плотность потока событий (среднее число событий, приходящееся на единицу времени) [30].

Определим случайные величины

$$\xi_k(m_0) = X_k + X_{k+1} + \dots + X_{k+m_0}, \quad k=1,2,\dots,$$

где  $X_s$  - промежуток времени между событием с номером  $s$  и событием с номером  $s + 1$ .

Обозначим далее:

$\eta_T$  - случайная величина равная общему числу событий за время  $T$ ,

$\{\chi_T\}$  - событие, состоящее в успешной реализации активного компьютерного события за время  $T$ .

Во введенных обозначениях вероятность  $p_{am}(T)$  проведения успешной реализации активного компьютерного события за время  $T$  имеет вид:

$$p_{am}(T) = \sum_{n=m_0+1}^T p\{\eta_T = n\} \cdot p\{\chi_T / \eta_T = n\}, \quad (2.1.6)$$

где  $p\{\eta_T = n\} = \frac{(\lambda T)^n \cdot e^{-\lambda T}}{n!}$ .

Для вычисления вероятности  $p\{\chi_T / \eta_T = n\}$ , аналогично определенному выше, определим последовательность индикаторов:

$$v_k(t_0) = \begin{cases} 1, & \text{если } \xi_{k-1}(m_0) \geq t_0 + 1, \xi_k(m_0) \leq t_0 \\ 0, & \text{в противном случае} \end{cases}$$

и случайную величину:

$$\tau(t_0) = \min \{k | v_k(t_0) = 1\}.$$

Тогда, для вероятности  $p\{\chi_T / \eta_T = n\}$  получаем:

$$p\{\chi_T / \eta_T = n\} = p\{\min_{1 \leq k \leq n-m_0} \xi_k(m_0) \leq t_0\}. \quad (2.1.7)$$

Нетрудно видеть, что во введенных обозначениях имеет место равенство:

$$p\{\chi_T / \eta_T = n\} = p\{\tau(t_0) \leq n - m_0\},$$

а для ее вычисления могут быть использованы соотношения (2.1.4) и (2.1.5).

Тогда вероятность  $p_{am}(T)$  проведения успешной реализации активного компьютерного события за время  $T$  имеет вид:

$$p_{am}(T) = \sum_{n=m_0+1}^T p\{\eta_T = n\} \cdot p\{\chi_T / \eta_T = n\}, \quad (2.1.8)$$

где  $p\{\eta_T = n\} = \frac{(\lambda T)^n \cdot e^{-\lambda T}}{n!}$ ,  $p\{\chi_T / \eta_T = n\} = p\{\tau(t_0) \leq n - m_0\}$ .

Принимая во внимание, сказанное выше,  $T_0$  - время в течение, которого обеспечивается эффективность процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях является корнем уравнения:

$$p_{am}(t) \cdot U(t) = R_0,$$

где  $U(t)$  - функция потерь (ущерба),

$R_0$  - максимально допустимое значение потерь от реализации активного компьютерного события.

## § 2.2. Модель управления рисками, учитывающая нарастающую сумму ущерба

Предположим, что в случайные дискретные моменты времени в вычислительных машинах, комплексах и компьютерных сетях происходят события  $I_k$ ,  $k=1,2,\dots$ , приводящие к нарушению информационной безопасности: к нарушению эффективности процесса обработки данных: активные компьютерные события, сбой в работе, нарушение правил эксплуатации и т.д. Каждое событие сопровождается ущербом [28].

Положим  $u = \{u_1, \dots, u_m\}$ ,  $u_i > 0$ ,  $i=1,2,\dots,m$  - множество возможных вариантов величины ущерба при появлении активных событий. Программная система реагирует на события, и возможные действия могут осуществляться по нескольким сценариям.

Один из возможных вариантов сценария состоит в последовательном сравнении величин ущерба  $u_k$ , при возникновении очередного события  $I_k$  с максимально допустимой величиной ущерба  $R_0$ . В случае

$$u_k < R_0$$

система продолжает работу в штатном режиме. В противном случае производится корректировка политики управления событиями, введение дополнительных мер по обеспечению эффективности обработки данных и других аналогичных мероприятий.

Второй вариант сценария состоит в следующем:

При появлении в моменты времени  $i_1, i_2, \dots, i_N, \dots$  событий  $I_1, I_2, \dots, I_N, \dots$  производится последовательное суммирование значений соответствующих ущербов  $u_{i_1}, u_{i_2}, \dots, u_{i_N}, \dots$ , вычисление статистики

$$S_N = \sum_{j=1}^N u_{i_j}, \quad (2.2.1)$$

и сравнение ее значения с максимально допустимой величиной ущерба  $R_0$ .

Если при появлении инцидента  $I_{i_N}$  выполняется условие:  $S_N < R_0$ , программная система продолжает работу в штатном режиме. В противном случае, аналогично первому сценарию делается вывод о необходимости корректировке политики управления событиями и введения дополнительных мер по обеспечению эффективной обработки данных.

Рассмотрим математическую модель, соответствующую первому сценарию.

Пусть  $\{\xi_t\}$ ,  $t = 1, 2, \dots, N$  - последовательность независимых, одинаково распределенных индикаторов,

$$p\{\xi_t = 1\} = p, \quad p\{\xi_t = 0\} = 1 - p,$$

характеризующих моменты поступления событий.

Предполагаем, что события появляются независимо друг от друга, вероятность появления каждого равна  $p$ .

Далее, как и ранее, положим  $u = \{u_1, \dots, u_m\}$  - значения возможных значений ущерба и будем полагать, что при наступлении очередного события вероятность возникновения ущерба  $u_k$  равна  $p_k$ ,  $\sum_{k=1}^m p_k = 1$ .

Пусть так же для значений  $k_1, k_2, \dots, k_s$ ,  $k_j \in \{1, \dots, m\}$ , имеет место неравенство

$$u_{k_j} \geq R_0, \quad j \in \{1, \dots, s\},$$

а для значений  $k \in \{1, \dots, m\} \setminus \{k_1, k_2, \dots, k_s\}$  имеет место неравенство

$$u_k < R_0,$$

Положим далее  $q = \sum_{j=1}^s p_{k_j}$

С последовательностью  $\{\xi_t\}$  свяжем последовательность индикаторов  $\{\zeta_t\}$

$$\zeta_t = \begin{cases} 1, & \text{если } \xi_t = 1 \text{ и наступил инцидент } I_k, \quad k \in \{k_1, k_2, \dots, k_s\} \\ 0, & \text{в противном случае} \end{cases}$$

и случайную величину  $\tau$ :

$$\tau = \min\{t \mid \zeta_t = 1\},$$

характеризующую момент наступления критического события, при котором вычислительные машины, комплексы и компьютерные сети нуждаются в дополнительных мерах по обеспечению эффективности процесса обработки данных.

В рамках рассматриваемой модели вероятностное распределение случайной величины  $\tau$  имеет вид:

$$p\{\tau=k\} = p \cdot q \cdot (1-p \cdot q)^{k-1}, \quad k=1,2,\dots, \quad (2.2.2)$$

что является геометрическим распределением.

Математическое ожидание случайной величины  $\tau$  равно  $E\tau = \frac{1}{p \cdot q}$

может служить оценкой среднего времени эффективной обработки данных.

Рассмотрим второй вариант сценария, при котором происходит суммирование значений ущерба от активных событий.

С последовательностью  $\{\xi_t\}$  свяжем последовательность индикаторов  $\{\eta_t\}$

$$\eta_t = \begin{cases} 1, & \text{если } S_{t-1} < R_0, S_t \geq R_0 \\ 0, & \text{в противном случае} \end{cases}, t = 1, 2, \dots,$$

характеризующую момент превышения суммы ущерба установленной границы  $R_0$  и, следовательно, момент наступления критического события, при котором система нуждается в усилении защиты, и определим случайную величину  $\chi$ :

$$\chi = \min\{t \mid \eta_t = 1\},$$

характеризующую момент первого наступления критического события.

Для решения задачи в рамках введенной теоретико-вероятностной модели найдем распределение случайной величины  $\chi$ :

$$p\{\chi=N\} = p_N.$$

Обозначим через  $\bar{\alpha}_N = (\alpha_0, \alpha_1, \dots, \alpha_m)$  первичную спецификацию исходов последовательности  $\{\xi_t\}_t^N$ , где  $\alpha_0$  - число исходов равных 0,  $\alpha_k$  число исходов равных  $u_k$ ,  $\alpha_0 + \alpha_1 + \dots + \alpha_m = N$  и положим:

$$M_{N-1}(u_r) = \left\{ \bar{\alpha}_{N-1} \mid \alpha_0 + \alpha_1 + \dots + \alpha_m = N-1, R_0 - u_k \leq \sum_{k=1}^m \alpha_k \cdot u_k < R_0 \right\}.$$

Тогда для вероятности первого наступления критического события имеем:

$$p\{\chi=N\} = \sum_{k=1}^m P_k \cdot \sum_{\alpha_{N-1} \in M_{N-1}(u_k)} (1-p)^{\alpha_0} \cdot p^{N-\alpha_0} \cdot \prod_{k=1}^m p_k^{\alpha_k}(\alpha_{N-1}). \quad (2.2.3)$$



**§ 2.3. Модель активного компьютерного события в условиях ограниченных возможностей по обеспечению эффективности обработки данных и модель попыток доступа к элементам программного обеспечения вычислительных машин, комплексов и компьютерных сетей**

*2.3.1. Модель активного компьютерного события в условиях ограниченных возможностей по обеспечению эффективности обработки*

Рассмотрим следующую модель появления активного компьютерного события и возможностей механизмов по обеспечению эффективности обработки данных.

В дискретные моменты времени активные компьютерные события в вычислительных машинах, комплексах и компьютерных сетях [29]. События появляются независимо друг от друга, вероятность появления активного компьютерного события в фиксированный момент времени равна  $p$ .

Полагаем, что на нейтрализацию активного компьютерного события и устранение его последствий требуется  $m$  единиц времени, а возможности программного обеспечения имеют ограниченный ресурс и одновременно могут обеспечить эффективность процесса обработки данных только при наступлении одного события. Таким образом, обеспечивается эффективность процесса обработки данных, только если временные интервалы между последовательными активными событиями превышают величину  $m$ .

Для вычисления вероятности  $p_N(m)$  успешной реализации активного компьютерного события в течение  $N$  единиц времени, обозначим через  $n$  общее количество активных компьютерных событий за рассматриваемое время  $N$  и  $p_N(m, n)$  - вероятность успешной реализации активного компьютерного события в этих условиях.

Нетрудно видеть, что в условиях введенной модели вероятность  $p_N(m, n)$  имеет вид:

$$p_N(m, n) = 1 - \varpi_N(m, n) \cdot p^n \cdot (1 - p)^{N-n}, \quad (2.3.1.1)$$

где  $\varpi_N(m, n)$  - число решений уравнения:

$$x_1 + x_2 + \dots + x_n = N, \quad (2.3.1.2)$$

где  $x_i$  - неизвестные целые числа,  $x_i \geq m$ .

Количество вариантов, при котором в  $N$  испытаниях появляется  $n$  единиц и расстояние между соседними не менее  $m$  является числом решений уравнения (2.3.1.2).

Обозначим  $C_{n,N}(\Lambda)$  - число  $N$  выборок в коммутативном несимметричном  $n$ -базисе с первичными спецификациями вида:

$$| a_1^{\alpha_1} a_2^{\alpha_2} \dots a_n^{\alpha_n} |,$$

где  $\alpha_1 + \alpha_2 + \dots + \alpha_n = N$ ,  $\alpha_j \in \Lambda_j$ ,  $\Lambda = (\Lambda_1, \Lambda_2, \dots, \Lambda_n)$ ,  $\Lambda_j$  - область изменения величины  $\alpha_j$ ,  $\alpha_j$  - количество знаков  $a_j$  в общей выборке объема  $N$ .

Нетрудно видеть, что в рамках введенных обозначений имеет место равенство:

$$C_{n,N}(\Lambda) = \sum_{\alpha_1 + \alpha_2 + \dots + \alpha_n = N} 1,$$

где  $\alpha_j \in \Lambda_j$ .

Используя это равенство можно получить выражение для производящей функции  $\varphi_n(t, \Lambda)$ :

$$\varphi_n(t, \Lambda) = \sum_{N=0}^{\infty} C_{n,N}(\Lambda) \cdot t^N = \prod_{j=1}^n \sum_{\alpha_j \in \Lambda_j} t^{\alpha_j}.$$

Для величины  $\Lambda_j$  в нашем случае имеем:

$$\Lambda_j = \{m, m+1, \dots\}, \quad C_{n,N}(\Lambda) = C_{n,N}(m).$$

Тогда

$$\varphi_n(t, m) = \sum_{N=n \cdot m}^{\infty} C_{n,N}(m) \cdot t^N = t^{n \cdot m} (1-t)^{-n},$$

Откуда следует соотношение

$$C_{n,N}(m) = C_{N+n-m-n-1}^{n-1}.$$

В случае

$$\Lambda_j = (m+1, m+2, \dots)$$

получаем:

$$C_{n,N}(m+1) = C_{N+n-(m+1)n-1}^{n-1} = C_{N-n\cdot m-1}^{n-1}.$$

В результате, вероятность  $p_N(m, n)$  успешной реализации активного компьютерного события в случае совершения  $n$  попыток равна:

$$p_N(m, n) = 1 - C_{N-n\cdot m-1}^{n-1} \cdot p^n \cdot (1-p)^{N-n}. \quad (2.3.1.3)$$

Рассмотрим асимптотическое поведение величины в правой части соотношения (2.3.1.3) [28].

*Утверждение 2.3.1.1.* Пусть  $m = \text{const}$ , а величины  $N, n \rightarrow \infty$  так, что  $n^2 = o(N)$ . Тогда

$$C_{N-n\cdot m-1}^{n-1} \cdot p^n \cdot (1-p)^{N-n} = R^{n-1} \cdot e^{-s} \left( \sqrt{2\pi n} \right)^{-1} \cdot (1 + o(1)),$$

где  $R = \frac{N}{n \cdot e}$ ,  $s = N \cdot p + n \cdot \frac{1-2p}{1-p}$ .

*Доказательство.* По определению величины  $C_{N-n\cdot m-1}^{n-1}$ , имеет место равенство:

$$C_{N-n\cdot m-1}^{n-1} = \frac{(N-n\cdot m-1) \cdot (N-n\cdot m-2) \dots (N-n\cdot m-n+1)}{(n-1)!}.$$

Из последнего равенства следует, что

$$\frac{(N-n\cdot(m+1)+1)^{n-1}}{(n-1)!} \leq C_{N-n\cdot m-1}^{n-1} \leq \frac{(N-n\cdot m-1)^{n-1}}{(n-1)!}.$$

Заметим, что

$$(N-n\cdot m-1)^{n-1} = N^{n-1} \cdot \left[ 1 - \frac{n\cdot m+1}{N} \right]^{n-1}.$$

Воспользовавшись известным соотношением

$$(1-x)^{\frac{1}{x} \cdot N} = e^{-N} \cdot (1 + o(1)), \quad x \rightarrow 0, \quad (2.3.1.4)$$

для последнего выражения в условиях утверждения 2.3.1.1 получаем:

$$(N - n \cdot m - 1)^{n-1} = N^{n-1} \exp\left\{-\frac{(n \cdot m - 1) \cdot (n - 1)}{N}\right\} \cdot (1 + o(1)) =$$

$$= N^{n-1} \cdot (1 + o(1)). \quad (2.3.1.5)$$

Далее заметим, что

$$p^n \cdot (1 - p)^{N-n} = \left(\frac{p}{1-p}\right)^n \cdot (1-p)^N. \quad (2.3.1.6)$$

С учетом (2.3.1.4) для второго сомножителя в правой части соотношения (2.3.1.6) получаем:

$$(1-p)^N = (1-p)^{\frac{N \cdot p}{p}} = e^{-N \cdot p} \cdot (1 + o(1)). \quad (2.3.1.7)$$

Обозначим далее

$$\alpha = 1 - \frac{p}{1-p} = \frac{1-2 \cdot p}{1-p}.$$

Тогда

$$\left(\frac{p}{1-p}\right)^n = (1-\alpha)^n = e^{-n \cdot \alpha} \cdot (1 + o(1)).$$

(2.3.1.8)

Из соотношений (2.3.1.7) - (2.3.1.8) далее получаем,

$$p^n \cdot (1-p)^{N-n} = \exp\{-(N \cdot p + n \cdot \alpha)\} \cdot (1 + o(1)) = e^{-s} \cdot (1 + o(1)),$$

(2.3.1.9)

где  $s = N \cdot p + n \cdot \alpha$ .

Применяя к величине  $(n-1)!$  формулу Стирлинга [108], получаем:

$$(n-1)! = \left[(n-1) \cdot e^{-1}\right]^{n-1} \cdot \sqrt{2\pi \cdot n} \cdot (1 + o(1)). \quad (2.3.1.10)$$

Окончательно, из соотношений (2.3.1.5) – (2.3.1.10) получаем,

$$C_{N-nm-1}^{n-1} \cdot p^n \cdot (1-p)^{N-n} = R^{n-1} \cdot e^{-s} \cdot (\sqrt{2\pi n})^{-1} \cdot (1 + o(1)),$$

где  $R = \frac{N}{n \cdot e}$ ,  $s = N \cdot p + n \cdot \frac{1-2p}{1-p}$ ,

что доказывает утверждение 2.3.1.1.

*Замечание.* В общем случае, когда средства обеспечивается эффективность процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях имеют возможность нейтрализовать не более  $m_0$  активных компьютерных событий в течение фиксированного промежутка времени  $t_0$ , задача вычисления вероятности успешной реализации активного события за некоторое время  $T$  сводится к задаче о распределении максимума процесса скользящего суммирования, рассмотренной ранее в параграфе 2.1.

*2.3.2. Модель попыток доступа к элементам программного обеспечения вычислительных машин, комплексов и компьютерных сетей*

Предположим, что осуществляются попытки доступа к некоторым элементам программного обеспечения. Для входа в систему ему необходимо представить правильное значение пароля. Рассмотрим ситуацию, когда злоумышленник имеет множество из  $K$  различных вариантов пароля, каждый из которых может быть истинным с одинаковой вероятностью  $\frac{1}{K}$  и количество попыток для его определения не ограничено [29].

В рамках введенной модели обозначим через  $\xi$  случайную величину равную количеству попыток до определения истинного варианта пароля.

Нетрудно видеть, что

$$P\{\xi = k\} = \left[ \prod_{i=1}^{k-1} \left(1 - \frac{1}{K-i+1}\right) \right] \cdot \left( \frac{1}{K-k+1} \right) = \frac{1}{K}, \quad (2.3.2.1)$$

при этом произведение от 1 до 0 считается равным 1.

Для вычисления среднего значения  $M\xi$  и дисперсии  $D\xi$  случайной величины  $\xi$  докажем утверждение.

*Утверждение 2.2.1.1.* В рамках введенной теоретико-вероятностной модели имеют место соотношения:

$$M\xi = \frac{K+1}{2}, \quad (2.3.2.2)$$

$$D\xi = \frac{K^2-1}{12}. \quad (2.3.2.3)$$

*Доказательство.* Нетрудно видеть, что

$$M\xi = \sum_{k=1}^K k \cdot p\{\xi = k\} = \frac{1}{K} \cdot \sum_{k=1}^K k = \frac{K+1}{2}. \quad (2.3.2.4)$$

Для дисперсии случайной величины  $\xi$  имеем:

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{1}{K} \cdot \sum_{k=1}^K k^2 - \frac{(K+1)^2}{4}. \quad (2.3.2.5)$$

Далее, несложно проверить справедливость равенства:

$$\sum_{k=1}^K k^2 = \frac{2 \cdot K^3 + 3 \cdot K^2 + K}{6} \quad (2.3.2.6)$$

Подставляя выражение (2.2.1.6) в (2.2.1.5) получаем:

$$D\xi = \frac{2 \cdot K^2 + 3 \cdot K + 1}{6} - \frac{(K+1)^2}{4} = \frac{K^2 - 1}{12}, \quad (2.3.2.7)$$

что завершает доказательство утверждения.

Доказанное утверждение позволяет получить оценки отклонения случайной величины  $\xi$  от среднего значения. В соответствии с неравенством Чебышева [10], имеем:

$$p\left\{|\xi - M\xi| \geq \frac{K}{s}\right\} \leq \frac{D\xi \cdot s^2}{K^2} = \frac{(K^2 - 1) \cdot s^2}{12 \cdot K^2} = \frac{s^2}{12} \cdot \left(1 + O\left(\frac{1}{K^2}\right)\right), \quad (2.3.2.8)$$

где  $s > 0$  - некоторая константа.

Обозначим далее через  $w$  – производительность вычислительной системы применяемой для определения пароля и получения доступа к программному обеспечению вычислительных машин, комплексов и компьютерных сетей. Величина  $w$  может быть представлена как количество вариантов пароля, которые могут быть опробованы в единицу времени.

Тогда среднее время  $T_{\text{взл}}$ , необходимое для получения доступа к программному обеспечению имеет вид:

$$T_{\text{взл}} = \frac{K+1}{2w}. \quad (2.3.2.9)$$

Обозначим через  $t$  некоторый временной интервал, а  $p_t$  - вероятность получения доступа к программному обеспечению в течение времени  $t$ .

Тогда, очевидно,

$$p_t = P\{\xi \leq r(t)\}, \quad (2.3.2.10)$$

где  $r(t) = t \cdot w$ .

Учитывая сделанные выше замечания, для вероятности  $p_t$  получаем:

$$p_t = \begin{cases} \frac{t \cdot w}{K}, & \text{если } t \leq \frac{K}{w}, \\ 0, & \text{в противном случае.} \end{cases} \quad (2.3.2.11)$$

Если, как и ранее, положить  $U(t)$  – функция ущерба при несанкционированном доступе к программному обеспечению, то  $T_0$  - время в течение, которого обеспечивается эффективность процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях с учетом данной угрозы находится из уравнения:

$$\frac{t \cdot w}{K} \cdot U(t) = R_0, \quad (2.3.2.12)$$

в частности, при  $U(t) = U = const$ , имеем:

$$T_0 = \frac{K \cdot R_0}{w \cdot U}. \quad (2.3.2.13)$$

Из последнего соотношения следует, что время эффективной обработки данных в вычислительных машинах, комплексах и компьютерных сетях относительно угрозы несанкционированного доступа к их элементам, и, следовательно, уровень обеспечения эффективной обработки данных, прямо пропорционален надежности парольной системы (количеству вариантов пароля) и обратно пропорционально наносимому ущербу.

### **Выводы по 2 главе.**

В настоящей главе рассмотрены вопросы построения теоретических моделей активных компьютерных событий в вычислительных машинах, комплексах и компьютерных сетях. В частности, построены вероятностные модели активных событий в результате взаимодействия программных си-

стем, модель управления рисками учитывающую нарастающую сумму ущерба, модель попыток доступа к элементам программного обеспечения вычислительных машин, комплексов и компьютерных сетей, а также модели активного компьютерного события в условиях ограниченных возможностей противодействия.

Полученные результаты применены для разработки методов оценки важного параметра, характеризующего процесс обработки данных в вычислительных машинах, комплексах и компьютерных сетях - время в течение, которого обеспечивается эффективность процесса обработки указанных данных.

Приведенные выше результаты показывают, что даже в простых теоретических моделях компьютерных событий выражения для вероятности их успешной реализации имеют весьма сложный вид. В связи с этим для оценки времени эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях и предложены экспериментальные методы построения указанных событий.



### **ГЛАВА 3. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ МЕТОДОВ ПРОГНОЗИРОВАНИЯ ИНТЕНСИВНОСТИ КОМПЬЮТЕРНЫХ СОБЫТИЙ.**

Данная глава посвящена изложению результатов исследований по моделированию компьютерных событий и применению полученных результатов к построению оценок эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях. В ходе экспериментальных исследований рассмотрены вопросы прогнозирования компьютерных событий в части построения непрерывных функций прогнозирования, зависящих от времени на основе данных об интенсивности негативных событий.

Выше в работе были построены теоретико-вероятностные модели ряда активных компьютерных событий и, было показано, что даже в наиболее простых ситуациях численные расчеты искомых параметров, в частности вероятности успешной реализации активных компьютерных событий, чрезвычайно сложны и громоздки для получения практических оценок информационных рисков. В связи с этим, в последующих разделах рассмотрен подход к оценке исследуемых параметров на основе применения экспериментальных методов. Данные методы позволяют строить оценки важного параметра, характеризующего процесс обработки данных в вычислительных машинах, комплексах и компьютерных сетях - время в течение, которого обеспечивается эффективность процесса обработки указанных данных.

#### **§ 3.1. Прогнозирование интенсивности активных событий.**

В настоящем разделе приведены результаты прогнозирования интенсивности активных событий в вычислительных машинах, комплексах и компьютерных сетях, основанной на данных об ее предыдущих значениях, опубликованных на сайте Лаборатории Касперского [113]. По указанным данным, с применением метода прогнозирования, изложенного ниже в работе, строится прогнозная модель интенсивности активных событий в виде аппроксимиру-

ющей непрерывной функции [30]. При заданной величине ущерба от реализации активных событий, построенная модель позволяет оценить время в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях.

В таблице 3.1.1 приведена статистика числа активных событий за период с 1 января 2017 года по 27 января 2017 года с интервалом 3-4 дня, представленная на сайте Лаборатории Касперского [113].

Таблица 3.1.1. Динамика числа активных событий.

Период	Динамика числа активных событий
1 января 2017 г.	793
3 января 2017 г.	100
7 января 2017 г.	560
11 января 2017 г.	300
15 января 2017 г.	600
19 января 2017 г.	500
23 января 2017 г.	450
25 января 2017 г.	120
27 января 2017 г.	400

В рассматриваемом эксперименте для построения прогнозной функции  $F(x) = F_1(x)$  использовались данные за период с 1 по 27 января 2017 года, за нулевое значение по оси ОХ принята дата 1 по 27 января 2017 года.

Для данного эксперимента аппроксимирующая функция  $F(x)$  имела вид:

$$F(x) = 69.5 \cdot \sin[2x] + 89.51 \cdot \cos[3 \cdot x] - 71.83 \cdot \cos[x] + 43.27 \cdot \cos[2 \cdot x] + 114.55 \cdot \ln x + \frac{926.25}{x^2} + 201.53 \cdot \sin[3x] + 185.67 \cdot \sin[4x] - 144.54 \cdot \cos[4 \cdot x].$$

График, прогнозной функции  $F(x)$  представлен ниже на рисунке 3.1.1. На данном графике по оси X расположены даты, начиная с 01.01.2017 г., по оси Y – количество активных событий. Точками на графике являются, приведенные для наглядности, значения исходных данных, на основании которых строится сама аппроксимирующая функция.

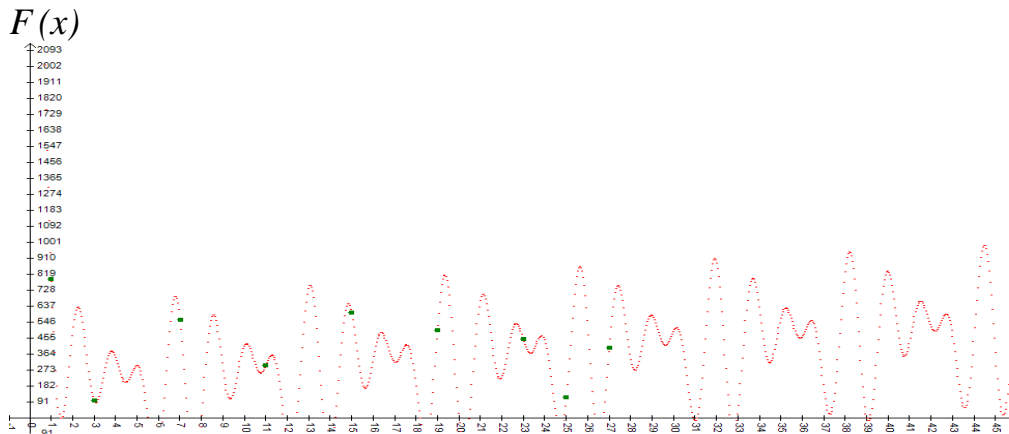


Рис. 3.1.1. График «аппроксимирующей» прогнозной функции  $y = F(x)$ .

В таблице 3.1.2 приведены результаты эксперимента по построению прогнозных значений динамики числа активных событий на период с 1 января 2017 года по 24 апреля 2017 года с интервалом от 2 до 10 дней.

Таблица 3.1.2. Прогноз динамики числа активных событий.

Период	Динамика числа активных событий
1 января 2017 г.	792,9
3 января 2017 г.	100
7 января 2017 г.	560
11 января 2017 г.	300
15 января 2017 г.	600
19 января 2017 г.	500
23 января 2017 г.	450
25 января 2017 г.	120
27 января 2017 г.	399,9
11 февраля 2017 г.	642

Период	Динамика числа активных событий
20 февраля 2017 г	760
1 марта 2017 г.	334
3 марта 2017 г.	369
6 марта 2017 г.	244
10 марта 2017 г.	405
14 марта 2017 г.	75
23 марта 2017 г.	530
30 марта 2017 г.	737
4 апреля 2017 г.	665
7 апреля 2017 г.	379
10 апреля 2017 г.	732
13 апреля 2017 г.	868
22 апреля 2017 г.	589
24 апреля 2017 г.	600

В соответствии с примененным выше методом построения прогнозной модели динамики числа активных событий, функция их числа имеет вид:

$$f(t) = \alpha_1 \cdot \sin[2 \cdot t] + \alpha_2 \cdot \cos[3 \cdot t] - \alpha_3 \cdot \cos[t] + \alpha_4 \cdot \cos[2 \cdot t] + \alpha_5 \cdot \ln[t] + \frac{\alpha_6}{t^2} + \alpha_7 \cdot \sin[3 \cdot t] + \alpha_8 \cdot \sin[4 \cdot t] - \alpha_9 \cdot \cos[4 \cdot t]. \quad (3.1.1)$$

где  $\alpha_i$ ,  $i = 1, 2, \dots, 9$  - соответствующие константы.

При  $t \geq 1$  для функции  $f(t)$  имеет место неравенство:

$$f(t) \leq \alpha_{10} + \alpha_5 \cdot \ln(t), \quad (3.1.2)$$

где  $\alpha_{10} = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_6 + \alpha_7 + \alpha_8 + \alpha_9$ .

Обозначим через  $U$  – средний ущерб, наносимый одной программной системе, а через  $N$  – количество программных систем, для которых проводились исследования.

Тогда прогнозная модель среднего ущерба программной системы имеет вид:

$$\frac{U \cdot f(t)}{N} \leq \frac{U}{N} \cdot (\alpha_{10} + \alpha_5 \cdot \ln(t)), \quad (3.1.3)$$

При этом оценка времени  $T_0$  в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях может быть найдена из уравнения:

$$\frac{U}{N} \cdot (\alpha_{10} + \alpha_5 \cdot \ln(t)) = R_0,$$

откуда получаем:

$$T_0 \geq \exp\left(\frac{R_0 \cdot N / U - \alpha_{10}}{\alpha_5}\right), \quad (3.1.4)$$

где  $R_0$  - максимально допустимый ущерб программной системе.

### **§ 3.2. Прогнозирование негативных событий связанных с несанкционированными операциями со счетами юридических лиц с использованием платежных карт.**

*3.2.1. Прогнозирование негативных событий связанных с несанкционированными операциями со счетами юридических лиц с использованием платежных карт.*

В данном эксперименте прогнозирование проводилось с целью изучения динамики негативных событий, связанных с несанкционированными операциями со счетами юридических лиц с использованием платежных карт [28]. В таблице 3.2.1.1 приведена статистика количества несанкционированных операций со счетами юридических лиц за период с 1 квартала 2016 года по 4 квартал 2017 года, представленная на официальном сайте Банка России [60].

Таблица 3.2.1.1. Количество несанкционированных операций (хищений).

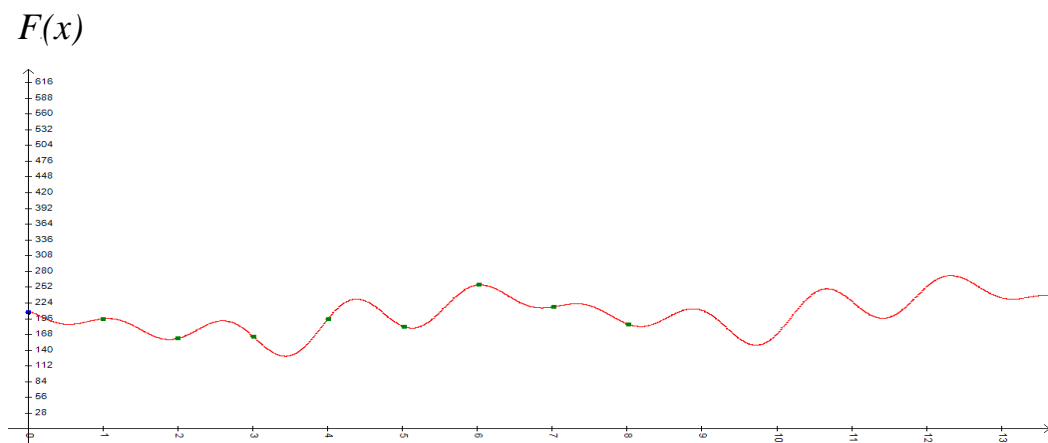
Период	Количество несанкционированных операций
1 квартал 2016 г.	196
2 квартал 2016 г.	161
3 квартал 2016 г.	164
4 квартал 2016 г.	196
1 квартал 2017 г.	182
2 квартал 2017 г.	256
3 квартал 2017 г.	217
4 квартал 2017 г.	186

В эксперименте для построения прогнозной аппроксимирующей функции  $F(x)$  использовались данные за период с 1 квартала 2016 года по 4 квартал 2017 года, за нулевое значение по оси ОХ принята дата 4 квартал 2015 года.

Для данного эксперимента аппроксимирующая функция  $F(x)$  имела вид:

$$F(x) = 164.19 + 15.63\sqrt{x} - 3.41 \cdot \sin[x] - 0.31 \cdot \sin[2 \cdot x] - 23.48 \cdot \sin[4 \cdot x] + 8.3 \cdot \sin[3 \cdot x] + 15.13 \cdot \cos[3 \cdot x] + 28.21 \cdot \cos[x] \quad (3.2.1.1)$$

График, прогнозной функции  $F(x)$  представлен ниже на рисунке 3.2.1.1.

Рис. 3.2.1.1. График функции  $y = F(x)$ .

На данном графике по оси X расположены даты, начиная с 01.01.2016 г., одно деление соответствует одному кварталу, по оси Y – количество несанкционированных операций. Точками на графике являются, приведенные для наглядности, значения исходных данных, на основании которых строится сама аппроксимирующая функция.

В таблице 3.2.1.2 приведены результаты эксперимента по построению прогнозных значений количества несанкционированных операций с использованием платежных карт на период с 1 квартала 2018 года по 2 квартал 2019 года.

Таблица 3.2.1.2. Прогноз количества несанкционированных операций (хищений).

Период	Количество несанкционированных операций
1 квартал 2018 г.	211
2 квартал 2018 г.	168
3 квартал 2018 г.	227
4 квартал 2018 г.	252
1 квартал 2019 г.	233
2 квартал 2019 г.	221
3 квартал 2019 г.	223
4 квартал 2019 г.	162

В данном примере, в соответствии с примененным выше методом построения прогнозной модели динамики негативных событий, связанных с несанкционированными операциями со счетов юридических лиц с использованием платежных карт, функция их числа  $f(t)$  в зависимости от времени  $t$  имеет вид:

$$f(t) = \alpha_1 + \alpha_2 \sqrt{t} - \alpha_3 \cdot \sin[t] - \alpha_4 \cdot \sin[2 \cdot t] - \alpha_5 \cdot \sin[4 \cdot t] + \alpha_6 \cdot \sin[3 \cdot t] + \alpha_7 \cdot \cos[3 \cdot t] + \alpha_8 \cdot \cos[t], \quad (3.2.1.2)$$

где  $\alpha_i, i=1,2,\dots,8$  - соответствующие константы

При  $t \geq 0$  для функции  $f(t)$  имеет место неравенство:

$$f(t) \leq \alpha_9 + \alpha_2 \sqrt{t}, \quad (3.2.1.3)$$

где  $\alpha_9 = \alpha_1 + \alpha_3 + \dots + \alpha_8$ .

Обозначим через  $U$  – средний ущерб от несанкционированной операции и, как и ранее, через  $N$  - число юридических лиц (программные системы), понесших ущерб от программной системы.

Тогда прогнозная модель среднего ущерба юридического лица имеет вид:

$$\frac{U}{N} \cdot f(t) \leq \frac{U}{N} \cdot (\alpha_9 + \alpha_2 \sqrt{t}), \quad (3.2.1.4)$$

При этом оценка времени  $T_0$  в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях может быть найдена из уравнения:

$$\frac{U}{N} (\alpha_9 + \alpha_2 \sqrt{t}) = R_0, \quad (3.2.1.5)$$

откуда получаем:

$$T_0 \geq \left( \frac{N \cdot R_0 / U - \alpha_9}{\alpha_2} \right)^2, \quad (3.2.1.6)$$

где  $R_0$  - максимально допустимый ущерб программной системе.

В таблицах 3.2.1.3 и 3.2.1.4 представлены примеры расчетов параметра  $T_0$  - времени (в годах) в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях, примеры рассмотрены для  $N=10$  и  $N=20$ .



Таблица 3.2.1.3. Время  $T_0$  (в годах)

$R_0$ (тыс. руб.) \ $U$ (тыс. руб.)				
	150	160	170	180
10	3,3	6	9,6	14

Таблица 3.2.1.4. Время  $T_0$  (в годах)

$R_0$ тыс.руб. \ $U$ (тыс. руб.)						
	260	280	300	320	340	350
20	0,3	1,4	3,3	6	9,6	10,2

### 3.2.2. Прогнозирование негативных событий связанных с несанкционированными операциями (хищениями) со счетов юридических лиц

Очевидно, что финансовые организации принимают различные меры для противодействия попыткам создания негативных событий в виде хищений денежных средств со счетов, как юридических, так и физических лиц [37]. С этой целью каждая транзакция проверяется на легитимность путем применения ряда процедур. В случае выявления негативных событий, проведение транзакции останавливается для осуществления дополнительных процедур подтверждения ее легитимности. Однако в ряде случаев все же удается совершить несанкционированные операции, далее называемые неостановленными.

В рассматриваемом эксперименте прогнозная модель строилась для изучения динамики негативных событий, связанных с количеством неостановленных несанкционированных операций с использованием платежных карт. В таблице 3.2.2.1 приведена статистика в процентном соотношении количества неостановленных несанкционированных операций за период с 1 квартала 2016 года по 4 квартал 2017 года, представленная в [60].

Таблица 3.2.2.1. Процент неостановленных несанкционированных операций.

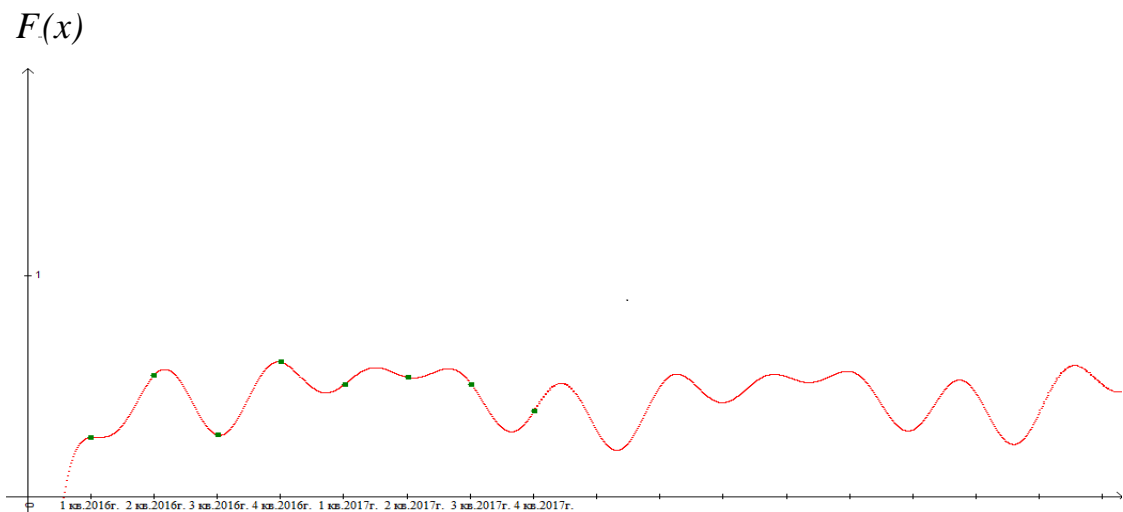
Период	Процент неостановленных несанкционированных операций
1 квартал 2016 г.	0,27
2 квартал 2016 г.	0,55
3 квартал 2016 г.	0,28
4 квартал 2016 г.	0,61
1 квартал 2017 г.	0,51
2 квартал 2017 г.	0,54
3 квартал 2017 г.	0,51
4 квартал 2017 г.	0,39

В рассматриваемом эксперименте для построения прогнозной функции  $F(x)$  использовались данные за период с 1 квартала 2016 года по 4 квартал 2017 года, за нулевое значение по оси ОХ принята дата 4 квартал 2015 года.

Для данного эксперимента аппроксимирующая функция  $F(x)$  имела вид:

$$F(x) = \frac{1.626}{\sqrt{x}} - 0.053 \cdot \sin[x] + 0.075 \cdot \cos[x] - \frac{1.294}{x} + 0.075 \cdot \cos[3 \cdot x] + 0.0005 x^2 - 0.039 \cdot \sin[4 \cdot x] - 0.07 \cdot \cos[4 \cdot x]. \quad (3.2.2.1)$$

График, прогнозной функции  $F(x)$  представлен ниже на рисунке 3.2.2.1.

Рис. 3.2.2.1. График функции  $y = F(x)$ .

На данном графике по оси X расположены даты, начиная с 01.01.2016 г., одно деление соответствует одному кварталу, по оси Y – количество неостановленных несанкционированных операций с использованием платежных карт, поделенное на 100. Точками на графике являются, приведенные для наглядности, значения исходных данных, на основании которых строится сама аппроксимирующая функция.

В таблице 3.2.2.2 приведены результаты эксперимента по построению прогнозных значений количества неостановленных несанкционированных операций с использованием платежных карт на период с 1 квартала 2018 года по 2 квартал 2019 года.

Таблица 3.2.2.2 . Прогноз количества неостановленных несанкционированных операций.

Период	Процент неостановленных несанкционированных операций
1 квартал 2018 г.	0,3
2 квартал 2018 г.	0,5
3 квартал 2018 г.	0,43
4 квартал 2018 г.	0,54
1 квартал 2019 г.	0,57
2 квартал 2019 г.	0,3
3 квартал 2019 г.	0,47
4 квартал 2019 г.	0,39

В соответствии с примененным выше методом построения прогнозной модели динамики негативных событий, связанных с относительным количеством неостановленных несанкционированных операций с использованием платежных карт, функция их числа имеет вид:

$$f(t) = \frac{\alpha_1}{\sqrt{t}} - \alpha_2 \cdot \sin[t] + \alpha_3 \cdot \cos[t] - \frac{\alpha_4}{t} + \alpha_5 \cdot \cos[3 \cdot t] + \alpha_6 \cdot t^2 - \alpha_7 \cdot \sin[4 \cdot t] - \alpha_8 \cdot \cos[4 \cdot t] \quad (3.2.2.2)$$

где  $\alpha_i$ ,  $i=1,2,\dots,8$  - соответствующие константы.

При  $t \geq 1$  для функции  $f(t)$  имеет место неравенство:

$$f(t) \leq \alpha_9 + \alpha_6 \cdot t^2, \quad (3.2.3.3)$$

где  $\alpha_9 = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + \alpha_7 + \alpha_8$ .

Обозначим через  $N$  - количество несанкционированных операций с использованием платежных карт, а через  $U$  – средний ущерб, наносимый одной транзакции.

Тогда прогнозная модель среднего ущерба в платежной системе за исследуемый промежуток времени имеет вид:

$$U \cdot N \cdot f(t) \leq U \cdot N \cdot (\alpha_9 + \alpha_6 \cdot t^2), \quad (3.2.2.4)$$

При этом оценка времени  $T_0$  в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях может быть найдена как корень уравнения относительно неизвестного аргумента  $t$ :

$$U \cdot N \cdot (\alpha_9 + \alpha_6 \cdot t^2) = R_0, \quad (3.2.2.5)$$

откуда получаем:

$$T_0 \geq \sqrt{\left( \frac{R_0/U \cdot N - \alpha_9}{\alpha_6} \right)}, \quad (3.2.2.6)$$

где  $R_0$  - максимально допустимый ущерб программной системе.

В таблице 3.2.2.3 представлены примеры расчетов параметра  $T_0$  - времени (в годах) в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях, примеры рассмотрены для  $N=1000$  несанкционированных операций с целью использования платежных карт.

Таблица 3.2.2.3. Время  $T_0$  (в годах)

$R_0$ (млн.руб)	20	30	40	50	60	70	80	90	100
$U$ (тыс. руб).									
1	0,5	0,62	0,73	0,82	0,91	1	1,06	1,33	1,2

### § 3.3. Прогнозирование динамики появления веб-угроз.

Для прогнозирования динамики появления веб-угроз [37] воспользуемся данными об их появлении, размещенными на сайте Лаборатории Касперского [113].

В таблице 3.3.1 представлена статистика веб-угроз программным системам за февраль-март 2018 года.

Таблица 3.3.1 Веб-угрозы за период с 13.02.2018 г по 12.03.2018г.

Тип веб-угрозы	Процент от общего числа угроз
Trojan.Script.Generic	58,8
Trojan.Script.Agent.gen	20,6
Trojan.JS.Miner.m	8,0
Trojan-Clicker.HTML.Iframe.dg	2,4
Trojan.Win32.Miner.ays	1,5
Trojan.Win64.Shelma.a	0,9
Trojan.JS.Agent.eak	0,9
Trojan.JS.Miner.o	0,4
Trojan-Dropper.VBS.Agent.bp	0,4
Packed.Multi.MultiPacked.gen	0,3

На рисунке 3.3.1 представлена динамика абсолютного числа установленных веб-угроз за февраль-март 2018 года.

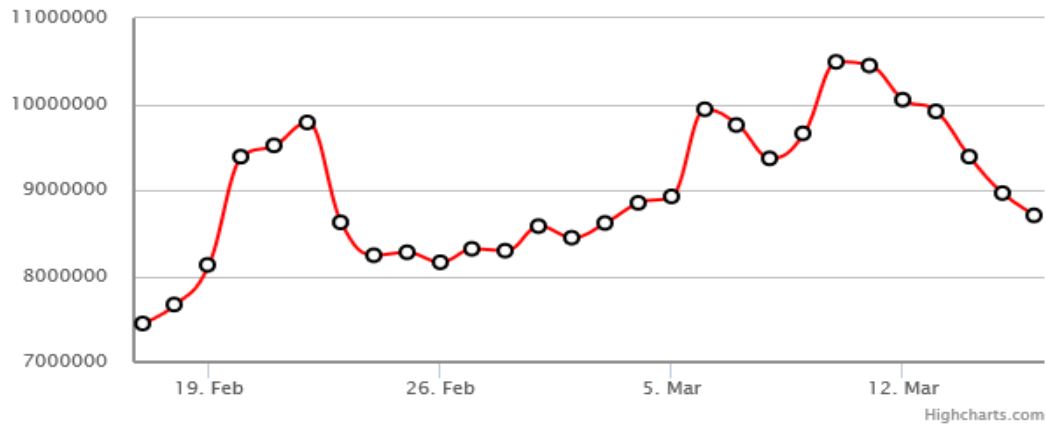


Рис. 3.3.1. График динамики абсолютного числа веб-угроз за период с 13.02.2018 года по 12.03.2018 года.

В эксперименте по прогнозированию веб-угроз при построении прогнозной функции  $F(x)$  использовались данные с 13.02.2018 года по 12.03.2018 года, для удобства представления исходные данные поделены на 100, за нулевое значение по оси  $Ox$  принята дата 01.02.2018 года. Для данного эксперимента аппроксимирующая функция  $F(x)$  имеет вид:

$$F(x) = 46007.45 + \frac{3460901.55}{x^2} + 1336.59 \cdot x - 5589.69 \sin\left[\frac{x}{2}\right] \quad (3.3.1)$$

График прогнозной функции  $y = F(x)$  представлен ниже на рисунке 3.3.2.

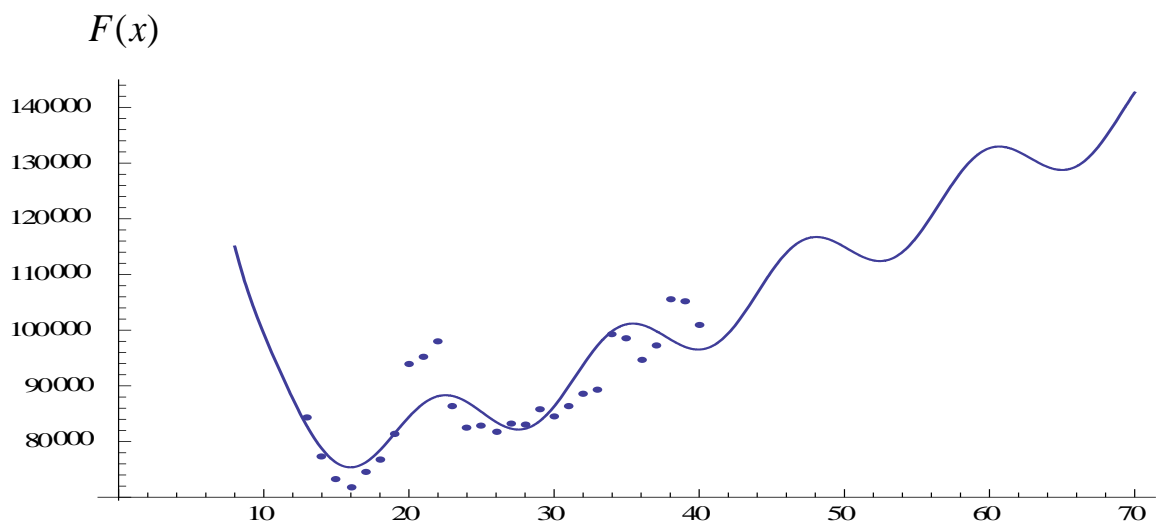


Рис. 3.3.2. График функции  $F(x)$

На данном графике по оси X расположены даты, начиная с 01.02.2018 г., одно деление соответствует двум дням, по оси Y – количество веб-угроз, поделенное на 100. Точками на графике являются, приведенные для наглядности, значения исходных данных, на основании которых строится сама аппроксимирующая функция.

В таблице 3.3.2 приведены результаты эксперимента по построению прогнозных значений веб-угроз на период с 13.03.2018 г. по 11.04.2018 года.

Таблица 3.3.2 Прогноз веб-угроз на период с 13.03.2018 г. по 11.04.2018 г.

Период	Прогнозное значение	Период	Прогнозное значение	Период	Прогнозное значение
13.03.2018	9729460	23.03.2018	11349700	02.04.2018	13290300
14.03.2018	9942970	24.03.2018	14252800	03.04.2018	13203500
15.03.2018	10271600	25.03.2018	11260500	04.04.2018	13061563
16.03.2018	10665540	26.03.2018	11402500	05.04.2018	12931200
17.03.2018	11058645	27.03.2018	11675600	06.04.2018	12876500
18.03.2018	11385640	28.03.2018	12044656	07.04.2018	12942800
19.03.2018	11597300	29.03.2018	12450900	08.04.2018	13146100
20.03.2018	11673800	30.03.2018	12826800	09.04.2018	13468700
21.03.2018	11624700	31.03.2018	13112100	10.04.2018	13863800
22.03.2018	11496100	01.04.2018	13268700	11.04.2018	14566900

Применим полученные данные для оценки времени время в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях.

В соответствии с примененным выше методом построения прогнозной модели появления веб-угроз (негативных событий), функция их числа имеет вид:

$$f(t) = \alpha_1 + \frac{\alpha_2}{t^2} + \alpha_3 t + \alpha_4 \sin \frac{t}{2}, \quad (3.3.2)$$

где  $\alpha_1, \alpha_2, \alpha_3, \alpha_4$  - соответствующие константы.

Обозначим через  $N$  общее число программных систем, для которых проводилось наблюдение. Тогда среднее число негативных событий, приходящееся на одну программную систему в момент времени  $t$  можно оценить величиной:

$$\frac{1}{N} f(t) = \frac{1}{N} \left( \alpha_1 + \frac{\alpha_2}{t^2} + \alpha_3 t + \alpha_4 \sin \frac{t}{2} \right) \quad (3.3.3)$$

Будем считать, что величина ущерба при возникновении веб-угроз (негативных событий) одинаковая и равна  $U$ .

Тогда, уравнение, определяющее время в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях:

$$R(t) = \sum_{i=1}^m p_i(t) \cdot u_i(t) = R_0 \quad (3.3.4)$$

принимает вид:

$$\frac{1}{N} f(t) \cdot U = R_0, \quad (3.3.5)$$

или

$$\frac{1}{N} \cdot \left( \alpha_1 + \frac{\alpha_2}{t^2} + \alpha_3 \cdot t + \alpha_4 \cdot \sin \frac{t}{2} \right) \cdot U = R_0 \quad . \quad (3.3.6)$$

Заметим далее, что для функции  $f(t)$ , при  $t \geq 1$ , имеет место неравенство:

$$f(t) \leq \alpha_5 + \alpha_3 t,$$

где  $\alpha_5 = \alpha_1 + \alpha_2 + \alpha_4$ .

Тогда, очевидно,  $T_0$  - время в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях может быть найдено из уравнения:

$$\frac{1}{N} (\alpha_5 + \alpha_3 \cdot t) \cdot U = R_0, \quad (3.3.7)$$



откуда получаем выражение для величины  $T_0$ :

$$T_0 = \frac{NR_0/U - \alpha_5}{\alpha_3},$$

где  $R_0$  - максимально допустимый ущерб программной системе.

В таблицах 3.3.3, 3.3.4 представлены примеры расчеты параметра  $T_0$  - времени (в годах) в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях в зависимости от величины ущерба от веб-угроз (негативного события)  $U$  и границы максимально допустимых потерь  $R_0$ .

В таблице 3.3.3 представлены расчеты  $T_0$  - время в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях, пример рассмотрен для  $N=1000$ .

Таблица 3.3.3. Время  $T_0$  (в годах)

$R$ (млн. руб) \ $U$ (тыс.руб)						
	40	50	60	70	80	85
1	1	3,05	5,1	7,15	9,2	10,2

В таблице 3.3.4 представлены расчеты  $T_0$  - время в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях пример рассмотрен для  $N=10000$ .

Таблица 3.3.4. Время  $T_0$  (в годах)

$R_0$ (млн. руб) \ $U$ (тыс.руб)						
	4,0	5,0	6,0	7,0	8,0	8,5
10	1	3,05	5,1	7,15	9,2	10,2

### § 3.4. Прогнозирование появления уязвимостей программных систем.

В следующем эксперименте прогноз строился для изучения динамики появления уязвимостей в программных системах, приводящих к нарушению эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях [29]. В таблице 3.4.1 приведена статистика уязвимостей (негативных событий), размещенная на сайте Лаборатории Касперского [113].

Таблица 3.4.1. Уязвимости за период с 13.02.2018 г по 12.03.2018 г.

Тип уязвимости	Процент от общего числа уязвимостей
Exploit.Win32.ShadowBrokers.a	10,0
Exploit.Win32.ShadowBrokers.z	2,6
Exploit.Win64.ShadowBrokers.c	6,1
Exploit.Win32.ShadowBrokers.ab	6,0
Exploit.Win32.ShadowBrokers.aa	5,9
Exploit.Win64.ShadowBrokers.d	5,9
Exploit.Win32.ShadowBrokers.ad	5,9
Exploit.Java.Generic	5,8
Exploit.Script.Generic	4,5
Exploit.Script.Blocker	3,5

На рисунке 3.4.1 представлена динамика абсолютного числа зафиксированных негативных событий, связанных с уязвимостями программных систем.

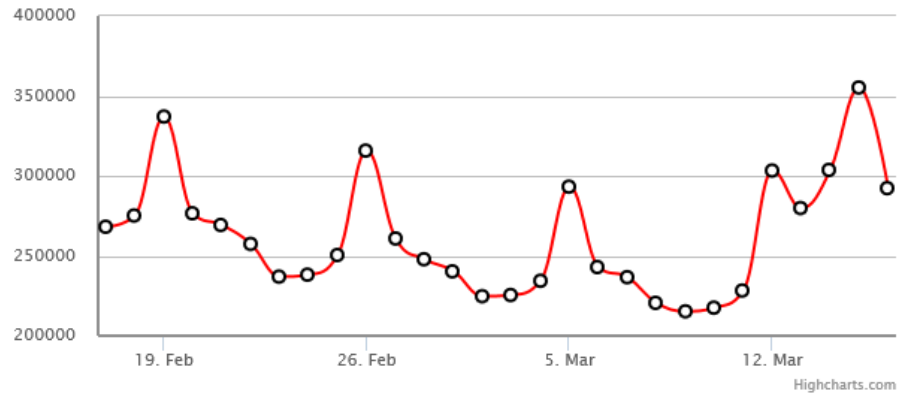


Рис. 3.4.1. График динамики абсолютного числа негативных событий, связанных с уязвимостями программных систем за период с 13.02.2018 г. по 12.03.2018 г.

В эксперименте для построения прогнозной функции  $F(x)$  использовались данные за период с 13.02.2018 года по 12.03.2018 года, за нулевое значение по оси ОХ принята дата 01.02.2018 года.

Для данного эксперимента аппроксимирующая функция  $F(x)$  имела вид:

$$F(x) = 2463152.87 - \frac{6.9581 \times 10^7}{x^2} + 19435.06 \cdot x - 801271.89 \ln[x] + 20916.82 \sin[x]$$

График функции  $y = F(x)$  представлен ниже на рис. 3.4.2.

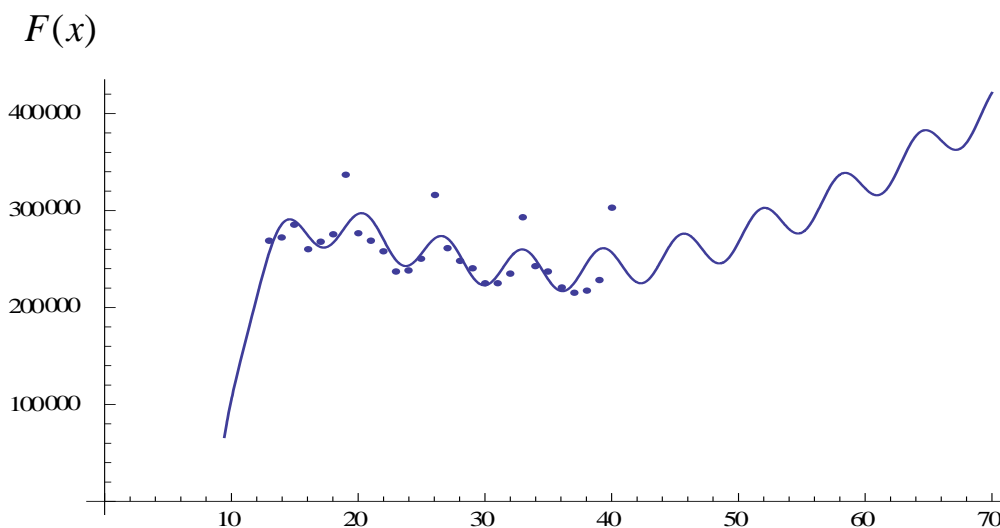


Рис. 3.4.2. График функции  $y = F(x)$ .

На данном графике по оси X расположены даты, начиная с 13.02.2018 г., одно деление соответствует одному дню, по оси Y – число негативных событий, связанных с уязвимостями программных систем. Точками на графике являются, приведенные для наглядности, значения исходных данных, на основании которых строится сама аппроксимирующая функция.

В таблице 3.4.2 приведены результаты эксперимента по построению прогнозных значений появления числа негативных событий, связанных с уязвимостями программных систем на период с 13.03.2018 года по 11.04.2018 года.

Таблица 3.4.2. Прогноз появления числа негативных событий, связанных с уязвимостями программных систем на период с 13.03.2018 г. по 11.04.2018 г.

Период	Прогнозное значение	Период	Прогнозное значение	Период	Прогнозное значение
13.03.2018	239699	23.03.2018	291147	02.04.2018	315856
14.03.2018	225920	24.03.2018	302660	03.04.2018	327607
15.03.2018	230086	25.03.2018	295438	04.04.2018	353754
16.03.2018	250560	26.03.2018	280835	05.04.2018	376857
17.03.2018	270996	27.03.2018	277204	06.04.2018	382438
18.03.2018	275362	28.03.2018	293018	07.04.2018	372285
19.03.2018	262671	29.03.2018	319075	08.04.2018	362805
20.03.2018	247881	30.03.2018	336951	09.04.2018	369934
21.03.2018	248135	31.03.2018	335935	10.04.2018	394486
22.03.2018	266991	01.04.2018	322870	11.04.2018	421394

В соответствии с рассматриваемым выше методом построения прогнозной модели появления числа негативных событий, связанных с уязвимостями программных систем, функция их числа имеет вид:

$$f(t) = \alpha_1 + \frac{\alpha_2}{t^2} + \alpha_3 \cdot t + \alpha_4 \cdot \sin t + \alpha_5 \cdot \ln[t] \quad (3.4.1)$$

где  $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$  - соответствующие константы.

Обозначим через  $N$  общее число программных систем, для которых проводилось наблюдение. Тогда среднее число негативных событий, приходящееся на одну программную систему в момент времени  $t$  можно оценить величиной:

$$\frac{1}{N} f(t) = \frac{1}{N} \left( \alpha_1 + \frac{\alpha_2}{t^2} + \alpha_3 \cdot t + \alpha_4 \cdot \sin t + \alpha_5 \cdot \ln [t] \right)$$

Будем считать, что величина ущерба при возникновении негативных событий одинаковая и равна  $U$ .

Тогда, уравнение (1.1.1) принимает вид:

$$\frac{1}{N} f(t) \cdot U - R_0 = 0, \quad (3.4.2)$$

или

$$\frac{1}{N} \left( \alpha_1 + \frac{\alpha_2}{t^2} + \alpha_3 \cdot t + \alpha_4 \cdot \sin t + \alpha_5 \cdot \ln [t] \right) \cdot U = R_0 \quad . \quad (3.4.3)$$

Заметим далее, что для функции  $f(t)$  при  $t \geq 1$  имеет место неравенство:

$$f(t) \leq \alpha_6 + \alpha_3 \cdot t + \alpha_5 \cdot t,$$

где  $\alpha_6 = \alpha_1 + \alpha_2 + \alpha_4$ .

Тогда, очевидно,  $T_0$  - время в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях может быть найдено из уравнения:

$$\frac{1}{N} (\alpha_6 + \alpha_3 \cdot t + \alpha_5 \cdot t) \cdot U = R_0, \quad (3.4.4)$$

откуда получаем выражение для величины  $T_0$ :

$$T_0 = \frac{N \cdot R_0 / U - \alpha_6}{\alpha_3 + \alpha_5},$$

где  $R_0$  - максимально допустимый ущерб программной системе.

В таблице 3.4.3. представлены примеры расчетов параметра  $T_0$  - времени (в годах) в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях в зависимости от величины ущерба от негативного события  $U$  и границы максимально допустимых потерь  $R_0$  для  $N=10000000$ .

Таблица 3.4.3. Время  $T_0$  (в годах)

$R_0$ (тыс.руб) \ $U$ (тыс.руб)	100	200	300	400	500	600	700
10	0,1	0,44	0,77	1,1	1,43	1,77	2,1

### § 3.5. Прогнозирование сетевых активных событий

В настоящем параграфе, на основании информации о сетевых активных событиях, строится прогнозная функция интенсивности указанных событий и оценивается время в течение, которого обеспечивается эффективность процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях [29].

Данные по сетевым активным событиям взяты на период с 20.10.2017 года по 19.11.2017 года, нормирующий коэффициент 100, за нулевое значение по оси ОХ принята дата 01.10.2017 года. Численные значения количества сетевых активных событий приведены в таблице 3.5.1.

Таблица 3.5.1. Данные по количеству сетевых активных событий

№	Дата	Количество событий/100	№	Дата	Количество событий/100
1	20.10.2017	91377,25	16	05.11.2017	92813,82
2	21.10.2017	85131,41	17	06.11.2017	93711,32
3	22.10.2017	87141,84	18	07.11.2017	95256,71
4	23.10.2017	92470,24	19	08.11.2017	86375,12
5	24.10.2017	83229,57	20	09.11.2017	84278,18
6	25.10.2017	79377,41	21	10.11.2017	89004,67

№	Дата	Количество событий/100	№	Дата	Количество событий/100
7	26.10.2017	83686,1	22	11.11.2017	77437,1
8	27.10.2017	84575,85	23	12.11.2017	76052,22
9	28.10.2017	81498,51	24	13.11.2017	83234,03
10	29.10.2017	84483,92	25	14.11.2017	91015,43
11	30.10.2017	86028,51	26	15.11.2017	90821,68
12	01.11.2017	87110,32	27	16.11.2017	93739,14
13	02.11.2017	88947,97	28	17.11.2017	93207,71
14	03.11.2017	95292,49	29	18.11.2017	85580,87
15	04.11.2017	98674,87	30	19.11.2017	82027,35

Аппроксимирующая функция, построенная по данным таблицы 3.5.1, имеет вид:

$$F(x) = 86618.3117 - \frac{640598.1148}{x^2} + 48.1926 \cdot x + 1694.8417 \cdot \sin[x] \quad (3.5.1)$$

График функции  $F(x)$  представлен ниже на рис. 3.5.1.

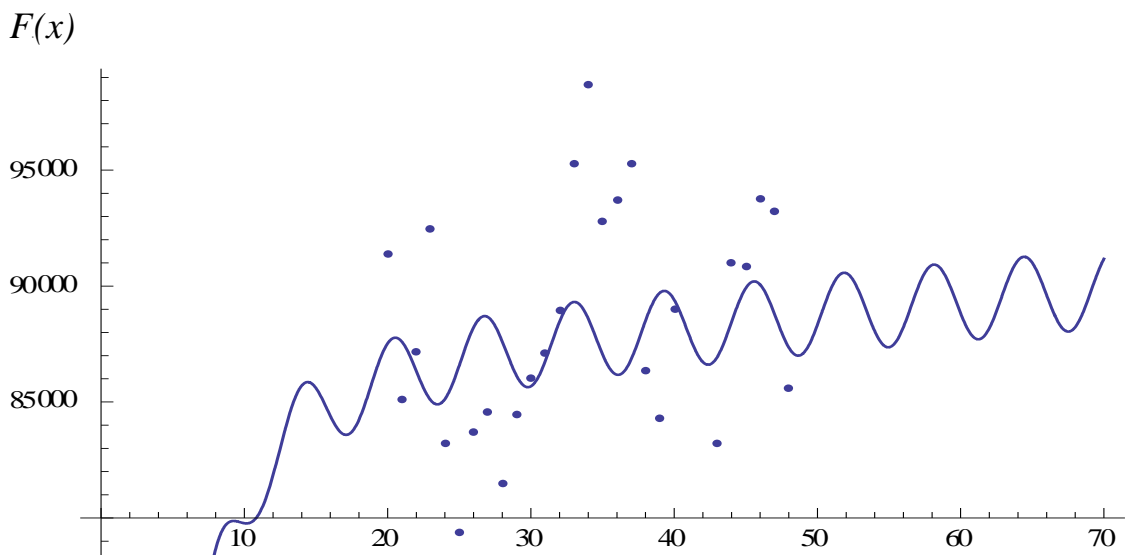


Рис. 3.5.1. График аппроксимирующей функции  $F(x)$  - прогноз сетевых активных событий.

На данном графике по оси X расположены даты, начиная с 20.10.2017 г., одно деление соответствует одному дню, по оси Y – количество сетевых

активных событий, поделенное на 100. Точками на графике являются, приведенные для наглядности, значения исходных данных, на основании которых строится сама аппроксимирующая функция.

В таблице 3.5.2 приведены результаты эксперимента по построению прогнозных значений сетевых активных событий на период с 20.11.2017 года по 19.12.2017 года.

Таблица 3.5.2. Прогноз сетевых активных событий на период с 20.11.2017 г. по 19.12.2017

Период	Прогнозное значение	Период	Прогнозное значение	Период	Прогнозное значение
20.11.2017	7945740	30.11.2017	9035680	10.12.2017	8961450
21.11.2017	8832700	01.12.2017	8881530	11.12.2017	9117270
22.11.2017	8996580	02.12.2017	8774850	12.12.2017	9152480
23.11.2017	9055960	03.12.2017	8818680	13.12.2017	9039480
24.11.2017	8961550	04.12.2017	8977670	14.12.2017	8886910
25.11.2017	8805400	05.12.2017	9110550	15.12.2017	8839790
26.11.2017	8736270	06.12.2017	9100060	16.12.2017	8946170
27.11.2017	8822890	07.12.2017	8960700	17.12.2017	9112950
28.11.2017	8990740	08.12.2017	8825450	18.12.2017	9191510
29.11.2017	9090580	09.12.2017	8823500	19.04.2018	9114320

В соответствии с примененным выше методом построения прогнозной модели появления сетевых активных событий, функция их числа имеет вид:

$$f(t) = \alpha_1 + \frac{\alpha_2}{t^2} + \alpha_3 \cdot t + \alpha_4 \sin t, \quad (3.5.2)$$

где  $\alpha_1, \alpha_2, \alpha_3, \alpha_4$  - соответствующие константы.

Обозначим через  $N$  общее число программных систем, для которых проводилось наблюдение. Тогда среднее число негативных событий, входя-



щееся на одну программную систему в момент времени  $t$  можно оценить величиной

$$\frac{1}{N} f(t) = \frac{1}{N} \left( \alpha_1 + \frac{\alpha_2}{t^2} + \alpha_3 \cdot t + \alpha_4 \cdot \sin t \right)$$

Будем считать, что величина ущерба при возникновении негативных событий одинаковая и равна  $U$ .

Тогда, уравнение (1.1.1) принимает вид:

$$\frac{1}{N} f(t) \cdot U = R_0, \quad (3.5.3)$$

или

$$\frac{1}{N} \cdot \left( \alpha_1 + \frac{\alpha_2}{t^2} + \alpha_3 \cdot t + \alpha_4 \cdot \sin t \right) \cdot U = R_0 \quad (3.5.4)$$

Заметим далее, что для функции  $f(t)$  при  $t \geq 1$  имеет место неравенство:

$$f(t) \leq \alpha_5 + \alpha_3 \cdot t,$$

где  $\alpha_5 = \alpha_1 + \alpha_2 + \alpha_4$ .

Тогда, очевидно, оценка величины  $T_0$  - времени в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях может быть найдена из уравнения:

$$\frac{1}{N} (\alpha_5 + \alpha_3 \cdot t) \cdot U = R_0, \quad (3.5.5)$$

откуда получаем выражение для величины  $T_0$ :

$$T_0 = \frac{NR_0/U - \alpha_5}{\alpha_3},$$

где  $R_0$  - максимально допустимый ущерб программной системе.

В таблицах 3.5.3, 3.5.4 представлены примеры расчета параметра  $T_0$  - времени (в годах) в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях ИС в зависимости от величины ущерба от негативного события  $U$  и

границы максимально допустимых потерь  $R_0$ . Пример рассмотрен для  $N=10000000$ .

Таблица 3.5.3. Время  $T_0$  (в годах)

$R_{(тыс. руб)}$						
$U_{(тыс.руб)}$	146	150	160	170	180	190
20	0,06	1,2	4,04	6,88	9,73	12,57

Таблица 3.5.3. Время  $T_0$  (в годах)

$R_{(тыс. руб)}$						
$U_{(тыс.руб)}$	220	230	240	250	260	270
30	0,25	2,15	4,04	5,94	7,83	9,73

### § 3.6. Прогнозирование интенсивности компьютерных активных событий с целью майнинга криптовалют.

В настоящем параграфе исследуются компьютерные активные события, направленные на перехват управления и использования вычислительных ресурсов данной программной системы для майнинга криптовалют [29]. Как и ранее, на основании имеющихся данных строится прогнозная функция интенсивности данных компьютерных активных событий с целью майнинга криптовалют, задается функция ущерба и оценивается время в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях.

С появлением цифровых валют, требующих для выработки (майнинга) новых единиц денежных знаков, значительных вычислительных ресурсов, появился новый вид негативных событий - несанкционированный доступ к компьютерам пользователей – перехват управления и использования вычислительных мощностей компьютеров для реализации указанной процедуры майнинга криптовалют.

В данном эксперименте прогнозная модель строилась с целью изучения динамики интенсивности компьютерных активных событий осуществляемы-

ми вредоносными программами для майнинга криптовалют. В таблице 3.6.1 приведена статистика негативных событий за период с августа 2017 года по апрель 2018 года, представленная в статье «Ландшафт угроз для систем промышленной автоматизации, Первое полугодие 2018», Kaspersky Lab ICS CERT [113].

Таблица 3.6.1. Процент негативных событий.

Период	Процент негативных событий
Август 2017	0,9
Сентябрь 2017	1,3
Октябрь 2017	2,1
Ноябрь 2017	1,1
Декабрь 2017	1,6
Январь 2018	1,7
Февраль 2018	1,1
Март 2018	1,6
Апрель 2018	3,5

В этом эксперименте для построения прогнозной функции  $F(x)$  использовались данные за период с августа 2017 года по апрель 2018 года, за нулевое значение по оси ОХ принята дата июль 2017 года.

Аппроксимирующая функция, построенная по данным таблицы 3.6.1, имеет вид:

$$F(x) = -0.02 \cdot \sin[x] - 0.74 \cdot \ln[x] - 0.36 \cdot \cos[x] - 0.58 \cdot \sin[2 \cdot x] - 0.01 \cdot \cos[4 \cdot x] + 1.22\sqrt{x} - 0.45 \cdot \sin[4 \cdot x] - 0.05 \cdot \cos[2 \cdot x] + 0.34 \cdot \sin[3 \cdot x].$$

График, прогнозной функции  $F(x)$  представлен ниже на рисунке 3.6.1. На данном графике по оси Х расположены даты, начиная с 01.08.2017 г., одно деление соответствует одному месяцу, по оси Y – количество компьютерных активных событий осуществляемых вредоносными программами для

майнинга криптовалют. Точками на графике являются, приведенные для наглядности, значения исходных данных, на основании которых строится сама аппроксимирующая функция.

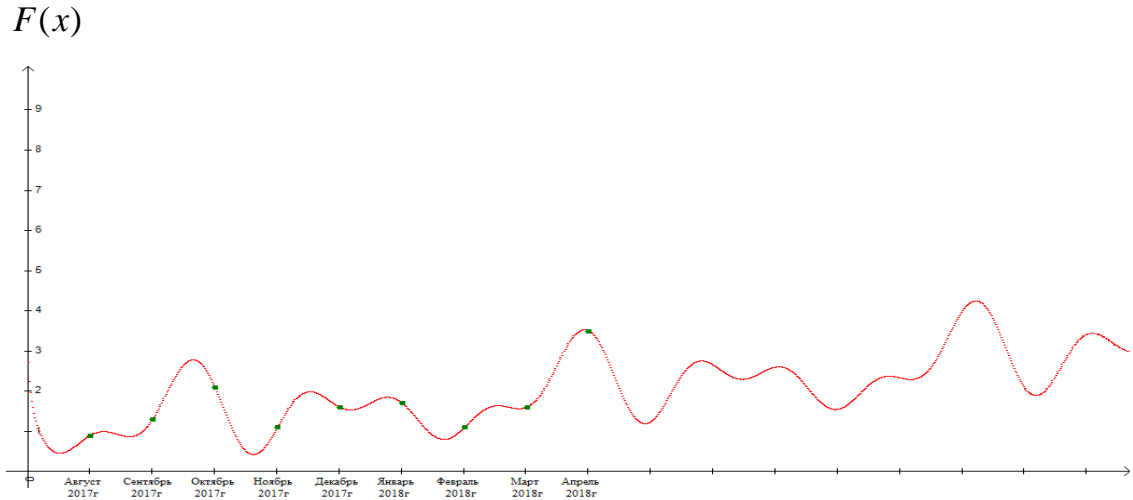


Рис. 3.6.1. График функции  $y = F(x)$ .

В таблице 3.6.2 приведены результаты эксперимента по построению прогнозных значений интенсивности компьютерных активных событий осуществляемыми вредоносными программами для майнинга криптовалют на период с июня 2018 года по январь 2019 года.

Таблица 3.6.2. Прогноз интенсивности активных событий на период с июня 2018 г. по январь 2019 г.

Период	Процент активных событий
Июнь 2018	2,7
Июль 2018	2,6
Август 2018	1,6
Сентябрь 2018	2,3
Октябрь 2018	3,9
Ноябрь 2018	2,1
Декабрь 2018	3,4
Январь 2019	3,1

В данном примере, в соответствии с примененным выше методом построения прогнозной модели интенсивности компьютерных активных событий осуществляемыми вредоносными программами для майнинга криптовалют, функция их интенсивности (процент негативных событий)  $f(t)$  в зависимости от времени  $t$  имеет вид:

$$f(t) = -\alpha_1 \cdot \sin[t] - \alpha_2 \cdot \cos[t] - \alpha_3 \cdot \sin[2 \cdot t] - \alpha_4 \cdot \cos[4 \cdot t] - \alpha_5 \cdot \ln[t] + \alpha_6 \sqrt{t} - \alpha_7 \cdot \sin[4 \cdot t] - \alpha_8 \cdot \cos[2 \cdot t] + \alpha_9 \cdot \sin[3 \cdot t], \quad (3.6.1)$$

где  $\alpha_i$ ,  $i = 1, 2, \dots, 9$  - соответствующие константы.

При  $t \geq 1$  для функции  $f(t)$  имеет место неравенство:

$$f(t) \leq \alpha_{10} + \alpha_6 \sqrt{t}, \quad (3.6.2)$$

где  $\alpha_{10} = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + \alpha_7 + \alpha_8 + \alpha_9$ .

Обозначим через  $U$  - средний ущерб, наносимый одному компьютеру (программной системе).

Тогда прогнозная модель среднего ущерба программной системе имеет вид:

$$U \cdot f(t) \leq U \cdot (\alpha_{10} + \alpha_6 \sqrt{t}), \quad (3.6.3)$$

При этом оценка для  $T_0$  - времени в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях может быть найдена из уравнения:

$$U \cdot (\alpha_{10} + \alpha_6 \sqrt{t}) = R_0, \quad (3.6.4)$$

откуда получаем:

$$T_0 \geq \left( \frac{R_0/U - \alpha_{10}}{\alpha_6} \right)^2, \quad (3.6.5)$$

где  $R_0$  - максимально допустимый ущерб программной системе.

В таблицах 3.6.3 – 3.6.5 представлены примеры расчетов параметра  $T_0$  - времени (в годах) в течение, которого обеспечивается эффективность про-

цесса обработки в вычислительных машинах, комплексах и компьютерных сетях в зависимости от величины ущерба от негативного события  $U$  и границы максимально допустимых потерь  $R_0$ .

Таблица 3.6.3. Время  $T_0$  (в годах)

$R_0$ (тыс.руб) \ $U$ (тыс.руб)							
	50	60	70	80	90	100	110
10	1	2	3,3	5	7	9,3	12

Таблица 3.6.4. Время  $T_0$  (в годах)

$R_0$ (тыс.руб) \ $U$ (тыс.руб)							
	100	120	130	150	170	180	200
20	1	2	2,6	4,1	6	7	9,3

Таблица 3.6.5. Время  $T_0$  (в годах)

$R_0$ (тыс.руб) \ $U$ (тыс.руб)							
	100	120	130	150	170	180	200
30	0,1	0,35	0,5	1	1,6	2	2,9

### § 3.7. Прогнозирование негативных событий, связанных с ущербом от хищений со счетов юридических лиц с использованием платежных карт.

В следующем эксперименте прогнозирование проводилось для изучения динамики негативных событий связанных с ущербом от проведенных несанкционированных операций (хищений) со счетов юридических лиц с использованием платежных карт [28]. В таблице 3.7.1. приведена статистика объема несанкционированных операций за период с 1 квартала 2016 года по 4 квартал 2017 года. Данные взяты из официального сайта Банка России представленная в [60].

Таблица 3.7.1. Объем несанкционированных операций (хищений)

Период	Объем несанкционированных операций (хищений, млн. руб.)
1 квартал 2016 г.	352,6
2 квартал 2016 г.	342,5
3 квартал 2016 г.	428,7
4 квартал 2016 г.	770,6
1 квартал 2017 г.	405,6
2 квартал 2017 г.	440,3
3 квартал 2017 г.	343,7
4 квартал 2017 г.	380,0

В рассматриваемом эксперименте для построения прогнозной функции  $F(x)$  использовались данные за период с 1 квартала 2016 года по 4 квартал 2017 года, за нулевое значение по оси ОХ принята дата 4 квартал 2015 года.

Для данного эксперимента «аппроксимирующая» функция  $F(x)$  имела вид:

$$F(x) = 510.34 + 26.72\sqrt{x} + 167.25 \cdot \sin[2 \cdot x] - 6.05 \cdot \cos[2 \cdot x] + 75.92 \cdot \cos[4 \cdot x] + 100.62 \cdot \cos[3 \cdot x] - 125.08 \cdot \sin[x] - 58.22 \cdot \cos[x].$$

График прогнозной функции  $F(x)$  представлен ниже на рисунке 3.7.1.

$F(x)$

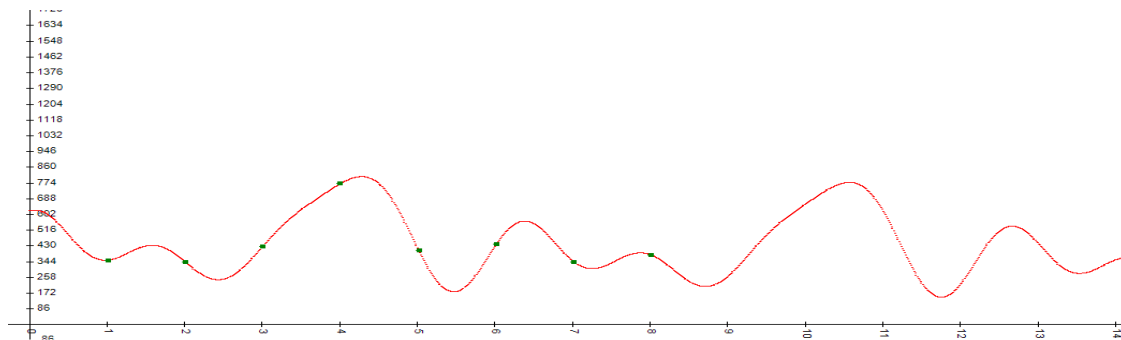


Рис. 3.7.1. График «аппроксимирующей» функции  $y = F(x)$ .

На данном графике по оси X расположены даты, начиная с 01.01.2016 г., одно деление соответствует одному кварталу, по оси Y – количество (объем несанкционированных операций). Точками на графике являются, приведенные для наглядности, значения исходных данных, на основании которых строится сама аппроксимирующая функция.

В таблице 3.7.2 приведены результаты эксперимента по построению прогнозных значений негативных событий связанных с ущербом от несанкционированных операций (хищений) со счетов юридических с использованием платежных карт на период с 1 квартала 2018 года по 2 квартал 2019 года.

Таблица 3.7.2. Прогноз объема несанкционированных операций.

Период	Объем несанкционированных операций (млн. руб.)
1 квартал 2018 г.	383
2 квартал 2018 г.	262
3 квартал 2018 г.	657
4 квартал 2018 г.	625
1 квартал 2019 г.	220
2 квартал 2019 г.	446

В соответствии с описанным выше методом, функция числа негативных событий связанных с ущербом от несанкционированных операций со счетов юридических лиц с использованием платежных карт  $f(t)$  в зависимости от времени  $t$  имеет вид:

$$f(t) = \alpha_1 + \alpha_2 \sqrt{t} + \alpha_3 \cdot \sin[2 \cdot t] - \alpha_4 \cdot \cos[2 \cdot t] + \alpha_5 \cdot \cos[4 \cdot t] + \alpha_6 \cdot \cos[3 \cdot t] - \alpha_7 \cdot \sin[t] - \alpha_8 \cdot \cos[t], \quad (3.7.1)$$

где  $\alpha_i$ ,  $i=1,2,\dots,8$  - соответствующие константы.

При  $t \geq 0$  для функции  $f(t)$  имеет место очевидное неравенство:

$$f(t) \leq \alpha_9 + \alpha_2 \sqrt{t}, \quad (3.7.2)$$

где  $\alpha_9 = \alpha_1 + \alpha_3 + \dots + \alpha_8$ .



Обозначим через  $N$  - число юридических лиц (программных систем), понесших ущерб от действий другой программной системой.

Тогда прогнозная модель среднего ущерба программной системы имеет вид:

$$\frac{1}{N} f(t) \leq \frac{1}{N} (\alpha_9 + \alpha_2 \sqrt{t}), \quad (3.7.3)$$

При этом оценка времени  $T_0$  в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях может быть найдена из уравнения:

$$\frac{1}{N} (\alpha_9 + \alpha_2 \sqrt{t}) = R_0, \quad (3.7.4)$$

откуда получаем оценку для величины  $T_0$ :

$$T_0 \geq \left( \frac{NR_0 - \alpha_9}{\alpha_2} \right)^2, \quad (3.7.5)$$

где  $R_0$  - максимально допустимый ущерб программной системе.

В таблице 3.7.3 представлен пример расчета параметра  $T_0$  - времени (в годах) в течение, которого обеспечивается эффективность процесса обработки в вычислительных машинах, комплексах и компьютерных сетях.

Таблица 3.7.3. Время  $T_0$  (в годах)

$R_0$ (млн.руб.) \ $N$	1100	1150	1200	1250	1300
1	1,1	4	8,6	15	23

### Выводы по 3 главе.

В настоящей главе рассмотрены вопросы экспериментальных исследований по прогнозированию компьютерных событий в вычислительных машинах, комплексах и компьютерных сетях, на основе построения непрерывных аппроксимирующих функций. Приведены результаты прогнозирования активных событий в вычислительных машинах, комплексах и компьютерных сетях; негативных событий связанных с несанкционированными операциями

со счетами юридических лиц с использованием платежных карт; негативных событий связанных с несанкционированными операциями со счетами юридических лиц; динамики появления веб-угроз; динамики появления уязвимостей в программных системах; сетевых активных событий; компьютерных активных событий, направленных на перехват управления и использования вычислительных ресурсов данной программной системы для майнинга криптовалют; динамики негативных событий, связанных с ущербом от проведенных несанкционированных операций со счетами юридических лиц с использованием платежных карт.

Полученные результаты применяются для разработки методов оценки параметра, характеризующего процесс обработки данных в вычислительных машинах, комплексах и компьютерных сетях - время в течение, которого обеспечивается эффективность процесса обработки указанных данных.

## ГЛАВА 4. РАЗРАБОТКА МЕТОДОВ ПРОГНОЗИРОВАНИЯ КОМПЬЮТЕРНЫХ СОБЫТИЙ НА ОСНОВЕ ПОСТРОЕНИЯ НЕПРЕРЫВНЫХ АППРОКСИМИРУЮЩИХ ФУНКЦИЙ

В настоящей главе представлено решение задачи построения методики прогнозирования компьютерных событий, приводящих к нарушению эффективности процесса обработки данных, которые могут быть рассмотрены как динамические процессы, состояния которых заданы в виде таблицы значений. Поставленная задача решается с помощью построения непрерывных аппроксимирующих функций, наиболее близко отстоящих от заданных значений состояний системы. Разработана методика построения аппроксимирующих функций в виде линейной комбинации базовых функций на основе модифицированного метода наименьших квадратов. Дано алгоритмическое описание разработанного инструментария и в рамках определенной теоретико-вероятностной модели построены теоретические оценки точности прогнозирования предлагаемым методом.

### § 4.1. Выбор математического аппарата

В данном параграфе рассматривается постановка задачи построения методики прогнозирования состояния динамического процесса и указаны возможные пути ее решения [31-34].

Задача построения методики прогнозирования состоит в следующем. Пусть на отрезке  $[a, b]$ , задана одномерная сетка:

$$X = \{x_i / x_i = x_{i-1} + h_i, h_i > 0, i = 1, 2, \dots, n; x_0 = a, x_n = b\}, \quad (4.1.1)$$

в узлах  $x_i$  которой заданы значения  $y_i = f(x_i), i = 0, 1, 2, \dots, n$  – соответствующие значения некоторой неизвестной функции  $f(x)$ .

Будем далее рассматривать данную таблицу как таблицу состояний некоторого динамического процесса. При этом величины  $x_i$  означают моменты времени при наблюдении состояния рассматриваемой системы, а величины  $y_i$  – сами эти состояния.

Метод наименьших квадратов состоит в минимизации выражения:

$$\sum_{i=0}^n (y_i - f(x_i))^2 \rightarrow \min .$$

Пусть также для аппроксимации табличных данных выбран некоторый класс функций  $F(x, c_0, c_1, \dots, c_m)$ ,  $m < n$ , где  $c_0, c_1, \dots, c_m$  – коэффициенты, выбор значений которых позволяет определить конкретную функцию из выбранного класса. Требуется найти значения коэффициентов  $c_0, c_1, \dots, c_m$ , для которых выполнено условие:

$$\Phi(c_0, c_1, \dots, c_m) = \sum_{i=0}^n (y_i - F(x_i, c_0, c_1, \dots, c_m))^2 \rightarrow \min . \quad (4.1.2)$$

Выбранные в соответствии с критерием (4.1.2) значения коэффициентов, позволяют определить среди множества функций конкретную функцию, наиболее согласованную с табличными (экспериментальными) данными или, иначе говоря, обеспечивающую наилучшее среднеквадратическое приближение. При этом данная функция должна обладать свойствами минимизации отклонений от статистических данных на интервале  $[b, b + c]$ , то есть, должно осуществляться, в том числе, прогнозирование с помощью экстраполяции.

Функция  $F(x, c_0, c_1, \dots, c_m)$  называется моделью, а искомые коэффициенты  $c_0, c_1, \dots, c_m$  – параметрами модели.

Ниже в данной главе будут рассмотрены примеры применения для построения прогнозных аппроксимирующих функций различных методов: метода Лагранжа построения интерполяционного многочлена, метода наименьших квадратов (Least Square Method) и метода Берга.

*4.1.1. Интерполяционный многочлен Лагранжа и его применение для целей аппроксимации.*

Интерполяционный многочлен Лагранжа [107] — многочлен минимальной степени, принимающий значения  $y_i$  в узловых точках  $x_i$  – состояний динамического процесса.

Для  $n+1$  пар чисел  $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$ , где все значения  $x_j$  различны, существует единственный многочлен  $L(x)$  степени не более  $n$ , для которого  $L(x_j) = y_j$ .

Лагранж предложил искать такие многочлены в виде:

$$L(x) = \sum_{i=0}^n y_i l_i(x)$$

где базисные полиномы  $l_i(x)$  определяются по формуле:

$$l_i(x) = \prod_{j=0, j \neq i}^n \frac{x-x_j}{x_i-x_j} = \frac{x-x_0}{x_i-x_0} \cdots \frac{x-x_{i-1}}{x_i-x_{i-1}} \frac{x-x_{i+1}}{x_i-x_{i+1}} \cdots \frac{x-x_n}{x_i-x_n}$$

а полиномы  $l_i(x)$ , являющиеся многочленами степени  $n$ , обладают следующими свойствами:  $l_i(x_i) = 1$ ,  $l_i(x_j) = 0$ , при  $j \neq i$ .

В работе [107] приведен вывод полинома Лагранжа, имеющий следующий вид:

$$L_n(x) = \sum_{i=0}^n y_i \frac{\prod_{n+1}(x)}{(x-x_i) \prod_{n+1}^i(x_i)},$$

где  $\prod_{n+1}(x) = (x-x_0) \cdot (x-x_1) \cdots (x-x_n)$ ,

$$\prod_{n+1}^i(x) = (x-x_0) \cdot (x-x_1) \cdots (x-x_{i-1})(x-x_{i+1}) \cdots (x-x_n).$$

Очевидным недостатком данного метода является прямая зависимость степени полученного многочлена от количества узлов интерполяции, то есть от количества данных на входе. В этом случае даже для небольшого участка таблично заданной функции, состоящей, например, из 100 точек, степень полинома окажется равной  $100 \rightarrow (n+1) - 1$ . Это достаточно большие числа, оперирование которыми даже у современных компьютеров может привести к выходу за диапазон максимально доступных значений, и, как следствие, к неправильному расчету значения полинома.

Другим существенным недостатком данного метода является то обстоятельство, что за границей исследуемого диапазона данных  $[a, b]$  полином

Лагранжа, практически, во всех случаях, имеет большой разброс значений, и не может быть использован в качестве прогнозной функции.

#### 4.1.2. Метод наименьших квадратов.

При рассмотрении ряда задач возникает необходимость подбора на основе имеющихся экспериментальных данных, представленных в табличной форме, некоторой функциональной зависимости из заданного класса функций, свойства которого наиболее соответствуют предполагаемым физическим свойствам изучаемых процессов или позволяют выявить характерную для этих процессов тенденцию.

С математической точки зрения имеется функция с неизвестным (или известным, но слишком сложным с точки зрения проводимого исследования) аналитическим заданием, представленная в узлах заданной сетки значениями, которые были получены в результате прямых или косвенных измерений, или некоторого вычислительного эксперимента. Такие данные называются экспериментальными. Поскольку по ряду причин (например, в силу большой погрешности) интерполирование такой таблично заданной функции не является целесообразным, ставится задача ее аппроксимации функцией из заданного класса, наиболее согласованной в смысле некоторого критерия с имеющимися экспериментальными данными. Если в качестве критерия согласованности выбирается критерий минимума суммы квадратов отклонений модельных данных от экспериментальных, говорят об аппроксимации таблично заданной функции методом наименьших квадратов (МНК).

В настоящем исследовании мы ограничимся рассмотрением случая, когда искомая модель  $F(x, c_0, c_1, \dots, c_m)$  линейно зависит от параметров  $c_0, c_1, \dots, c_m$  и ее можно представить в виде линейной комбинации:

$$F(x, c_0, c_1, c_2, \dots, c_m) = c_0 \cdot \varphi_0(x) + c_1 \cdot \varphi_1(x) + \dots + c_m \cdot \varphi_m(x). \quad (4.1.2.1)$$

Модель вида (4.1.2.1) часто называют обобщенным полиномом. Здесь  $\varphi_0(x), \varphi_1(x), \dots, \varphi_m(x)$  – множество, так называемых, базисных функций. Базис-

ные функции могут быть как линейными, так и нелинейными функциями переменной  $x$ . Независимо от этого модель (4.1.2.1) остается линейной, поскольку она линейно зависит от модельных параметров  $c_0, c_1, \dots, c_m$ .

В качестве базисных могут быть выбраны, например, степенные функции

$$\varphi_0(x) = 1, \varphi_1(x) = x, \varphi_2(x) = x^2, \dots, \varphi_m(x) = x^m.$$

Тогда модель будет представлять собой полином степени  $m$ :

$$\varphi(x, c_0, c_1, c_2, \dots, c_m) = c_0 + c_1x + c_2x^2 + \dots + c_mx^m.$$

Очевидно, что в качестве базисных могут быть использованы и другие функции, необходимо лишь, чтобы они были линейно независимыми.

Таким образом, для линейной модели (4.1.2.1) требуется найти значения параметров  $c_0, c_1, \dots, c_m$ , обеспечивающих выполнение условия (4.1.2).

С математической точки зрения, поставленная задача является задачей минимизации функции нескольких переменных. Функция  $\Phi(c_0, c_1, \dots, c_m)$  является квадратичной, одно экстремальной, поэтому ее минимум можно искать, исходя из необходимых условий экстремума для функций нескольких переменных.

Для прогнозирования компьютерных событий был выбран именно этот метод и проведены предварительные экспериментальные исследования. Однако результаты экспериментов показали, что без учета статистических характеристик исходных данных, в ряде случаев, возникают значительные ошибки прогнозирования. В связи с этим было принято решение модификации данного метода, учитывающего характеристики начальных данных и реализуемого путем определения соответствующего набора базовых функций. Проведенные в главе 5 экспериментальные исследования точности прогнозирования модифицированным методом показали небольшие значения возникающих ошибок.

В части проведения теоретических исследований в данной главе работы в рамках определенной теоретико-вероятностной модели получены строгие оценки вероятности превышения ошибок заданной величины.

#### 4.1.3. Метод Берга

Метод Берга является одним из методов линейного предсказания, позволяющий построить прогноз значения некоторой последовательности данных на основании ее предыдущих значений, в предположении наличия в рассматриваемой последовательности рекуррентной зависимости [12]. Определение значений коэффициентов данной зависимости осуществляется при помощи МНК. Первым шагом рассматриваемого метода является построение модели предполагаемой рекуррентной зависимости, а вычисление прогнозных значений осуществляется далее на основании построенной рекурренты.

Пусть исходная последовательность состоит из  $N$  значений  $\{x_n\}_{n \in [0, N]}$ . Для построения модели рекуррентной зависимости будем использовать набор из  $k$  коэффициентов  $\{a_k\}_{k \in [1, k]}$ . Обозначим через  $y_n$  и  $z_n$  величины

$$y_n = -\sum_{i=1}^k a_i x_{n-i}, \quad z_n = -\sum_{i=1}^k a_i x_{n+i},$$

называемые передним и обратным линейным предсказанием. Другими словами, предположим, что каждая величина  $y_n$  это линейная комбинация из  $k$  предыдущих известных значений, а  $z_n$  — линейная комбинация из  $k$  последующих значений. При этом величина  $y_n$  определена для значений  $n \in [k, N]$ , а  $z_n$  определена для значений  $n \in [0, N-k]$ .

Стандартный метод определения значений коэффициентов для переднего и обратного линейного предсказания это метод наименьших квадратов:

$$F_k = \sum_{n=k}^N (x_n - y_n)^2 = \sum_{n=k}^N (x_n - (-\sum_{i=1}^k (a_i x_{n-i})))^2$$

$$B_k = \sum_{n=k}^N (x_n - z_n)^2 = \sum_{n=k}^N (x_n - (-\sum_{i=1}^k (a_i x_{n+i})))^2$$



В соответствии с МНК, задача нахождения неизвестных коэффициентов  $\{a_k\}, k \in [1, k]$  сводится к минимизации суммы  $F_k + B_k$ .

Очевидным недостатком данного метода является то, что из-за особенности рекурсии и ограничения на глубину линейной зависимости членов входной последовательности, предсказанные точки с ростом глубины предсказания приобретают вид гармонических затухающих колебаний.

Таким образом, данный метод рекомендуется использовать для предсказания поведения гладких гармонических и затухающих функций.

#### § 4.2. Построение аппроксимирующих функций на основе модифицированного метода наименьших квадратов

В данном разделе, пользуясь условиями экстремума функции нескольких переменных [107], определим метод формирования набора базовых функций  $\varphi_0(x), \varphi_1(x), \dots, \varphi_m(x)$ , для дальнейшего построения на его основе аппроксимирующей функции  $F(x, c_0, c_1, \dots, c_m)$  [31-35].

Как было отмечено выше, функция  $F(x, c_0, c_1, \dots, c_m)$  от переменных  $x, c_0, c_1, \dots, c_m$  ищется в виде линейной комбинации базовых функций:

$$F(x, c_0, c_1, c_2, \dots, c_m) = c_0 \varphi_0(x) + c_1 \varphi_1(x) + c_2 \varphi_2(x) + \dots + c_m \varphi_m(x),$$

где  $c_0, c_1, \dots, c_m$  – неизвестные коэффициенты, выбираемые из условия достижения минимального значения функции  $\Phi(c_0, c_1, \dots, c_m)$ :

$$\Phi(c_0, c_1, \dots, c_m) = \sum_{i=0}^n (y_i - F(x_i, c_0, c_1, \dots, c_m))^2 \rightarrow \min ,$$

то есть, определяются из условия наилучшего приближения функции  $F(x, c_0, c_1, \dots, c_m)$  к неизвестной функции  $y=f(x)$  в узловых точках  $(x_0, y_0) \dots (x_n, y_n)$

В соответствии с [146], экстремум функции многих переменных  $\Phi(c_0, c_1, \dots, c_m)$  определяется условием:

$$\frac{\partial \Phi(c_0, c_1, \dots, c_m)}{\partial c_i} = 0, \quad i = 0, 1, \dots, m.$$

Запишем эти условия в виде системы линейных уравнений и преобразуем ее к виду:

$$\left\{ \begin{array}{l} c_0 \sum_{i=0}^n \varphi_0^2(x_i) + c_1 \sum_{i=0}^n \varphi_0(x_i)\varphi_1(x_i) + c_2 \sum_{i=0}^n \varphi_0(x_i)\varphi_2(x_i) + \dots + c_m \sum_{i=0}^n \varphi_0(x_i)\varphi_m(x_i) = \sum_{i=0}^n y_i \varphi_0(x_i), \\ c_0 \sum_{i=0}^n \varphi_1(x_i)\varphi_0(x_i) + c_1 \sum_{i=0}^n \varphi_1^2(x_i) + c_2 \sum_{i=0}^n \varphi_1(x_i)\varphi_2(x_i) + \dots + c_m \sum_{i=0}^n \varphi_1(x_i)\varphi_m(x_i) = \sum_{i=0}^n y_i \varphi_1(x_i), \\ c_0 \sum_{i=0}^n \varphi_2(x_i)\varphi_0(x_i) + c_1 \sum_{i=0}^n \varphi_2(x_i)\varphi_1(x_i) + c_2 \sum_{i=0}^n \varphi_2^2(x_i) + \dots + c_m \sum_{i=0}^n \varphi_2(x_i)\varphi_m(x_i) = \sum_{i=0}^n y_i \varphi_2(x_i), \\ \dots \\ c_0 \sum_{i=0}^n \varphi_m(x_i)\varphi_0(x_i) + c_1 \sum_{i=0}^n \varphi_m(x_i)\varphi_1(x_i) + c_2 \sum_{i=0}^n \varphi_m(x_i)\varphi_2(x_i) + \dots + c_m \sum_{i=0}^n \varphi_m^2(x_i) = \sum_{i=0}^n y_i \varphi_m(x_i). \end{array} \right. \quad (4.2.1)$$

Матрица системы линейных уравнений (4.2.1) носит название матрицы Грама и имеет вид:

$$G = \begin{bmatrix} \sum_{i=0}^n \varphi_0^2(x_i) & \sum_{i=0}^n \varphi_0(x_i) \cdot \varphi_1(x_i) & \sum_{i=0}^n \varphi_0(x_i) \cdot \varphi_2(x_i) \dots \sum_{i=0}^n \varphi_0(x_i) \cdot \varphi_m(x_i) \\ \sum_{i=0}^n \varphi_1(x_i) \cdot \varphi_0(x_i) & \sum_{i=0}^n \varphi_1^2(x_i) & \sum_{i=0}^n \varphi_1(x_i) \cdot \varphi_2(x_i) \dots \sum_{i=0}^n \varphi_1(x_i) \cdot \varphi_m(x_i) \\ \sum_{i=0}^n \varphi_2(x_i) \cdot \varphi_0(x_i) & \sum_{i=0}^n \varphi_2(x_i) \cdot \varphi_1(x_i) & \sum_{i=0}^n \varphi_2^2(x_i) & \dots \sum_{i=0}^n \varphi_2(x_i) \cdot \varphi_m(x_i) \\ \dots & \dots & \dots & \dots \\ \sum_{i=0}^n \varphi_m(x_i) \cdot \varphi_0(x_i) & \sum_{i=0}^n \varphi_m(x_i) \cdot \varphi_1(x_i) & \sum_{i=0}^n \varphi_m(x_i) \cdot \varphi_2(x_i) \dots \sum_{i=0}^n \varphi_m^2(x_i) \end{bmatrix}$$

Очевидно, что матрица  $G$  симметричная. Можно показать, что она положительно определена. Определитель матрицы  $G$  отличен от нуля в силу линейной независимости базисных функций  $\varphi_0(x), \varphi_1(x), \dots, \varphi_m(x)$ . Следовательно, система (4.2.1) имеет единственное решение, которое может быть найдено любым из известных методов решения такого вида систем. В результате решения системы (4.2.1) будут найдены значения модельных параметров  $(c_0^*, c_1^*, \dots, c_m^*)$  и определена наилучшая в указанном выше смысле модель  $F(x, c_0^*, c_1^*, \dots, c_m^*)$ .

При этом минимум функции  $\Phi(c_0, c_1, \dots, c_m)$  достигается при условии положительной определенности квадратичной формы  $A$ :

$$A = \sum_{i,j} \frac{\partial^2 \Phi(c_1, \dots, c_m)}{\partial c_i \partial c_j} > 0 \quad (4.2.3)$$

Поскольку

$$F(x, c_0, c_1, \dots, c_m) = c_0 \varphi_0(x) + c_1 \varphi_1(x) + c_2 \varphi_2(x) + \dots + c_m \varphi_m(x),$$

то нетрудно видеть, что

$$\frac{d^2 \Phi}{dc_i dc_j} = 2 \sum_{k=0}^n \varphi_i(x_k) \cdot \varphi_j(x_k),$$

где  $x_i$  - значения аргументов узловых точек  $(y_0, y_1, \dots, y_m, x_0, x_1, \dots, x_m)$ .

Тогда, квадратичная форма  $A$  имеет вид:

$$A = 2 \sum_{k=0}^n \sum_{i,j} \varphi_i(x_k) \cdot \varphi_j(x_k).$$

Если  $A > 0$ , то в точке экстремума  $c_0, c_1, \dots, c_m$  функция  $\Phi(c_0, c_1, \dots, c_m)$  достигает минимума, в противном случае необходима замена набора базовых функций.

Как было отмечено выше, прогнозной моделью рассматриваемого динамического процесса является аппроксимирующая функция:

$$F(x) = F(x, c_0, c_1, c_2, \dots, c_m) = c_0 \varphi_0(x) + c_1 \varphi_1(x) + c_2 \varphi_2(x) + \dots + c_m \varphi_m(x),$$

где  $\varphi_0(x), \varphi_1(x), \dots, \varphi_m(x)$  – множество базисных функций,  $c_0, c_1, \dots, c_m$  – коэффициенты найденные путем решения системы линейных уравнений.

Процедура прогнозирования предполагает далее задание множества точек (моментов времени), для которых предполагается построение прогнозных значений:

$$\check{T} = \{t_1, t_2, \dots, t_k\}, \quad \text{где } t_i < t_j, \text{ при } i < j, t_i \in [b, b + c].$$

Процедуру построения прогнозной модели, а именно аппроксимирующей функции  $F(x, c_0, c_1, c_2, \dots, c_m)$  дополним далее следующим условием.

Для каждого значения величины  $t_i$  (момента времени, для которого строится прогноз) будем проверять выполнение неравенства:

$$|F(t_i) - \lambda| \leq \delta \cdot \sigma, \tag{4.2.4}$$

где  $\delta$  – параметр алгоритма построения прогнозной функции,

$\lambda = \frac{1}{n+1} \cdot \sum_{i=0}^n y_i$  - эмпирическое среднее состояний исходного процесса,

$\sigma^2 = \frac{1}{n+1} \cdot \sum_{i=0}^n (y_i - \lambda)^2$  - эмпирическая дисперсия состояний исходного

процесса.

В случае нарушения условия (4.2.4), хотя бы для одного значения величин  $t_i, i=1,2,\dots,n$  производится замена множества базисных функций  $\varphi_0(x), \varphi_1(x), \dots, \varphi_m(x)$  [35].

При переборе множеств базисных функций, удовлетворяющих условию (4.2.4), выбирается то множество, при котором эмпирическое среднее

$\bar{\lambda} = \frac{1}{n+1} \cdot \sum_{i=1}^k F(t_i), t_i \in \check{T}$ , построенное по прогнозным значениям с помощью

аппроксимирующей функции  $F(x)$ , наиболее близко к эмпирическому среднему исходных данных  $\lambda$ .

### **§ 4.3. Теоретическая оценка точности прогнозирования, модифицированного МНК.**

В данном разделе исследуются теоретические вопросы точности прогнозирования в результате применения рассматриваемого модифицированного МНК в рамках определенной теоретико-вероятностной модели [35].

В соответствии с рассматриваемой в настоящей работе исходной задачей, состояние процесса, будущее, которого предстоит спрогнозировать, задается в виде таблично заданной функции  $y = f(x)$ , а именно, набором из  $n+1$  узловых точек  $(x_0, y_0), \dots, (x_n, y_n)$  [35].

Для оценки точности прогнозирования и, соответственно, эффективности рассматриваемой прогнозной модели будем считать, что значение функции  $f(x)$ , описывающей исходное состояние процесса, есть реализация некоторого случайного стационарного процесса  $\zeta(t)$ , то есть, в момент времени  $t_i, (t_i = x_i)$ , последовательность величин:

$$y_i = f(x_i) = \zeta(t_i)$$

есть реализация некоторой случайной величины  $\xi(t)$ , имеющей математическое ожидание:

$$\lambda = M\xi(t) = \frac{1}{n+1} \cdot \sum_{i=0}^n y_i$$

и дисперсию:

$$\sigma^2 = D\xi(t) = \frac{1}{n+1} \cdot \sum_{i=0}^n (y_i - \lambda)^2.$$

В соответствии с предлагаемой процедурой построения прогнозной модели, для получения прогнозных значений рассматриваемого процесса строится аппроксимирующая функция:

$$F(x) = \sum_{i=0}^m c_i \varphi_i(x)$$

в виде линейной комбинации базовых функций  $\varphi_i(x)$ ,  $i=0, 1, \dots, m$ .

При этом в соответствии с алгоритмом построения прогнозной функции  $F(x)$  на интервале  $[b, b+c]$  (см. §4.2), имеет место неравенство:

$$|F(t) - \lambda| \leq \delta \sigma. \quad (4.3.1)$$

С целью оценки устойчивости предлагаемой модели и ошибок прогнозирования, в условиях введённой теоретико-вероятностной модели оценим вероятность отклонения прогнозных значений рассматриваемого процесса от ее действительных значений:

$$P\left\{|\xi(t) - F(t)| \leq \varepsilon\right\}, \quad (4.3.2)$$

где  $\varepsilon > 0$  - некоторая фиксированная величина,  $t \in [b, b+c]$ .

$$F(x) = \sum_{i=0}^m c_i \cdot \varphi_i(x).$$

Оценку точности прогнозирования предлагаемой процедуры МНК даёт следующее утверждение.

*Утверждение 4.3.1.* В условиях введённой теоретико-вероятностной модели при фиксированных значениях величин  $\varepsilon$ ,  $\delta$  справедливо неравенство:

$$P\left\{|\xi(t) - F(t)| < \varepsilon\right\} \geq 1 - \left(\frac{\sigma}{\varepsilon - \delta\sigma}\right)^2. \quad (4.3.3)$$

*Доказательство.* Поскольку  $\lambda = M\xi(t)$ , а  $\sigma^2 = D\xi(t)$ , то в соответствии с неравенством Чебышева [14] имеет место неравенство:

$$P\{|\xi(t) - \lambda| \geq \Omega\} \leq \frac{\sigma^2}{\Omega^2},$$

где  $\Omega > 0$  – некоторая константа.

Тогда, очевидно,

$$P\{|\xi(t) - \lambda| < \Omega\} \geq 1 - \frac{\sigma^2}{\Omega^2}. \quad (4.3.4)$$

Далее имеем:

$$\begin{aligned} P\{|\xi(t) - F(t)| < \varepsilon\} &= P\{|\xi(t) - \lambda + \lambda - F(t)| < \varepsilon\} \geq P\{|\xi(t) - \lambda| + |\lambda - F(t)| < \varepsilon\} = \\ &= P\{|\xi(t) - \lambda| < \varepsilon - |\lambda - F(t)|\} = P\{|\xi(t) - \lambda| < \varepsilon - |F(t) - \lambda|\}. \end{aligned}$$

Поскольку, в соответствии с алгоритмом построения прогнозной функции  $F(t)$ , выполняется условие (4.3.1), то из последнего соотношения получаем:

$$P\{|\xi(t) - \lambda| < \varepsilon - |F(t) - \lambda|\} \geq P\{|\xi(t) - \lambda| < \varepsilon - \delta \cdot \sigma\}. \quad (4.3.5)$$

Применяя неравенство Чебышева [14] к последнему соотношению, получаем при  $\Omega = \varepsilon - \delta\sigma > 0$ :

$$P\{|\xi(t) - \lambda| < \varepsilon - \delta\sigma\} \geq 1 - \left(\frac{\sigma}{\varepsilon - \delta\sigma}\right)^2. \quad (4.3.6)$$

Принимая во внимание соотношения (4.3.5) и (4.3.6), приходим к доказательству утверждения:

$$P\{|\xi(t) - F(t)| < \varepsilon\} \geq 1 - \left(\frac{\sigma}{\varepsilon - \delta\sigma}\right)^2.$$

В частности, если положить  $\varepsilon = k \cdot \sigma$ , где  $k > 0$ , то неравенство (4.3.6) принимает вид:

$$P\{|\xi(t) - F(t)| < k\sigma\} \geq 1 - \left(\frac{1}{k - \delta}\right)^2. \quad (4.3.7)$$

В частности, при  $k=1$  имеем:

$$P\{|\xi(t) - F(t)| < \sigma\} \geq 1 - \left(\frac{1}{1-\delta}\right)^2. \quad (4.3.8)$$

В общем случае, оценка (4.3.7) является содержательной при  $k - \delta > 1$ , то есть, при  $k > \delta + 1$ , и, следовательно, при выполнении условия  $\varepsilon > (1 + \delta) \cdot \sigma$ .

Заметим, что оценки вероятности ошибки прогнозирования, даваемые соотношением (4.3.3) не зависят от вида распределения случайной величины  $\xi(t)$ , являющейся вероятностной моделью исследуемого процесса.

В случае, когда распределение случайной величины  $\xi(t)$  известно, оценки (4.3.3), (4.3.7) и (4.3.8) могут быть уточнены.

Рассмотрим случаи, когда  $\xi(t)$  имеет нормальное распределение [14] и распределение Пуассона [14].

1. Пусть  $\xi(t)$  имеет нормальное распределение с параметрами  $\lambda$  и  $\sigma^2$ .

Тогда,

$$P\left\{|\xi(t) - \lambda| < \varepsilon\right\} = P\left\{\frac{|\xi(t) - \lambda|}{\sigma} < \frac{\varepsilon}{\sigma}\right\} = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\varepsilon/\sigma}^{\varepsilon/\sigma} e^{-t^2/2} dt$$

В этом случае для оценки вероятности ошибки прогнозирования имеем:

$$\begin{aligned} P\{|\xi(t) - F(t)| < \varepsilon\} &\geq P\{|\xi(t) - \lambda| < \varepsilon - \delta\sigma\} = P\left\{\frac{|\xi(t) - \lambda|}{\sigma} < \frac{\varepsilon}{\sigma} - \delta\right\} = \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\varepsilon/\sigma + \delta}^{\varepsilon/\sigma - \delta} e^{-t^2/2} dt. \end{aligned}$$

2. Пусть случайная величина  $\xi(t)$  имеет распределение Пуассона [19]:

$$P\{\xi(t) = k\} = \frac{\lambda^k e^{-\lambda}}{k!},$$

где  $k=0, 1, 2, \dots$ ,  $\lambda = M\xi(t)$ .

Из формулы остаточного члена Маклорена [14] следует справедливость неравенства:

$$P\{\xi(t) > L\} = \sum_{k=L}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \leq \frac{\lambda^{L+1}}{(L+1)!} e^{-\lambda},$$

где  $L$ - целое.

Тогда

$$P\{|\xi(t) - \lambda| \geq L\} = P\{\xi(t) \geq \lambda + L\} < \frac{\lambda^{\lambda+L+1}}{(\lambda+L+1)!} e^{-\lambda},$$

и

$$P\{|\xi(t) - \lambda| < L\} \geq 1 - \frac{\lambda^{\lambda+L+1}}{(\lambda+L+1)!} e^{-\lambda}.$$

Следовательно, справедливо соотношение:

$$\begin{aligned} P\{|\xi(t) - F(t)| < L\} &\geq P\{|\xi(t) - \lambda| < L - \delta\lambda\} \geq 1 - \frac{\lambda^{\lambda+L+1-\delta\lambda}}{(\lambda+L+1-\delta\lambda)!} e^{-\lambda} = \\ &= 1 - \frac{\lambda^{L+1+\lambda(1-\delta)}}{(L+1+\lambda(1-\delta))!} e^{-\lambda}. \end{aligned}$$

Следует отметить, что во всех рассмотренных в настоящей диссертационной работе прогнозных задач значения исследуемых величин хорошо согласовались с нормальным распределением.

Результаты экспериментальных исследований по оценке точности прогнозирования компьютерных событий приведены в главе 5.

#### **§ 4.4. Разработка инструментария для построения аппроксимирующих функций.**

В настоящей главе дается описание инструментальных средств для построения непрерывных аппроксимирующих функций модифицированным методом наименьших квадратов [31-34].

Как было отмечено выше, аппроксимирующая функция  $F(x, c_0, c_1, \dots, c_m)$  ищется в виде линейной комбинации базовых функций  $\varphi_0(x), \varphi_1(x), \dots, \varphi_m(x)$ . В этом случае, неизвестные коэффициенты  $c_0, c_1, \dots, c_m$  могут быть найдены известными методами поиска минимума функции нескольких переменных. Для решения вопроса о выборе базовых функций, при котором будет дости-



гаться минимум величины  $\Phi(c_0, c_1, \dots, c_m)$ , требуется применение вычислительных процедур.

В начале настоящих диссертационных исследований, при построении указанных прогнозных моделей и поиске соответствующих аппроксимирующих функций, применялся метод наименьших квадратов, реализованный в пакете прикладных программ «Wolfram Mathematica» [23], который, в свою очередь, имеет ряд ограничений в возможности применения. С целью преодоления возникающих ограничений и расширения возможностей применения метода аппроксимации автором разработано специализированное программное обеспечение, реализующее алгоритм модифицированного метода наименьших квадратов и имеющее значительно более широкий спектр возможностей, чем пакет «Wolfram Mathematica». В частности, значительно расширено множество базовых функций, предусмотрена как ручная, так и автоматическая процедура перебора значений параметров при поиске аппроксимирующей функции, обеспечивающей наилучшее приближение на заданном временном интервале.

Блок схема рассматриваемого алгоритма приведена на рисунке 4.4.1.

Рассмотрим краткое описание разработанной программы.

*Ввод данных.* На вход программе подаётся таблично заданная функция. В качестве источника используется файл Excel следующего вида таблица 4.4.1.

В таблице 4.4.1 первый столбец – значения величин  $x_i$ , второй столбец – значения величин  $y_i$ . Количество считываемых точек равно количеству использованных строк, поэтому ниже этой таблицы не должно быть занятых ячеек (правее таблицы могут быть не пустые ячейки, они никак не повлияют на работу программы).

Блок схема рассматриваемого алгоритма приведена на рисунке 4.4.1.

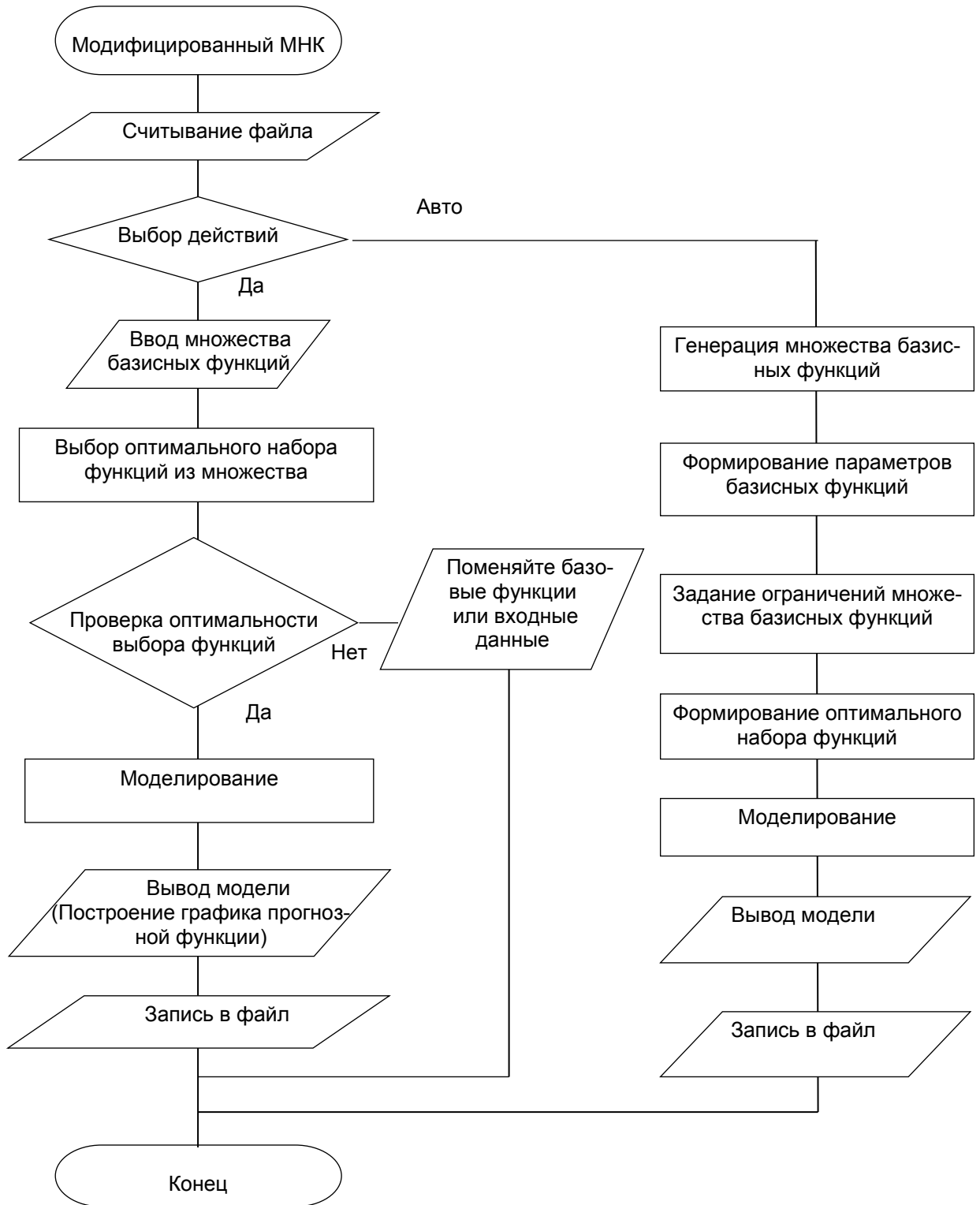


Рис. 4.4.1. Блок схема алгоритма, на основе модифицированного МНК.

Таблица 4.4.1. Формат исходных данных.

	A	B
1	1	0,841471
2	2	0,909297
3	3	0,14112
4	4	-0,7568
5	5	-0,95892
6	6	-0,27942
7	7	0,656987
8	8	0,989358
9	9	0,412118
10	10	-0,54402
11	11	-0,999999
12	12	-0,53657
13	13	0,420167
14	14	0,990607
15	15	0,650288
16	16	-0,2879
17	17	-0,9614
18	18	-0,75099
19	19	0,149877
20	20	0,912945

### Интерфейс программы

На рисунке 4.4.2 представлен интерфейс разработанной программы

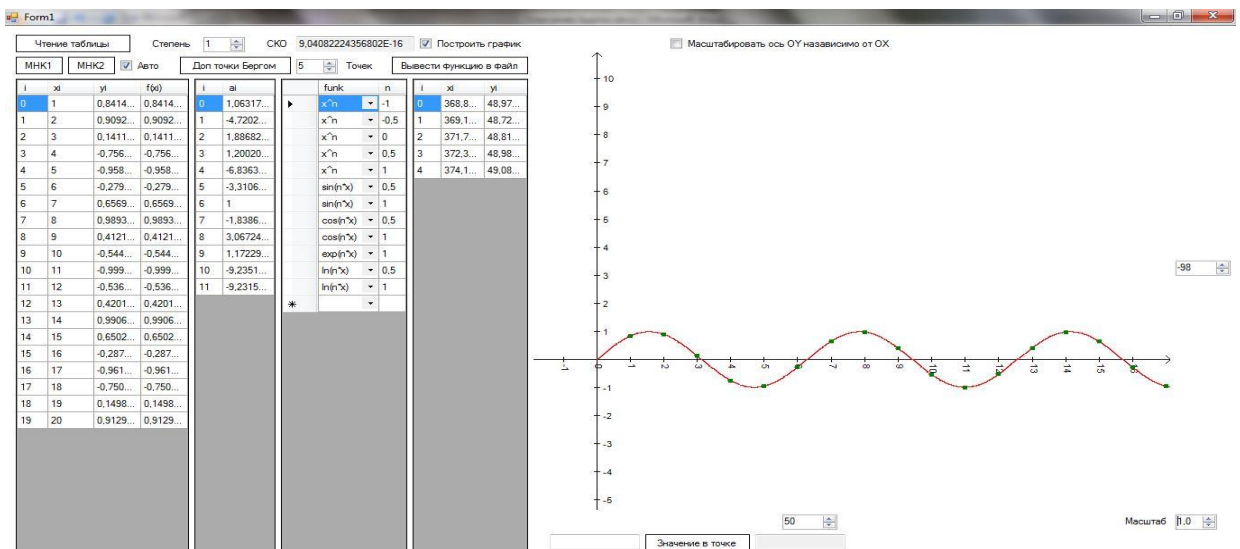


Рис. 4.4.2. Интерфейс программы.

Построение аппроксимирующей функции с помощью разработанной программы начинается с нажатия кнопки «Чтение таблицы» и выбора файла входных данных. В случае успешного чтения, первая таблица заполняется данными из Excel файла. Если при попытке чтения произошла ошибка, либо файл был пустой, появляется сообщение «Ошибка чтения из файла». После

успешного ввода данных становятся доступны остальные функции программы.

Кнопка «МНК 1» запускает алгоритм поиска многочлена аппроксимирующей функцией методом наименьших квадратов с базисными функциями  $x^i$ . Степень многочлена можно менять в поле «Степень». При этом в первой таблице в четвёртом столбце  $f(x_i)$  появляются значения аппроксимирующей функции в заданных точках, строится график и подсчитывается среднеквадратическое отклонение по узлам таблично заданной функции. Коэффициенты при соответствующих слагаемых многочлена выводятся во вторую таблицу. При изменении степени, многочлен пересчитывается автоматически и все изменения отображаются на графике и в таблицах.

Кнопка «МНК 2» без пометки «Авто» запускает алгоритм поиска аппроксимирующей функции модифицированным методом наименьших квадратов с базисными функциями, выбранными в третьей таблице (количество, порядок, комбинация и дополнительные параметры выбираются любыми, но в случае линейной зависимости программа выдаст ошибку). После успешного выполнения алгоритма, результат выводится аналогично, но коэффициенты во второй таблице уже соответствуют коэффициентам базисных функций в самой аппроксимирующей функции.

Если при нажатии кнопки «МНК 2» стоит пометка «Авто», то выбор базисных функций и параметров для них определяется автоматически, путём перебора используемых различных комбинаций базисных функций и параметров. Сложность рассматриваемых комбинаций регулируется значением в поле «Степень». Критерий для выбора оптимальной конфигурации описан в § 4.2. Вывод результатов аналогичен режиму без автоматического подбора базисных функций, разница заключается в том, что оптимальный набор базисных функций появится в третьей таблице автоматически.

Нажатие кнопки «Доп. точки Берга» запускает описанный выше алгоритм Берга и позволяет построить по нему заданное в поле «Точки» количе-

ство точек. В результате найденные точки появятся в четвёртой таблице и отобразятся на графике.

Если после построения точек методом Берга построить аппроксимирующую функцию, то построение будет производиться с расчётом этих точек. Если это не нужно, то точки сбрасываются построением нуля точек методом Берга.

Для вывода аппроксимирующей функции предусмотрена кнопка «Вывести функцию в файл», с помощью которой найденная функция выводится в указанный файл в виде, представленном на рисунке 4.4.3:

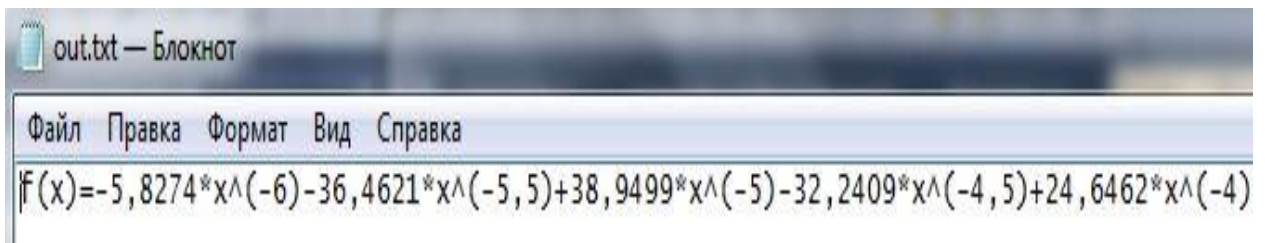


Рис. 4.4.3. Вид аппроксимирующей функции.

#### *Построение графиков аппроксимирующих функций*

Программа предусматривает возможность построения графиков аппроксимирующих функций, которые демонстрируют характер ее изменения.

Пример построения графика приведен на рисунке 4.4.4.

При вычислении заданного прогнозного значения оно указывается на графике аппроксимирующей функции.

В следующей главе диссертации работа предлагаемого метода рассматривается на примере построения технических и экономических показателей.

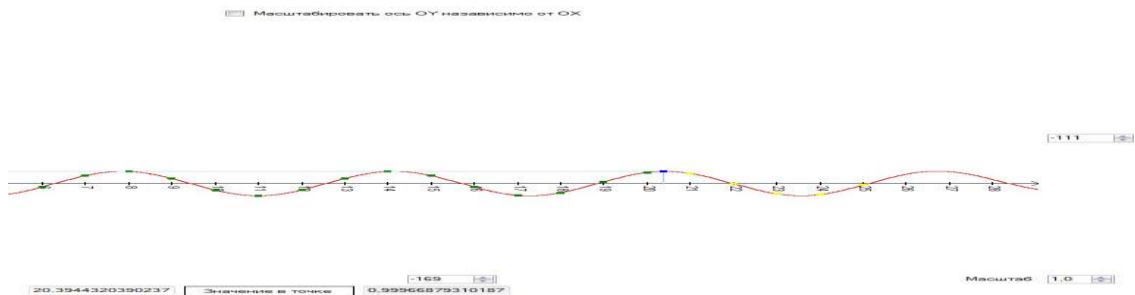


Рис. 4.4.4. График аппроксимирующей функции.

Построенные графики наглядно иллюстрируют поведение аппроксимирующей функции и позволяют делать качественные выводы о характере изменения исследуемых параметров.

Разработанный в работе алгоритм реализован в процедуре контроля эффективности процесса обработки данных, встроенного в виде отдельного модуля в общую схему вычислительного комплекса, представленного на рисунке 4.4.5. В соответствии с приведенной схемой, на него поступают данные с модуля сбора информации о произошедших событиях. Модуль производит их обработку и выдает решение о возможности продолжения штатной работы или о введении дополнительных механизмов для повышения эффективности процедуры обработки данных.

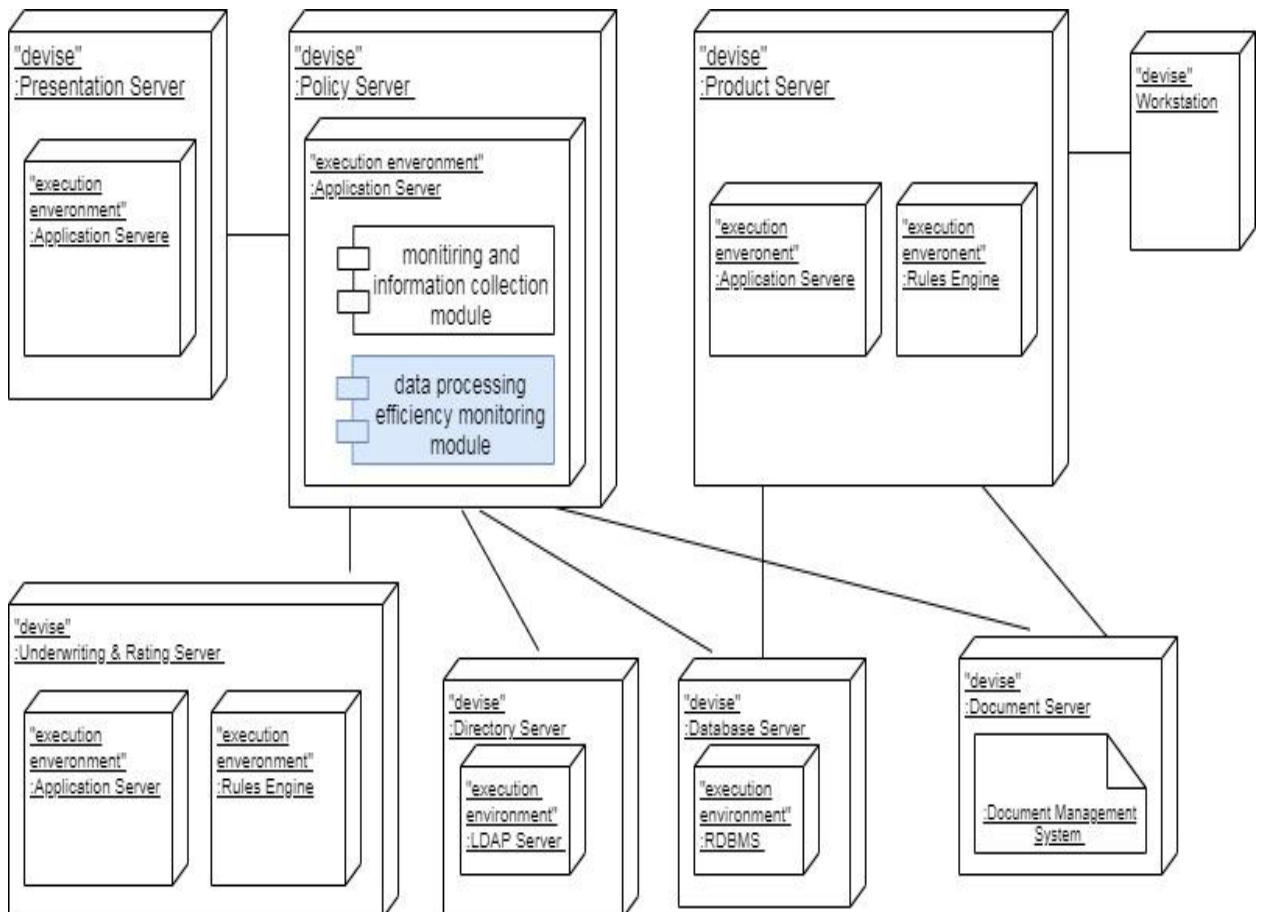


Рис. 4.4.5. Диаграмма (UML) развертывания и компонентов вычислительного комплекса со встроенным модулем контроля эффективности обработки данных

**Выводы по 4 главе.**

В данной главе рассмотрены вопросы разработки методики прогнозирования компьютерных событий. С этой целью данные о компьютерном событии рассматриваются как состояние динамического процесса, исходные значения которого заданы в виде табличной функции. Предложен метод, основанный на построении непрерывной аппроксимирующей функции, значения которой наиболее близко подходят к заданным значениям. За основу взят метод наименьших квадратов поиска аппроксимирующей функции в виде линейной комбинации базовых. Классический метод МНК при этом дополнен условием на выбор базовых функций. С учетом сделанного дополнения получены теоретические оценки точности прогнозирования при применении предлагаемого метода. Для проведения экспериментальных исследований разработано специализированное программное обеспечение, позволяющее строить прогнозные функции, как в интерактивном, так и в автоматическом режиме. Актуальность исследований главы 4 подтверждена ранее результатами главы 3, в которой на основании полученных результатов строятся экспериментальные оценки параметра, характеризующего процесс обработки данных в вычислительных машинах, комплексах и компьютерных сетях - время в течение, которого обеспечивается эффективность процесса обработки указанных данных.

## **ГЛАВА 5. ИССЛЕДОВАНИЕ ТОЧНОСТИ ПРИМЕНЯЕМОГО МЕТОДА ПРОГНОЗИРОВАНИЯ КОМПЬЮТЕРНЫХ СОБЫТИЙ.**

Настоящая глава посвящена вопросам исследования ошибок прогнозирования при применении метода, рассмотренного в главе 4. Как указано ранее, применение методов прогнозирования интенсивности компьютерных событий позволяет строить оценки важного параметра, характеризующего процесс обработки данных в вычислительных машинах, комплексах и компьютерных сетях - время в течение, которого обеспечивается эффективность процесса обработки указанных данных.

В качестве иллюстрации эффективности разработанного метода, приведены результаты экспериментальных исследований по прогнозированию биржевых курсов мировых валют и курсов драгоценных металлов. Данные примеры выбраны по следующим причинам:

- демонстрации универсальности предлагаемого метода прогнозирования, не зависящего от природы исследуемых статистических данных,
- недостаточностью необходимого количества накопленных статистических данных по конкретным компьютерным событиям, находящихся в открытом доступе в частности для какой-либо конкретного программного обеспечения вычислительных машинах, комплексах и компьютерных сетях,
- отнесением исследуемых данных, в частности нанесенного ущерба, к коммерческой тайне в конкретных организациях,
- большим количеством доступных фактических данных, позволяющих в полной мере применить методы статистического анализа.

Приведены результаты исследований ошибок прогнозирования на основе построения статистических критериев.



## § 5.1. Экспериментальные исследования по построению прогнозных моделей

### 5.1.1. Построение прогнозных моделей динамики изменения валютных курсов.

Для проведения данных экспериментальных исследований взяты архивные данные Сбербанка России о курсе покупки доллара США за наличную валюту РФ за периоды с 1 по 31 января 2011 года, с 1 января по 31 марта 2011 года и с 1 января по 30 апреля 2011 года [31]. Значения курсов за данные периоды размещены на сайте Сбербанка России [www.sbrf.ru](http://www.sbrf.ru), график динамики их изменения приведен ниже на рис. 5.1.1.1.



Рис. 5.1.1.1. Курсы Сбербанка России для проведения операций покупки наличной иностранной валюты (доллара США) за наличную валюту РФ за период январь - апрель 2011 года.

Для построения прогнозных моделей в рассматриваемом случае использовались данные за первый месяц 2011 года: с 1 по 31 января (рис. 5.1.1.1).

Как было отмечено выше, первым шагом на пути получения прогнозных оценок является построение по выбранным данным аппроксимирующей функции  $F(x)$ .

С помощью описанного выше метода для данного эксперимента построена аппроксимирующая функция  $F(x) = F_1(x)$  следующего вида:

$$F_1(x) = 30.53 - \frac{0.36}{x^2} - 0.042 \cdot x - 0.098 \cdot \sin x$$

График функции  $y = F_1(x)$  представлен ниже на рис. 5.1.1.2.

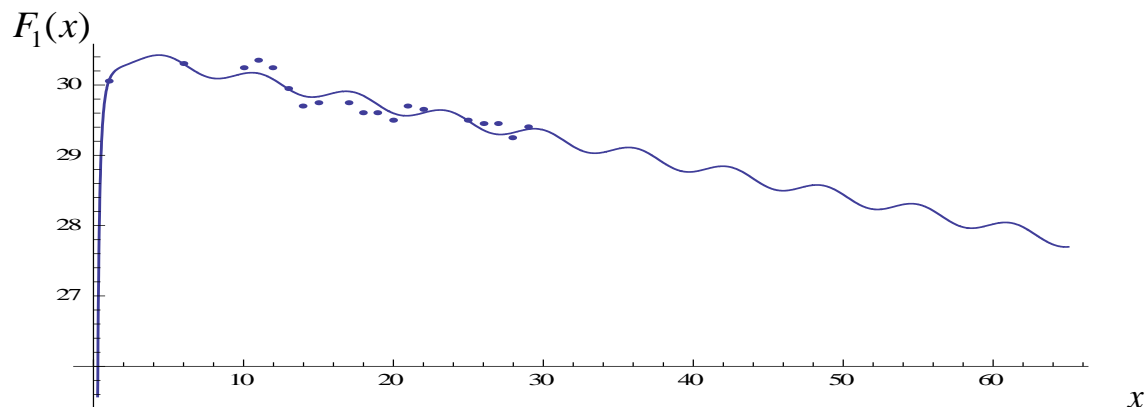


Рис. 5.1.1.2. График функции  $F_1(x) = 30.53 - \frac{0.36}{x^2} - 0.042 \cdot x - 0.098 \cdot \sin x$ .

Построенная модель позволяет вычислить прогнозные значения исследуемого параметра на определенный временной промежуток. В таблице 5.1.1.1 приведены прогнозные значения курса покупки Сбербанком доллара США с 1 февраля по 10 марта 2011 года и соответствующие ошибки прогнозирования.

Таблица 5.1.1.1. Прогноз курса покупки Сбербанком доллара США с 1 февраля по 10 марта 2011 года и ошибки прогнозирования.

Период	Прогнозное значение	Действительное значение	Абсолютное значение	Относительная погрешность	Относительная погрешность (%)
01.02.2011	29,26	29,40	0,14	0,004	0,4
02.02.2011	29,12	29,10	-0,02	-0,0006	-0,06
03.02.2011	29,04	29,05	0,01	0,0003	0,03
04.02.2011	29,04	29,00	-0,04	-0,001	-0,1
05.02.2011	29,09	29,00	-0,09	-0,003	-0,3
07.02.2011	29,03	29,00	-0,03	-0,001	-0,1
08.02.2011	28,90	29,00	0,10	0,003	0,3
09.02.2011	28,79	28,95	0,16	0,005	0,5
10.02.2011	28,77	28,95	0,18	0,006	0,6
11.02.2011	28,81	28,95	0,14	0,004	0,4
12.02.2011	28,85	28,95	0,10	0,003	0,3
14.02.2011	28,79	28,80	0,01	0,0003	0,03
15.02.2011	28,67	28,90	0,23	0,008	0,8

Период	Прогнозное значение	Действительное значение	Абсолютное значение	Относительная погрешность	Относительная погрешность (%)
16.02.2011	28,50	28,90	0,40	0,013	1,3
17.02.2011	28,53	28,90	0,37	0,012	1,2
18.02.2011	28,58	28,80	0,22	0,007	0,7
19.02.2011	28,55	28,80	0,25	0,008	0,8
21.02.2011	28,31	28,80	0,49	0,017	1,7
22.02.2011	28,23	28,80	0,57	0,019	1,9
23.02.2011	28,25	28,90	0,65	0,022	2,2
24.02.2011	28,30	28,60	0,30	0,01	1
25.02.2011	28,30	28,55	0,25	0,008	0,8
26.02.2011	28,21	28,55	0,34	0,01	1
28.02.2011	28,08	28,40	0,32	0,01	1
01.03.2011	27,97	28,40	0,43	0,015	1,5
02.03.2011	28,02	28,00	-0,02	-0,0007	-0,07
03.03.2011	28,04	27,85	-0,19	-0,006	-0,6
04.03.2011	27,97	27,85	-0,12	-0,004	-0,4
05.03.2011	27,85	27,85	0	0	0
09.03.2011	27,78	28,00	0,22	0,007	0,7
10.03.2011	27,74	28,20	0,44	0,015	1,5

В столбце 3 приведены действительные значения указанных выше курсов покупки доллара США за российские рубли, а в столбцах 4 - 6 - абсолютные и относительные значения ошибок прогнозирования.

Как следует из данных таблицы 5.1.1.1, относительная ошибка прогнозирования предлагаемым методом в рассматриваемом случае не превышает 3%.

Результаты исследования характеристик точности прогнозирования с применением статистических критериев приведены ниже в параграфе 5.2.

На рис. 5.1.1.3 представлены результаты первого эксперимента и приведены графики прогнозных и действительных значений валютных курсов

с 1 февраля по 10 марта 2011 года.

### Курсы покупки доллара США с 1.02 по 10.03.2011 года

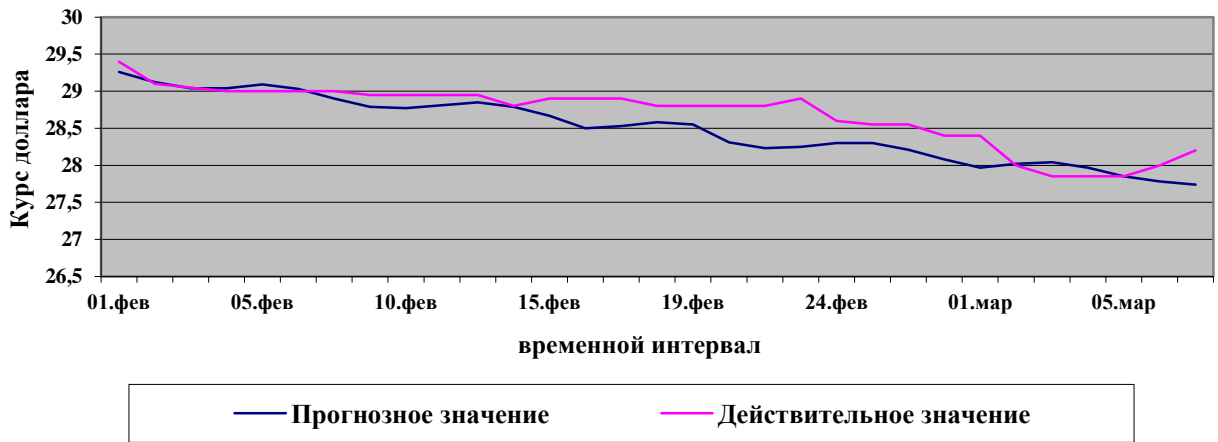


Рис. 5.1.1.3. Графики действительных и прогнозных значений курсов на февраль, март 2011 года.

В втором эксперименте для построения прогнозной функции  $F(x) = F_2(x)$  использовались данные за первый квартал 2011 года: с 1 января по 31 марта (рис. 5.1.1.1).

Аппроксимирующая функция  $F_2(x)$  при этом имела вид:

$$F_2(x) = 30.38 - 0.042 \cdot x + 0.000162 \cdot x^2 + 0.027 \cdot \sin x$$

График функции  $y = F_2(x)$  представлен ниже на рис. 5.1.1.4.

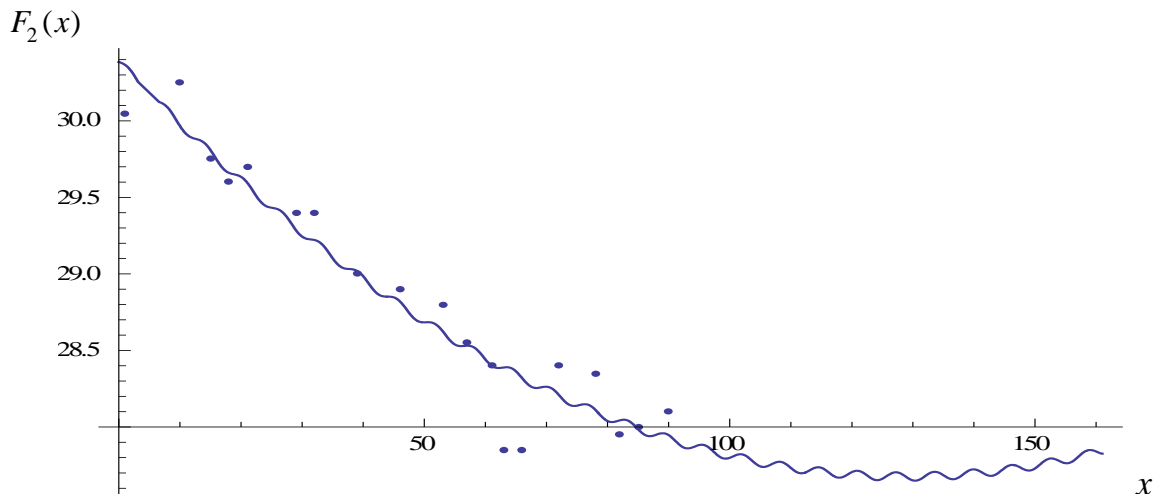


Рис. 5.1.1.4. График функции  $F_2(x) = 30.38 - 0.042 \cdot x + 0.000163 \cdot x^2 + 0.027 \cdot \sin x$

Прогнозные значения на рассматриваемый период с интервалом 3-5 дней приведены в таблице 5.1.1.2.

Таблица 5.1.1.2. Прогноз курсов для проведения Сбербанком России операций покупки иностранной валюты (доллара США) за наличную валюту РФ с 1 апреля по 30 июня 2011 года и ошибки прогнозирования.

Период	Прогнозное значение	Действительное значение	Абсолютная погрешность	Относительная погрешность	Относительная погрешность (%)
01.04.2011	27,91	28,15	0,24	0,008	0,8
05.04.2011	27,88	28,00	0,12	0,004	0,4
07.04.2011	27,85	27,90	0,05	0,001	0,1
09.04.2011	27,79	27,75	-0,001	-0,001	-0,1
12.04.2011	27,82	27,75	-0,07	-0,0025	-0,25
14.04.2011	27,77	27,80	0,03	0,001	0,1
16.04.2011	27,74	27,85	0,11	0,003	0,3
20.04.2011	27,73	28,00	0,27	0,009	0,9
22.04.2011	27,70	27,60	-0,1	-0,003	-0,3
26.04.2011	27,70	27,55	-0,15	-0,005	-0,5
28.04.2011	27,66	27,40	-0,26	-0,009	-0,9
04.05.2011	27,65	27,05	-0,6	-0,002	-2
06.05.2011	27,68	27,50	-0,18	-0,006	-0,6
07.05.2011	27,70	27,50	-0,2	-0,007	-0,7
10.05.2011	27,65	27,65	0	0	0
13.05.2011	27,70	27,65	-0,15	-0,005	-0,5
16.05.2011	27,66	27,80	0,14	0,005	0,5
19.05.2011	27,71	27,80	0,09	0,003	0,3
21.05.2011	27,71	27,70	-0,01	-0,0003	-0,03
23.05.2011	27,68	28,00	0,32	0,01	1
25.05.2011	27,73	28,05	0,32	0,01	1
27.05.2011	27,74	27,85	0,11	0,003	0,3
31.05.2011	27,76	27,75	-0,01	-0,0003	-0,03
03.06.2011	27,78	27,60	-0,18	-0,006	-0,6
07.06.2011	27,80	27,50	-0,3	-0,01	-1
11.06.2011	27,82	27,50	-0,32	-0,01	-1
17.06.2011	27,89	27,85	-0,04	-0,001	-0,1
21.06.2011	27,96	27,80	-0,16	-0,005	-0,5
24.06.2011	27,98	27,90	-0,08	-0,002	-0,2
30.06.2011	28,08	27,65	-0,43	-0,015	-1,5

В столбце 3 приведены действительные значения указанных выше курсов покупки доллара США за российские рубли, а в столбцах 4 - 6 - абсолютные и относительные значения ошибок прогнозирования.

Как следует из данных таблицы 5.1.1.2, относительная ошибка прогнозирования предлагаемым методом в рассматриваемом случае не превышает 1,5%.

На рис. 5.1.1.5 представлены результаты третьего эксперимента и приведены графики прогнозных и действительных значений валютных курсов на период с 1 апреля по 30 июня 2011 года.

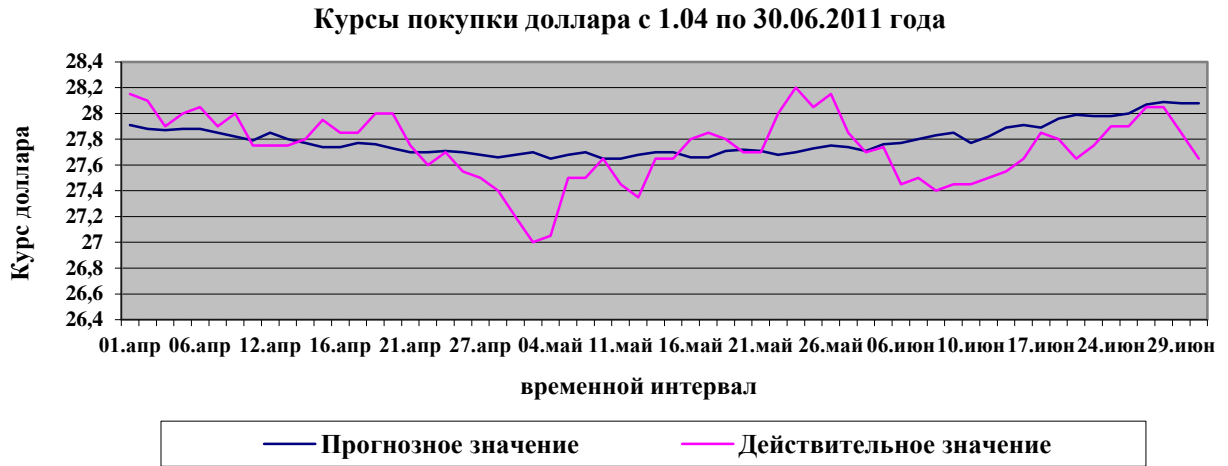


Рис. 5.1.1.5. Графики действительных и прогнозных значений курсов с 1.04 по 30.06.2011г.

В третьем эксперименте по данным с 1 января по 30 апреля 2011 года строился прогноз на 125 дней: с 1 мая по 2 сентября 2011 года.

Аппроксимирующая функция  $F_3(x)$  при этом имела вид:

$$F_3(x) = 30.38 - 0.042 \cdot x + 0.000163 \cdot x^2 + 0.027 \cdot \sin x .$$

График функции  $y = F_3(x)$  представлен на рис. 5.1.1.6.

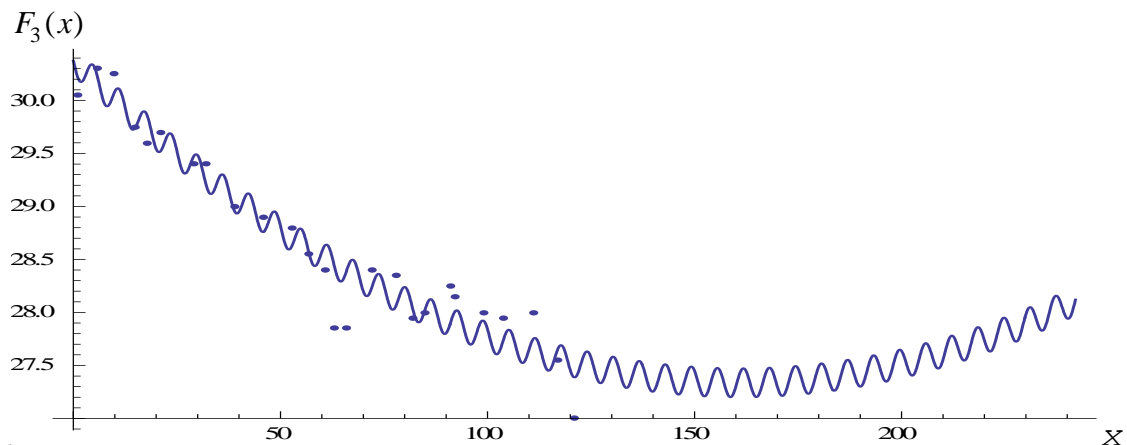


Рис. 5.1.1.6. График функции  $F_3(x) = 30.38 - 0.042 \cdot x + 0.000163 \cdot x^2 + 0.027 \cdot \sin x$

Данные эксперимента с интервалом 3-5 дней приведены в таблице 5.1.1.3.

Таблица 5.1.1.3. Прогноз курса покупки доллара с 4 мая по 2 сентября 2011 года и ошибки прогнозирования.

Период	Прогнозные значения	Действительные значения	Абсолютная погрешность	Относительная погрешность	Относительная погрешность (%)
04.05.2011	27,57	27,05	-0,52	-0,019	-1,9
06.05.2011	27,33	27,10	-0,23	-0,008	-0,8
10.05.2011	27,58	27,65	0,07	0,002	0,2
11.05.2011	27,42	27,35	-0,07	-0,002	-0,2
14.05.2011	27,30	27,65	0,34	0,012	1,2
17.05.2011	27,53	27,85	0,32	0,011	1,1
20.05.2011	27,25	27,70	0,45	0,016	1,6
23.05.2011	27,51	28,00	0,49	0,017	1,7
27.05.2011	27,28	27,85	0,57	0,02	2
31.05.2011	27,45	27,75	0,3	0,01	1
03.06.2011	27,46	27,60	0,14	0,005	0,5
06.06.2011	27,35	27,45	0,1	0,003	0,3
10.06.2011	27,47	27,45	-0,02	-0,0007	-0,07
15.06.2011	27,41	27,55	0,14	0,005	0,5
23.06.2011	27,38	27,65	0,27	0,009	0,9
30.06.2011	27,51	27,65	0,14	0,005	0,5
01.07.2011	27,41	27,60	0,19	0,006	0,6
5.07.2011	27,49	27,50	0,01	0,0003	0,03
6.07.2011	27,55	27,55	0	0	0
11.07.2011	27,50	27,80	0,3	0,01	1
15.07.2011	27,35	27,70	0,35	0,012	1,2
20.07.2011	27,53	27,75	0,22	0,007	0,7
23.07.2011	27,52	27,40	-0,12	-0,004	-0,4
27.07.2011	27,51	27,20	-0,31	-0,011	-1,1
30.07.2011	27,70	27,35	-0,35	-0,012	-1,2
4.08.2011	27,55	27,45	-0,1	-0,003	-0,3
8.08.2011	27,84	28,15	0,31	0,011	1,1
13.08.2011	27,92	28,40	0,48	0,016	1,6
17.08.2011	27,73	28,35	0,62	0,021	2,1
23.08.2011	27,83	28,60	0,77	0,026	2,6
26.08.2011	28,15	28,45	0,3	0,01	1
30.08.2011	27,94	28,30	0,36	0,012	1,2
02.09.2011	28,25	28,55	0,3	0,01	1

В столбце 3 приведены действительные значения указанных выше курсов покупки доллара США за российские рубли, а в столбцах 4 - 6 - абсолютные и относительные значения ошибок прогнозирования.

Как следует из данных таблицы 5.1.1.3, относительная ошибка прогнозирования предлагаемым методом в рассматриваемом случае не превышает 3%.

На рис. 5.1.1.7 приведены графики прогнозных и действительных значений валютного курса на период с 1 мая по 2 сентября 2011 года.

#### Курсы покупки доллара с 1 мая по 2 сентября 2011 года.

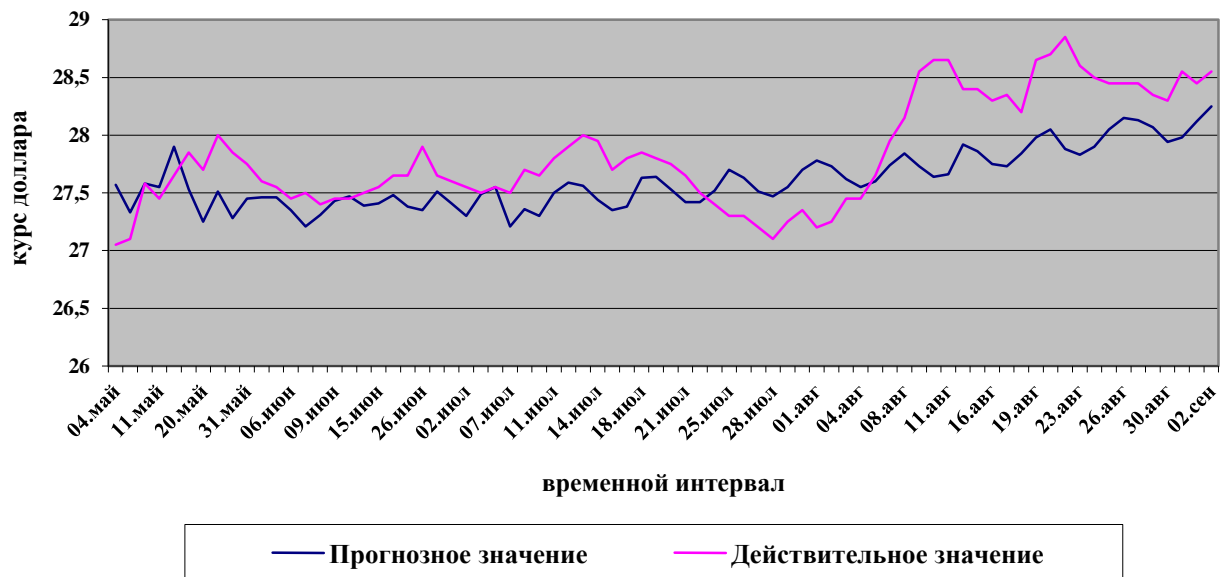


Рис. 5.1.1.7. Графики действительных и прогнозных значений курсов с 1 мая по 2 сентября 2011 г.

Сравнение данных, приведенных в таблицах 5.1.1.1 – 5.1.1.3, показывает, что точность прогнозирования предлагаемым методом в явном виде не зависит от количества дней, на которые стоит прогноз и от расположения временного промежутка [31].

Результаты исследования характеристик точности прогнозирования с применением статистических критериев приведены ниже в параграфе 5.2.



### 5.1.2. Прогнозирование курсов драгоценных металлов

Для проведения экспериментальных исследований по построению прогнозных моделей биржевых курсов драгоценных металлов были взяты архивные данные Сбербанка РФ по указанным выше котировкам за период с 1 января по 30 марта и с 3 по 31 мая 2011 года [32]. Значения котировок за данные периоды размещены на сайте Сбербанка России [www.sbrf.ru](http://www.sbrf.ru), график их изменения приведен ниже на рис. 5.1.2.1.

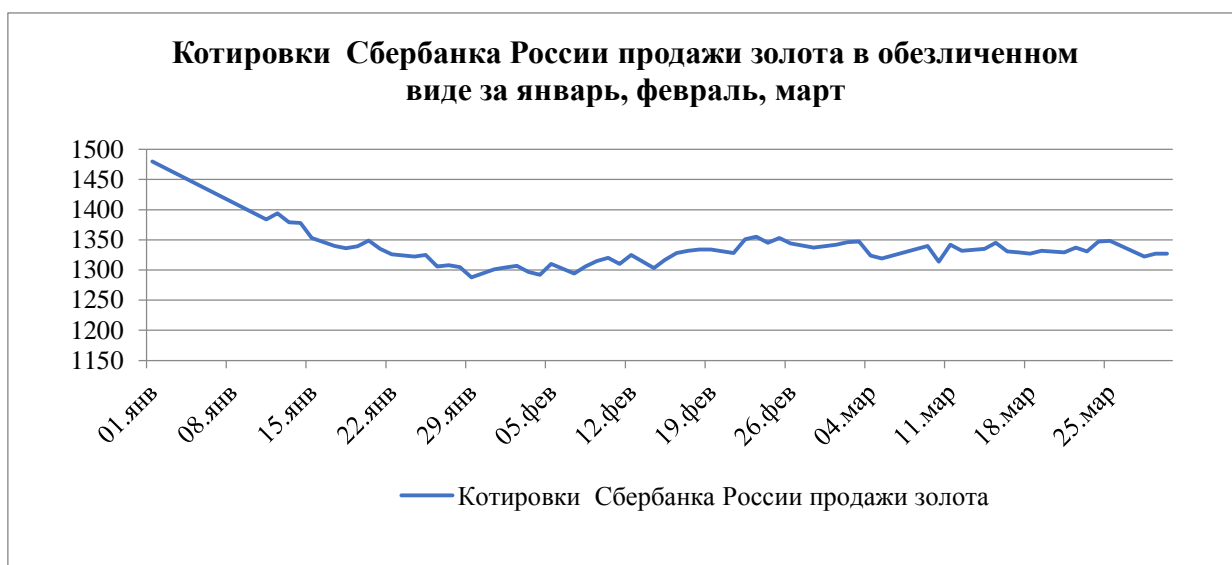


Рис. 5.1.2.1. Котировки Сбербанка России продажи золота в обезличенном виде за январь, февраль и март 2011 года.

Для построения прогнозных моделей в рассматриваемом случае использовались данные за первый месяц 2011 года: с 1 по 31 января (рис. 5.1.2.1).

Как было отмечено выше, первым шагом на пути получения прогнозных оценок является построение по выбранным данным аппроксимирующей функции  $F(x)$ .

С помощью описанного выше метода для данного эксперимента построена аппроксимирующая функция  $F_1(x)$  следующего вида:

$$F_1(x) = 1496.67 - \frac{8.41}{x^2} - 11.32 \cdot x + 0.16 \cdot x^2 + 3.36 \cdot \sin x$$

График функции  $F_1(x)$  представлен ниже на рис. 5.1.2.2.

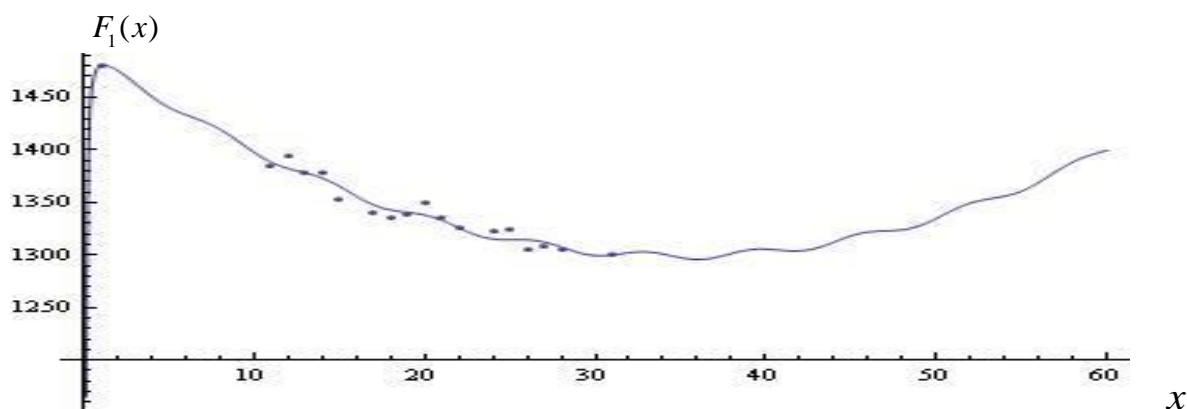


Рис. 5.1.2.2. График функции  $F_1(x) = 1496.67 - \frac{8.41}{x^2} - 11.32 \cdot x + 0.16 \cdot x^2 + 3.36 \cdot \sin x$

Построенная модель позволяет вычислить прогнозные значения исследуемого параметра на определенный временной промежуток. В таблице 5.1.2.1 приведены прогнозные значения курса продажи Сбербанком России золота в обезличенном виде США с 1 по 26 февраля 2011 года и соответствующие ошибки прогнозирования.

В таблице 5.1.2.1 приведены результаты эксперимента по построению прогноза котировок золота на период с 1 по 26 февраля 2011 г.

Таблица 5.1.2.1. Прогноз курса продажи Сбербанком России золота в обезличенном виде с 1 по 26 февраля 2011 года и ошибки прогнозирования.

Период	Прогнозное значение	Действительное значение	Абсолютное значение	Относительная погрешность	Относительная погрешность (%)
01.02.2011	1302,20	1304,00	1,8	0,001	0,1
02.02.2011	1302,90	1307,00	4,1	0,003	0,3
03.02.2011	1300,80	1297,00	-3,8	-0,003	-0,3
04.02.2011	1297,00	1292,00	-5	-0,003	-0,3
05.02.2011	1295,80	1310,00	14,2	0,01	1
07.02.2011	1301,50	1294,00	-7,5	-0,005	-0,5
08.02.2011	1304,90	1306,00	1,1	0,0008	0,08
09.02.2011	1305,60	1315,00	9,4	0,007	0,7
10.02.2011	1304,40	1320,00	15,6	0,01	1
11.02.2011	1306,70	1310,00	3,3	0,0025	0,25
12.02.2011	1312,30	1325,00	12,7	0,009	0,9
14.02.2011	1318,20	1303,00	-15,2	-0,01	-1

15.02.2011	1321,80	1317,00	-4,8	-0,003	-0,3
16.02.2011	1323,00	1328,00	5	0,003	0,3
17.02.2011	1324,00	1332,00	8	0,006	0,6
18.02.2011	1328,00	1334,00	6	0,004	0,4
19.02.2011	1334,00	1334,00	0	0	0
21.02.2011	1349,00	1328,00	-21	-0,015	-1,5
22.02.2011	1353,00	1351,00	-2	-0,001	-0,1
23.02.2011	1355,00	1355,00	0	0	0
24.02.2011	1360,00	1345,00	-15	-0,01	-1
25.02.2011	1369,00	1353,00	-16	-0,01	-1
26.02.2011	1379,00	1344,00	-35	-0,01	-2

Прогнозные значения по выбранным интервалам, полученные предлагаемым методом, приведены во втором столбце таблицы 5.1.2.1.

В последующих столбцах приведены действительные значения указанных выше дневных котировок продажи золота в обезличенном виде Сбербанком России, а также абсолютные и относительные значения ошибок прогнозирования.

Как следует из данных таблицы 5.1.2.1, относительная ошибка прогнозирования предлагаемым методом в рассматриваемом случае не превышает 2%.

На рис. 5.1.2.3 приведены графики прогнозных и действительных значений котировок на февраль 2011 года.

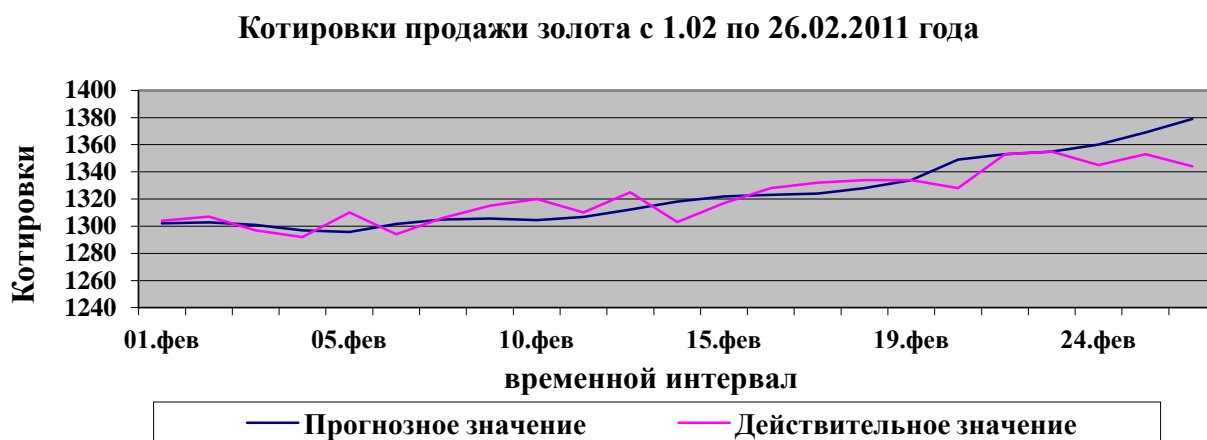


Рис. 5.1.2.3. Графики действительных и прогнозных значений котировок на февраль 2011 года.

Во втором эксперименте для построения прогнозной функции  $F(x) = F_2(x)$  использовались данные за три первых месяца 2011 года: с 1 января по 30 марта рис. 5.1.2.1.

Аппроксимирующая функция  $F_2(x)$  при этом имела вид:

$$F_2(x) = 1347.42 + \frac{145.71}{x^2} - 1.06 \cdot x + 0.01 \cdot x^2 - 10.61 \cdot \sin x$$

График функции  $F_2(x)$  представлен на рис.5.1.2.4.

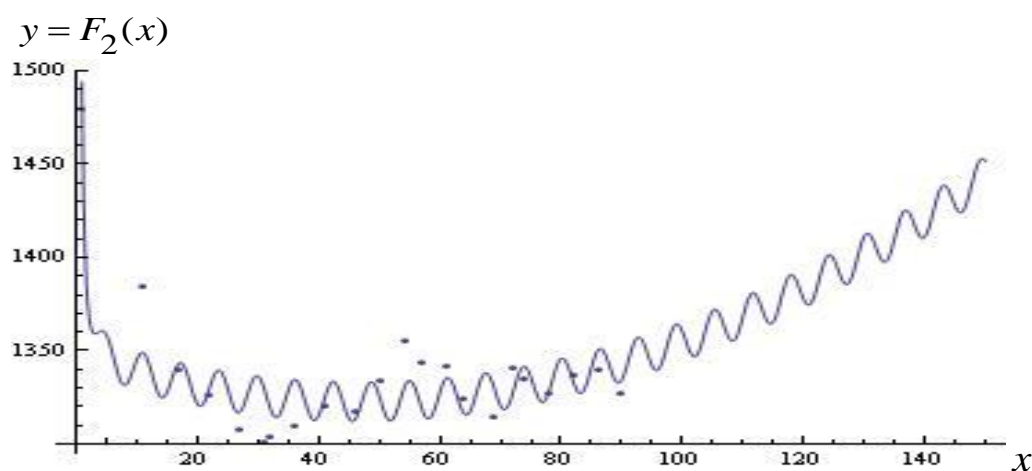


Рис. 5.1.2.4. График функции

$$F_2(x) = 1347.42 + \frac{145.71}{x^2} - 1.06 \cdot x + 0.01 \cdot x^2 - 10.61 \cdot \sin x$$

Прогнозные значения на рассматриваемый период приведены в таблице 5.1.2.2.

Таблица 5.1.2.2. Прогноз котировок продажи золота Сбербанком России в обезличенном виде с 31 марта по 3 июня 2011 года.

Период	Прогнозное значение	Действительное значение	Абсолютная погрешность	Относительная погрешность	Относительная погрешность (%)
31.03.2011	1343,70	1345,00	1,3	0,0009	0,09
05.04.2011	1341,80	1342,00	0,2	0,0001	0,01
07.04.2011	1347,00	1360,00	13	0,009	0,9
09.04.2011	1364,00	1364,00	0	0	0
11.04.2011	1351,00	1349,00	-2	-0,001	-0,1
13.04.2011	1352,00	1354,00	2	0,001	0,1
16.04.2011	1370,00	1374,00	4	0,0029	0,29
20.04.2011	1368,50	1384,00	15,5	0,01	1
22.04.2011	1380,50	1388,00	7,5	0,005	0,5

26.04.2011	1374,60	1389,00	14,4	0,01	1
28.04.2011	1390,80	1381,00	-9,8	-0,007	-0,7
30.04.2011	1398,00	1399,00	1	0,0007	0,07
03.05.2011	1393,60	1389,00	-4,9	-0,003	-0,3
04.05.2011	1401,00	1396,00	-5	-0,003	-0,3
07.05.2011	1385,00	1365,00	-20	-0,014	-1,4
11.05.2011	1411,00	1385,00	-26	-0,019	-1,9
13.05.2011	1398,00	1380,00	-18	-0,013	-1,3
14.05.2011	1399,80	1389,00	-10,8	-0,007	-0,7
19.05.2011	1412,00	1379,00	-33	-0,02	-2
24.05.2011	1435,00	1420,00	-15	-0,01	-1
25.05.2011	1427,40	1425,00	-2,4	-0,0016	-0,16
27.05.2011	1430,50	1418,00	-12,5	-0,008	-0,8
31.05.2011	1444,00	1425,00	-19	-0,013	-1,3
03.06.2011	1442,50	1420,00	-22,5	-0,015	-1,5

В столбце 3 приведены действительные значения указанных выше котировок продажи золота Сбербанком России в обезличенном виде за российские рубли, а в столбцах 4 - 6 - абсолютные и относительные значения ошибок прогнозирования.

Как следует из данных таблицы 5.1.2.2, относительная ошибка прогнозирования предлагаемым методом в рассматриваемом случае не превышает 2%. На рис. 5.1.2.5 приведены графики прогнозных и действительных значений котировок золота на период с 31 марта по 3 июня 2011 года.

**Котировки продажи золота с 31.03 по 03.06.2011 года**

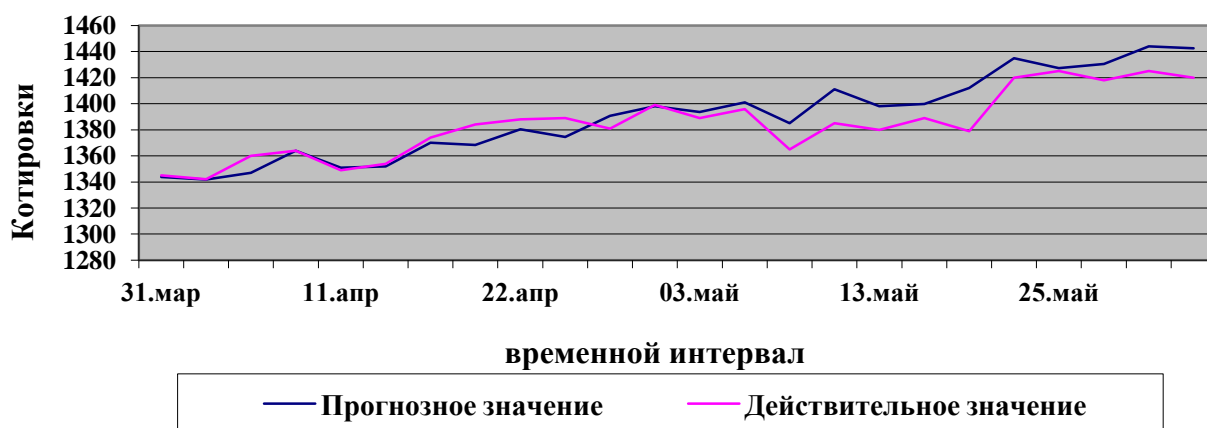


Рис. 5.1.2.5. Графики действительных и прогнозных значений котировок с 31.03 по 03.06. 2011 г.

В третьем эксперименте по данным с 03 по 31 мая 2011 года строился краткосрочный прогноз на первую половину июня 2011 года. Данный эксперимент проводился с целью установления зависимости точности прогнозирования от различных факторов, в частности, от количества прогнозируемых дней и от времени года. Данные эксперимента приведены в таблице 5.1.2.3.

Таблица 5.1.2.3. Прогноз котировок продажи Сбербанком России золота в обезличенном виде на период с 1 по 15 июня 2011 года.

Период	Прогнозные значения	Действительные значения	Абсолютная погрешность	Относительная погрешность	Относительная погрешность (%)
01.06.2011	1425,00	1420,00	-5	-0,003	-0,3
02.06.2011	1426,00	1417,00	-9	-0,006	-0,6
03.06.2011	1425,00	1420,00	-5	-0,003	-0,3
04.06.2011	1425,00	1416,00	-9	-0,006	-0,6
06.06.2011	1423,00	1405,00	-18	-0,01	-1
07.06.2011	1422,00	1420,00	-2	-0,001	-0,1
08.06.2011	1420,00	1416,00	-4	-0,002	-0,2
09.06.2011	1417,00	1410,00	-7	-0,004	-0,4
10.06.2011	1413,00	1414,00	1	0,000	0
11.06.2011	1409,00	1415,00	6	0,004	0,4
14.06.2011	1393,00	1388,00	-5	-0,003	-0,3
15.06.2011	1386,00	1397,00	11	0,007	0,7

В столбце 3 приведены действительные значения указанных выше котировок продажи золота Сбербанком России в обезличенном виде за российские рубли, а в столбцах 4 - 6 - абсолютные и относительные значения ошибок прогнозирования.

Как следует из данных таблицы 5.1.2.3, относительная ошибка прогнозирования предлагаемым методом в рассматриваемом случае не превышает 1%.

На рис. 5.1.2.6 приведены графики прогнозных и действительных значений котировок на период с 1 по 15 июня 2011 года.

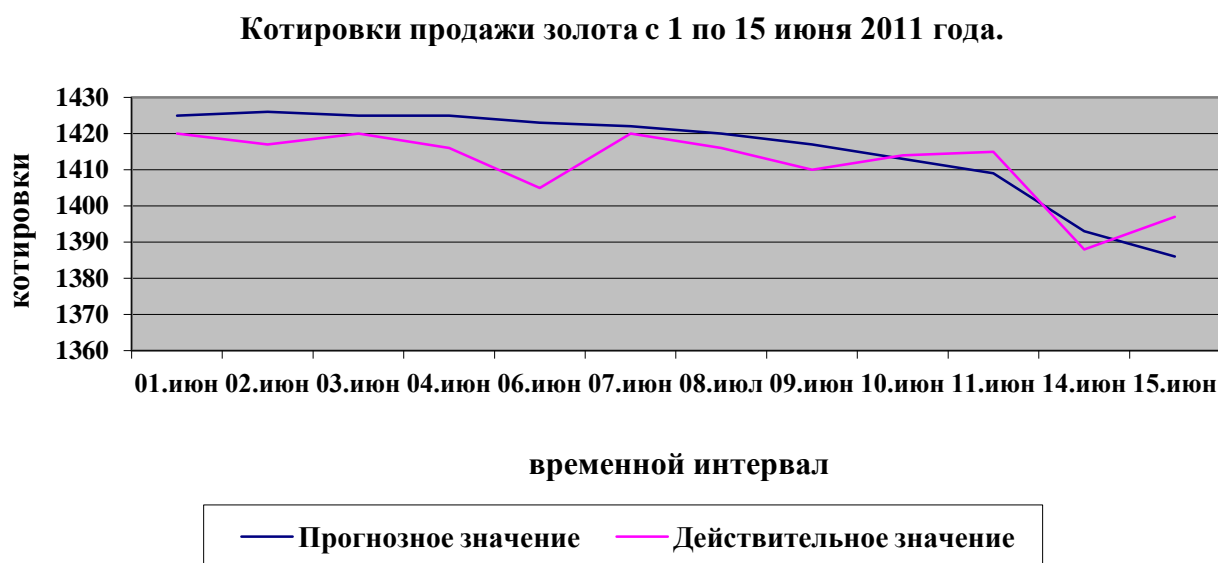


Рис. 5.1.2.6. Графики действительных и прогнозных значений котировок с 1 по 15 июня 2011 г.

Сравнение данных таблиц 5.1.2.1-5.1.2.3 показывает, что точность прогнозирования предлагаемым методом в явном виде не зависит от количества дней, на которые стоит прогноз и от расположения временного промежутка [32].

Результаты исследования характеристик точности прогнозирования с применением статистических критериев приведены ниже в параграфе 5.2.

### *5.1.3. Прогнозирование динамики изменения параметров средств вычислительной техники*

В данном разделе приведены результаты исследований по построению прогнозных моделей динамики изменения основных средств вычислительной техники – быстродействия и объема памяти. Прогнозирование указанных параметров необходимо для построения важного параметра, характеризующего процесс обработки данных в вычислительных машинах, комплексах и компьютерных сетях - время в течение, которого обеспечивается эффективность процесса обработки указанных данных [34].

В таблицах 5.1.3.1 и 5.1.3.2 приведены обобщенные данные динамики развития производительности ЭВМ, измеряемой во флопсах (flop/s), терафлопсах (Tflop/s), петафлопсах (Pflop/s) – количестве элементарных операций в секунду и объем памяти жестких дисков, измеряемой в байтах. Данные соответствуют ежегодным обзорам суммарной производительности наиболее мощных ЭВМ [161].

На рисунках 5.1.3.1 и 5.1.3.2 приведен явный вид аппроксимирующих функций, определяющих прогнозную модель и их графическое представление.

В таблицах 5.1.3.1 и 5.1.3.2 приведены прогнозные значения оцениваемых параметров на период до 2024 года.

*Прогноз суммарной производительности ЭВМ.*

Таблица 5.1.3.1. Динамика суммарной производительности ЭВМ.

№	Год выпуска	Производительность
1	1995	4,8 Tflop/s
2	1996	8 Tflop/s
3	1997	16,9 Tflop/s
4	1998	29,4 Tflop/s
5	1999	50,9 Tflop/s
6	2000	88,1 Tflop/s
7	2001	135 Tflop/s
8	2002	291,8 Tflop/s
9	2003	526,7 Tflop/s
10	2004	812,3 Tflop/s
11	2005	2,3 Pflop/s
12	2006	3,6 Pflop/s
13	2007	7,0 Pflop/s
14	2008	17,4 Pflop/s
15	2009	28,0 Pflop/s
16	2010	43,8 Pflop/s
17	2011	74,1 Pflop/s
18	2012	162,1 Pflop/s
19	2013	250,1 Pflop/s
20	2014	308,9 Pflop/s
21	2015	393,7 Pflop/s



Прогнозная функция  $F_1(x)$ , построенная по данным таблицы 5.1.3.1, имеет вид:

$$F_1(x) = -4.94 \cdot 10^6 + \frac{2.57 \cdot 10^7}{x^2} - 535190 \cdot x + 9523.34x^6 + 3.95 \cdot 10^6 \cdot \ln x$$

ее график представлен на рисунке 5.1.3.1.

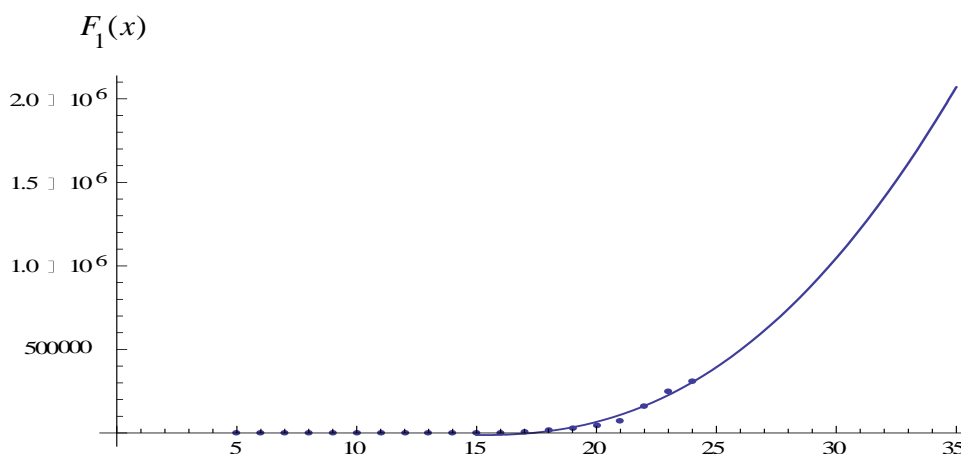


Рис. 5.1.3.1. График функции  $F_1(x)$ .

На представленном графике по оси OX за нулевое значение принят 1991 год, точками отмечены данные за прошлый период, последняя точка соответствует 2015 году. По оси OY данные приведены в терафлопсах (Tflops).

В таблице 5.1.3.2 представлены прогнозные значения суммарной производительности ЭВМ на период до 2024 года.

Таблица 5.1.3.2. Прогноз увеличения суммарной производительности ЭВМ

№	Год	Производительность (P flop/s)
1	2016	496,5
2	2017	612,9
3	2018	743,2
4	2019	887,7
5	2020	1046,9
6	2021	1220,9
7	2022	1410,1
8	2023	1614,6
9	2024	1834,7

Прогноз суммарного объема памяти жестких дисков ЭВМ.

Таблица 5.1.3.3. Динамика суммарного объема памяти жестких дисков ЭВМ.

№	Год выпуска	Емкость жестких дисков
1	1981	10 Мб
2	1990	320 Мб
3	1995	2 Гб
4	1997	10 Гб
5	2002	140 Гб
6	2005	500 Гб
7	2007	1 Тб
8	2009	2 Тб
9	2010	3 Тб
10	2012	4 Тб
11	2013	6 Тб
12	2014	8 Тб
13	2015	12.6 Тб

Прогнозная функция  $F_2(x)$ , построенная по данным таблицы 5.1.3.3, имеет вид:

$$F_2(x) = 4.38 \cdot 10^7 + 3.52 \cdot 10^{-9} \cdot e^x - \frac{2.78 \cdot 10^9}{x^2} - 2.97 \cdot 10^6 \cdot x + 56926.2x^2,$$

ее график представлен на рисунке 5.1.3.2.

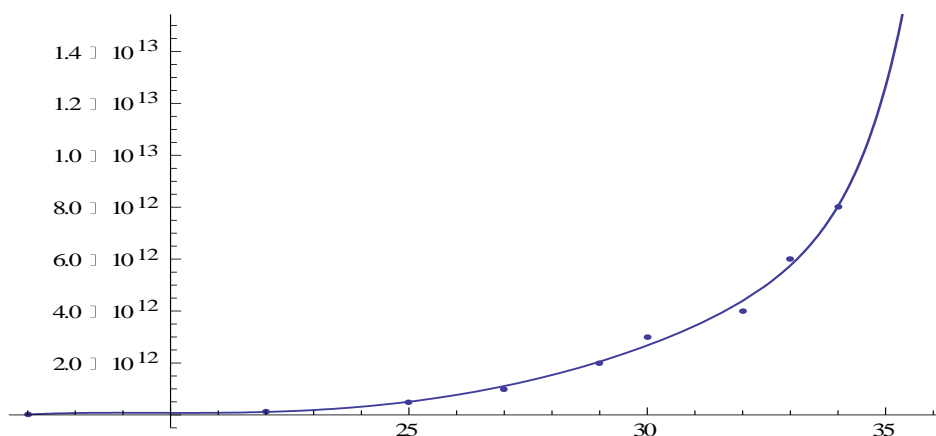


Рис. 5.1.3.2. График функции  $F_2(x)$ .

На представленном графике по оси ОХ за нулевое значение принят 1981 год, точками отмечены данные за прошлый период, последняя точка соответствует 2015 году. По оси ОУ данные приведены в байтах.

В таблице 5.1.3.4 представлены прогнозные значения суммарного объема жестких дисков на период до 2024 года.

Таблица 5.1.3.4. Прогноз увеличения суммарной производительности ЭВМ

№	Год выпуска	Емкость жестких дисков
1	2016	23.4 Тб
2	2017	50.9 Тб
3	2018	123 Тб
4	2019	318 Тб
5	2020	844 Тб
6	2021	2.27 Рб
8	2022	6.15 Рб
9	2023	16.7 Рб
10	2024	45.3 Рб

## **§ 5.2. Экспериментальное исследование точности прогнозирования на примерах прогнозирования курсов валют и драгоценных металлов**

В данном параграфе представлены результаты экспериментальных исследований точности прогнозирования предложенным выше методом. В качестве примера рассмотрена задача прогнозирования биржевых курсов валют и драгоценных металлов [31,32]. Указанные примеры выбраны в связи с достаточно большим объемом фактического материала позволяющего обеспечить необходимый объем выборки.

### *5.2.1. Оценка точности прогнозирования биржевых курсов валют.*

В таблице 5.1.1.3 приведены действительные и соответствующие прогнозные значения курса покупки Сбербанком доллара США с 4 мая по 2 сентября 2011 года с интервалом 3-5 дней, а также абсолютные и относительные ошибки прогнозирования.

С целью проверки качества прогнозирования была выдвинута гипотеза о том, что величины относительных ошибок прогноза являются случайными величинами, имеющими нормальное вероятностное распределение [10].

Для проверки гипотезы о нормальном распределении относительных ошибок прогнозирования был построен статистический критерий Пирсона [10]. Для выборки  $\{x_i\}$  данных об относительных ошибках приближения (выраженных в процентах) объема  $N=95$ , произведена оценка среднего значения

$\bar{x} = \frac{1}{N} \cdot \sum_{i=1}^N x_i = 0,95$ , среднеквадратичного отклонения  $\sigma = 1,23$ , и построена нормированная выборка  $X_i = \frac{x_i - \bar{x}}{\sigma}$ , значения которой приведены в

таблице 5.2.1.1.

Таблица 5.2.1.1. Нормированные упорядоченные значения выборки  $\{x_i\}$ .

№	Значение выборки $X_i$	№	Значение выборки $X_i$	№	Значение выборки $X_i$	№	Значение выборки $X_i$	№	Значение выборки $X_i$
1	-2,36	20	-0,682	39	-0,088	58	0,235	77	0,8
2	-2,36	21	-0,653	40	-0,008	59	0,235	78	0,804
3	-2,28	22	-0,577	41	-0,008	60	0,235	79	0,804
4	-2,12	23	-0,495	42	-0,008	61	0,235	80	0,88
5	-1,87	24	-0,495	43	-0,008	62	0,235	81	0,88
6	-1,79	25	-0,495	44	0,07	63	0,235	82	0,88
7	-1,715	26	-0,495	45	0,07	64	0,31	83	0,967
8	-1,7	27	-0,41	46	0,07	65	0,317	84	1,13
9	-1,634	28	-0,333	47	0,07	66	0,479	85	1,2
10	-1,6	29	-0,333	48	0,07	67	0,479	86	1,37
11	-1,39	30	-0,33	49	0,07	68	0,479	87	1,37
12	-1,22	31	-0,33	50	0,07	69	0,479	88	1,45
13	-1,06	32	-0,33	51	0,15	70	0,56	89	1,536
14	-0,983	33	-0,33	52	0,154	71	0,56	90	1,536
15	-0,98	34	-0,25	53	0,154	72	0,56	91	1,69
16	-0,902	35	-0,25	54	0,154	73	0,64	92	1,69
17	-0,796	36	-0,17	55	0,203	74	0,642	93	2,1
18	-0,739	37	-0,17	56	0,235	75	0,723	94	2,02
19	-0,7	38	-0,17	57	0,235	76	0,8	95	2,34

Для удобства дальнейшей работы значения выборки упорядочены по возрастанию с учетом знаков. Диаграмма частотных характеристик указанной выборки, сгруппированных по интервалам.

Диаграмма приведена на рис. 5.2.1.1.

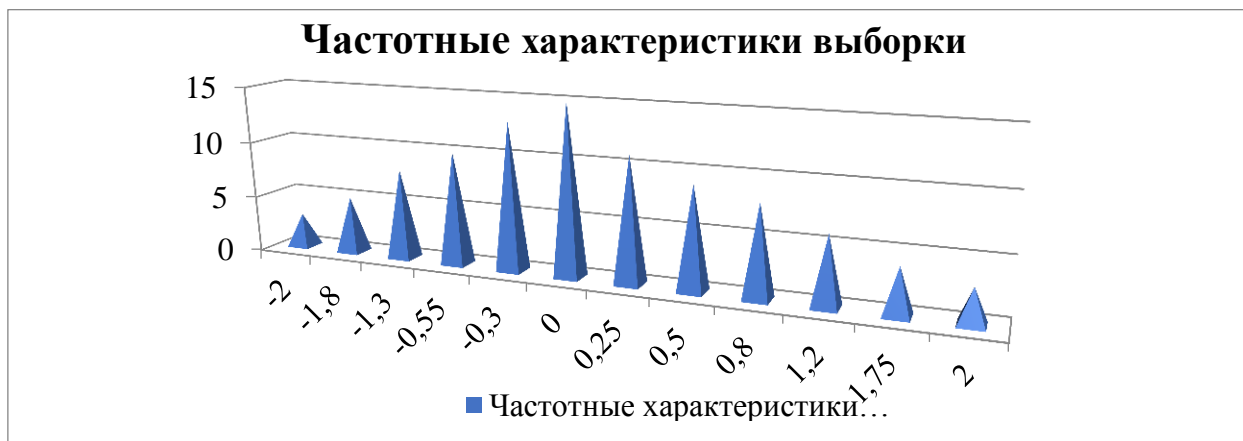


Рис. 5.2.1.1. Диаграмма частотных характеристик выборки.

Далее все значения полученной выборки были разделены на 12 интервалов, вычислены эмпирические и теоретические частоты попадания значений в соответствующий интервал, и произведено их сравнение по критерию Пирсона. Для этого вначале вычислено наблюдаемое значение критерия Пирсона:  $\chi^2_{набл} = 2,35$ . Затем с помощью таблицы критических точек критерия  $\chi^2$ , по уровню значимости  $\alpha = 0,95$  и числу степеней свободы  $k = s - 3 = 9$ , ( $s = 12$  – число интервалов), найдена крайняя точка правосторонней критической области  $\chi^2_{кр} (0,95; 9) = 3,53$ . Так как в нашем случае  $\chi^2_{набл} < \chi^2_{кр}$ , то гипотеза о нормальном распределении (с заданным уровнем значимости) рассматриваемой выборки принимается.

Подтверждение гипотезы о нормальном распределении ошибок построенной прогнозной модели позволяет вычислить границы вероятных ошибок. По таблицам нормального распределения [10] находим, что с вероятностью 0,95 ошибка прогноза составляет не более 3,37%.

*Оценка точности прогнозирования курса драгоценных металлов (золота)*

В таблице 5.1.2.2 приведены действительные и соответствующие прогнозные значения котировок продажи золота Сбербанком России в обезличенном виде с 31 марта по 3 июня 2011 года, а также абсолютные и относительные ошибки прогнозирования.

С целью проверки качества прогнозирования была выдвинута гипотеза о том, что величины относительных ошибок прогноза являются случайными величинами, имеющими нормальное вероятностное распределение [31,32].

Для проверки гипотезы о нормальном распределении относительных ошибок прогнозирования был построен статистический критерий  $\chi^2$  [10].

Для выборки  $\{v_i\}$  данных об относительных ошибках приближения (выраженных в процентах) объема  $N=65$ , произведена оценка среднего значения

$\bar{v} = \frac{1}{N} \cdot \sum_{i=1}^N v_i = -1.13$ , среднеквадратичного отклонения  $\sigma = 1.58$ , и построена

нормированная выборка  $V_i = \frac{v_i - \bar{v}}{\sigma}$ , значения которой приведены в таблице

5.2.1.2. Для целей дальнейшей работы значения выборки упорядочены по возрастанию с учетом знаков.

Таблица 5.2.1.2. Нормированные значения выборки

№	Значение выборки $V_i$	№	Значение выборки $V_i$	№	Значение выборки $V_i$	№	Значение выборки $V_i$	№	Значение выборки $V_i$
1	-1,82	14	-0,99	27	-0,23	40	0,52	53	0,97
2	-1,82	15	-0,93	28	-0,17	41	0,61	54	1,03
3	-1,82	16	-0,93	29	-0,11	42	0,65	55	1,03
4	-1,82	17	-0,87	30	-0,11	43	0,7	56	1,15
5	-1,69	18	-0,87	31	-0,11	44	0,7	57	1,22
6	-1,57	19	-0,55	32	-0,11	45	0,72	58	1,28
7	-1,18	20	-0,55	33	0,07	46	0,75	59	1,28
8	-1,18	21	-0,55	34	0,079	47	0,77	60	1,28
9	-1,18	22	-0,49	35	0,2	48	0,77	61	1,34
10	-1,18	23	-0,49	36	0,27	49	0,88	62	1,34
11	-1,18	24	-0,42	37	0,3	50	0,88	63	1,34
12	-1,18	25	-0,42	38	0,34	51	0,89	64	1,34
13	-1,06	26	-0,36	39	0,52	52	0,96	65	1,72

Для удобства дальнейшей работы значения выборки упорядочены по возрастанию с учетом знаков. Диаграмма частотных характеристик указанной выборки, сгруппированных по интервалам, приведена на рис. 5.2.1.2.



Рис. 5.2.1.2. Диаграмма частотных характеристик выборки.

Далее все значения полученной выборки делим на 9 интервалов, вычисляем эмпирические и теоретические частоты и сравниваем их по критерию Пирсона [10]. Для этого вычисляем наблюдаемое значение критерия Пирсона, в нашем случае оно равно  $\chi_{набл}^2 = 3,24$ . Затем по таблице критических точек  $\chi^2$  по уровню значимости  $\alpha = 0,05$  и числу степеней свободы  $k = s - 3$ , ( $s$  – число интервалов) находим критическую точку правосторонней критической области  $\chi_{кр}^2(0,05; 6) = 12,6$ . Так как в нашем случае  $\chi_{набл}^2 < \chi_{кр}^2$ , то гипотеза о нормальном распределении рассматриваемой выборки принимается.

Подтверждение гипотезы о нормальном распределении ошибок построенной прогнозной модели позволяет вычислить границы вероятных ошибок. По таблицам [10] нормального распределения находим, что с вероятностью 0,95 ошибка прогноза составляет не более 5,85%.

### § 5.3. Сравнение точности прогнозирования модифицированным методом МНК и методом полиномов Лагранжа.

В качестве исходных данных рассмотрим эксперименты по построению аппроксимирующих функций с помощью линейной комбинацией базовых функций и с помощью полиномов для случая прогнозирования курса покупки доллара США Сбербанком РФ [35].

В таблице 5.3.1. приведены данные значений курсов покупки доллара США Сбербанком России за наличную валюту РФ на период 1 по 29 января 2014 года.

Таблица 5.3.1. Данные покупки наличной иностранной валюты (доллара США) за наличную валюту РФ с 1 по 29 января 2014 года.

Период	Значение курса покупки в рублях за 1 доллар
01.01.2014	32,55
06.01.2014	32,55
10.01.2014	32,55
11.01.2014	32,65
14.01.2014	32,55
15.01.2014	32,70
17.01.2014	32,80
18.01.2014	32,80
20.01.2014	33,00
21.01.2014	33,20
22.01.2014	33,40
23.01.2014	33,35
24.01.2014	33,60
25.01.2014	33,85
27.01.2014	34,10
28.01.2014	33,85
29.01.2014	34,00

Степени полинома (количество узловых точек):  $n = 17, 15, 13$ .

**$n = 17$ .**

Аппроксимирующая функция, построенная по 17 узловым точкам, имеет вид:

$$F(x) = 31.66 + \frac{0.85}{x^2} + 0.077x - 0.02 \sin[x]$$



График ее функции представлен на рисунке 5.3.1.

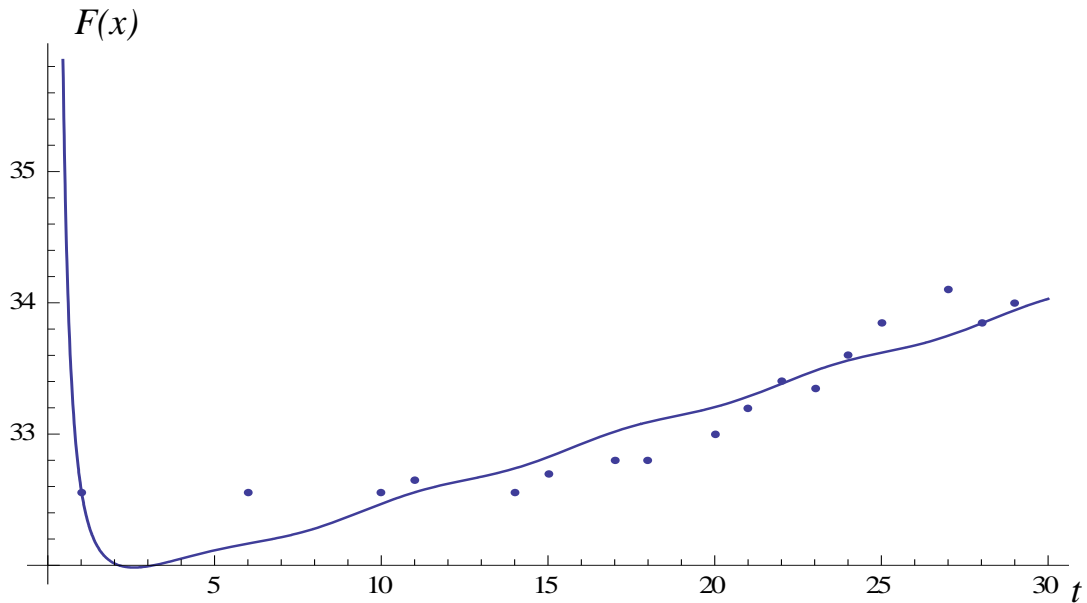


Рис. 5.3.1. Аппроксимирующая функция на основе базовых функций,  $n=17$ .

Интерполяционный многочлен степени  $n = 17$ , построенный по тем же точкам имеет вид:

$$F(x) = -92900.6157 + 168671.7742x - 104027.7666x^2 + 83.91867x^6 + 5.13689x^7 - \\ -0.18950x^8 + 0.002144 + 0.0001576x^{10} - 0.000007737x^{11} + 3.0258 \times 10^{-8}x^{12} + \\ + 9.3090 \times 10^{-9}x^{13} - 4.0577 \times 10^{-10}x^{14} + 8.4483 \times 10^{-12}x^{15} - 9.2643 \times 10^{-14}x^{16} \\ + 4.3109 \times 10^{-16}x^{17}$$

График его функции представлен на рисунке 5.3.2.

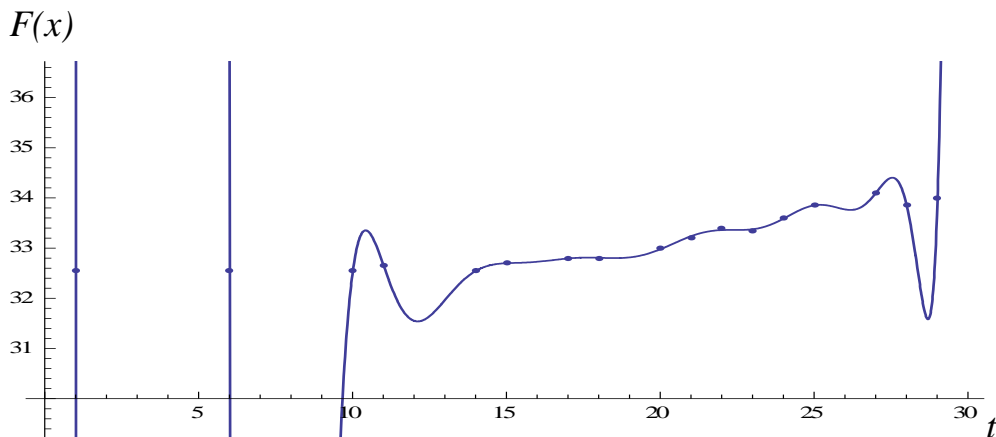


Рис. 5.3.2. Интерполяционный многочлен степени  $n=17$ .

$n = 15$ .

Аппроксимирующая функция, построенная по  $n = 15$  узловым точкам, имеет вид:

$$F(x) = 31.69 + \frac{0.82}{x^2} + 0.07x - 0.02\sin[x]$$

График ее функции представлен на рисунке 5.3.3.

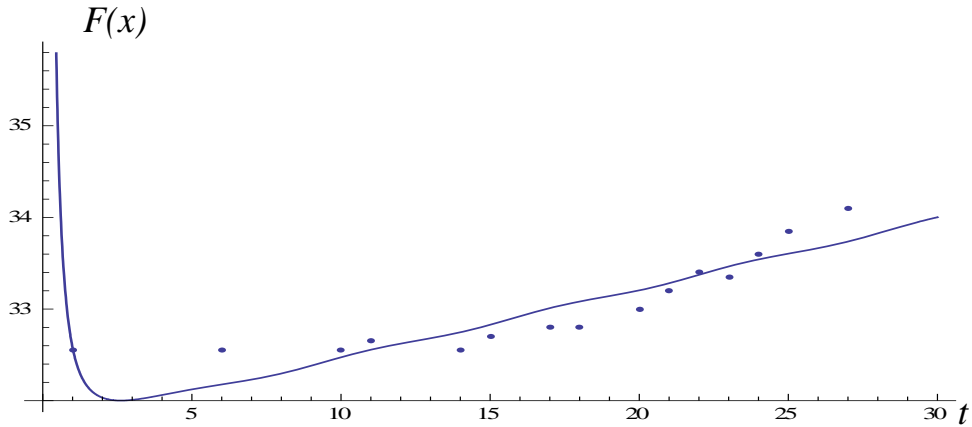


Рис. 5.3.3. Аппроксимирующая функция на основе базовых функций,  $n=15$ .

Интерполяционный многочлен степени  $n = 15$  имеет вид:

$$\begin{aligned} F(x) = & 103975.3213 - 189582.8973x + 118037.8007x^2 - 38965.8818x^3 + \\ & 7998.0047x^4 - 1097.8386x^5 + 103.7743x^6 - 6.7014x^7 + 0.2739x^8 \\ & - 0.004657x^9 - 0.0001905x^{10} + 0.00001634x^{11} - 5.6746 \times 10^{-7}x^{12} \\ & + 1.1292 \times 10^{-8}x^{13} - 1.2590 \times 10^{-10}x^{14} \\ & + 6.1421 \times 10^{-13}x^{15} \end{aligned}$$

График его функции представлен на рисунке 5.3.4.

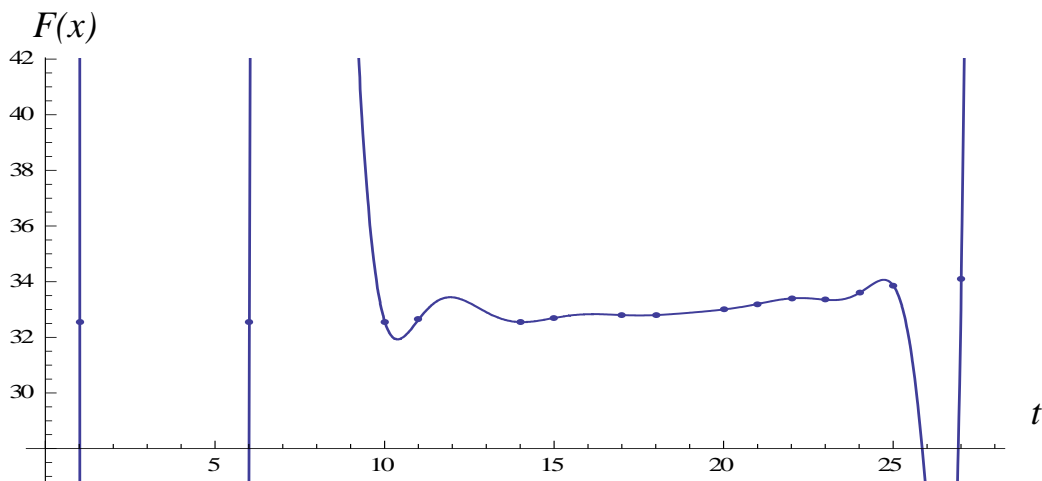


Рис. 5.3.4. Интерполяционный многочлен степени  $n=15$ .

$n = 13$ .

Аппроксимирующая функция, построенная по  $n = 13$  узловым точкам имеет вид:

$$F(x) = 31.91 + \frac{0.66}{x^2} + 0.06x - 0.07 \sin[x].$$

График ее функции представлен на рисунке 5.3.5.

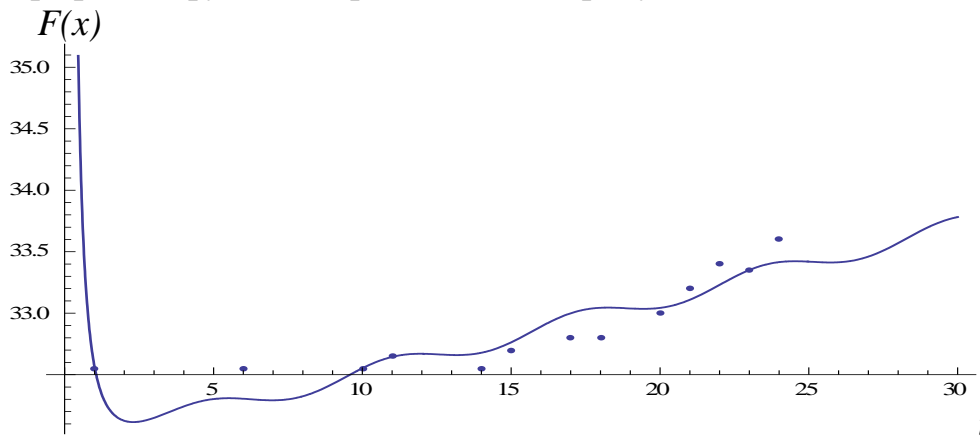


Рис. 5.3.5. Аппроксимирующая функция на основе базовых функций,  $n=13$ .

Интерполяционный многочлен степени  $n = 13$  имеет вид:

$$F(x) = 557.6435 - 958.008832x + 591.7635x^2 - 188.6841x^3 + 36.0240x^4 - 4.3657x^5 + 0.3354x^6 - 0.01473x^7 + 0.0001521x^8 + 0.00002265x^9 - 0.000001495x^{10} + 4.5006 \times 10^{-8}x^{11} - 7.04522 \times 10^{-10}x^{12} + 4.6392 \times 10^{-12}x^{13}$$

График его функции представлен на рисунке 5.3.6.

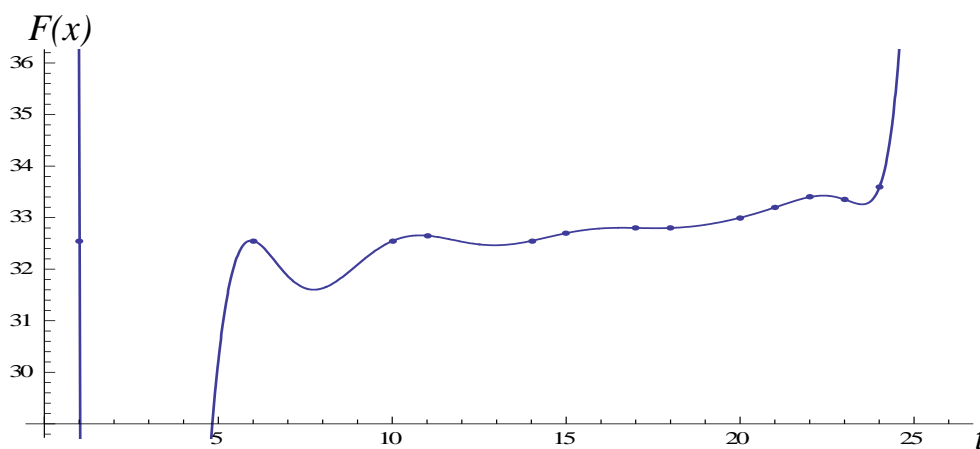


Рис. 5.3.6. Интерполяционный многочлен степени  $n=13$ .

Как видно из представленных рисунков, при интерполяции таблично заданной функции полиномами, за границей заданного интервала приближающая функция делает резкие скачки, что делает ее непригодной для построения прогнозной модели [35].

#### **§ 5.4. Методика оценки эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях.**

На основе разработанных программно-математических средств можно сформулировать новую методику оценки эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях.

Методика состоит из следующих этапов:

1. Ввод данных о произошедших компьютерных событиях.
2. Задание будущего прогнозного временного интервала.
3. Задание узловой точки будущего периода (при необходимости).
4. Выбор алгоритма (модифицированный МНК, Берг).
5. Задание параметров метода.
6. Применение алгоритма выбора базовых функций на основе минимального среднего квадратичного отклонения аппроксимирующей кривой от указанных выше статистических данных (значений в узловых точках).
7. Построение прогнозной модели, в виде непрерывной аппроксимирующей функции, интенсивности компьютерных событий.
8. Вывод значений в узловых точках.
9. Вывод значения в заданной точке будущего периода.
10. Анализ полученного решения и при необходимости переход к п.3.
11. На основе построенной функции вычисление времени эффективной обработки данных.

Предложенная методика позволяет осуществлять прогнозирование в виде построения аппроксимирующих непрерывных функций, характеризующих интенсивность компьютерных событий, приводящих к нарушению эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях. Прогнозирование осуществляется на основе ана-

лиза ограниченного количества статистических данных о произошедших негативных событиях, с применением алгоритма поиска оптимального набора базовых функций.

Разработанная методика позволяет строить прогнозные модели различных компьютерных событий, приводящих к нарушению функционирования программного обеспечения и на их основе оценивать сроки (время) эффективности процесса обработки данных в вычислительных машинах, комплексов и компьютерных сетей.

### **Выводы по главе 5.**

В данной главе рассмотрены вопросы точности прогнозирования при применении метода, описанного в главе 4. По мнению автора, вопросы величин ошибок при построении прогнозных оценок имеют первостепенное значение. К сожалению, как следует из результатов главы 1, большинство исследователей, занимающихся вопросами прогнозирования, не уделяют достаточного внимания данной проблеме. В связи с отсутствием в открытом доступе необходимых данных по компьютерным событиям, для исследования ошибок прогнозирования при применении рассматриваемого метода автором выбраны задачи прогнозирования экономических показателей, а именно: биржевых валютных курсов доллара США и курсов драгоценных металлов. Проведение указанных исследований обусловлено доступностью большого количества исходного материала и возможностью применения статистических методов. Проведен ряд экспериментальных исследований по построению методов прогнозирования рассматриваемых показателей в виде непрерывных аппроксимирующих функций и получены их прогнозные значения на период от одного временного интервала до 180, при этом временной интервал может составлять часы, дни, неделю и так далее. Проведены исследования ошибок предложенной процедуры. В частности, на основе построения статистического критерия Пирсона показано, что эмпирическое распределение ошибок прогнозирования в рассматриваемых примерах хорошо согласо-

ется с законом нормального распределения, что, в свою очередь, позволяет получить верхнюю границу для указанных ошибок.

В дополнение к перечисленным результатам, приведены результаты сравнения методов построения прогнозных функций предлагаемым методом и методом полиномов Лагранжа. Показано, что применение последнего метода для прогнозирования не представляется возможным в связи с резкими скачками прогнозной функции за границей интервала данных значений.

В завершении главы 5, на основе прогнозирования компьютерных событий, сформулирована методика исследования эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях, позволяющая строить оценки параметра, характеризующего процесс обработки данных в вычислительных машинах, комплексах и компьютерных сетях - время в течение, которого обеспечивается эффективность процесса обработки указанных данных.

## ЗАКЛЮЧЕНИЕ

Настоящая диссертационная работа посвящена актуальной задаче - разработке методов оценки эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях.

С целью решения данной задачи проведен содержательный научный анализ существующих подходов на основе оценки рисков, включающий вопросы построения моделей компьютерных событий в вычислительных машинах, комплексах и компьютерных сетях и вопросы оценки ущерба при их реализации. Исследованы существующие решения в части прогнозирования состояний динамических процессов и методов оценки их эффективности.

В ходе проведенного анализа установлено, что для применения рассмотренных выше методов прогнозирования необходима обработка большого количества статистических данных. Кроме того, ни один из рассмотренных методов не предполагает построения непрерывной прогнозной функции от времени, которая необходима для нахождения оценок эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях. И, что весьма существенно, при построении прогнозных оценок более, чем на 15 временных интервалов (дней, недель, месяцев), ошибки прогнозирования составляют более 10%.

С целью развития методов оценки эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях проведены исследования по построению новых теоретических и экспериментальных моделей компьютерных событий и оценке вероятности их реализации. В частности, построены: вероятностные модели активных компьютерных событий в результате взаимодействия программных систем; модель управления рисками учитывающие нарастающую сумму ущерба; модели активного компьютерного события в условиях ограниченных возможностей противодействия, и модель попыток доступа к элементам программного обеспечения вычислительных машин, комплексов и компьютерных сетей.

Проведены экспериментальные исследования по прогнозированию компьютерных событий в вычислительных машинах, комплексах и компьютерных сетях на основе построения непрерывных аппроксимирующих функций. Приведены результаты прогнозирования активных событий в вычислительных машинах, комплексах и компьютерных сетях; негативных событий связанных с несанкционированными операциями со счетами юридических лиц с использованием платежных карт; негативных событий связанных с несанкционированными операциями со счетами юридических лиц; динамики появления веб-угроз; динамики появления уязвимостей в программных системах; сетевых активных событий; компьютерных активных событий, направленных на перехват управления и использования вычислительных ресурсов данной программной системы для майнинга криптовалют; динамики негативных событий, связанных с ущербом от проведенных несанкционированных операций со счетами юридических лиц с использованием платежных карт.

Полученные результаты применяются для разработки методов оценки параметра, характеризующего процесс обработки данных в вычислительных машинах, комплексах и компьютерных сетях - время в течение, которого обеспечивается эффективность процесса обработки указанных данных.

С целью разработки процедуры прогнозирования предложен метод, основанный на построении непрерывных аппроксимирующих функций исследуемых параметров, вычисляемых по их значениям за предыдущий период. Разработано алгоритмическое и программное обеспечение для реализации процедуры построения указанных функций. Данное программное обеспечение предусматривает возможность выбора оптимального решения. Проведены экспериментальные исследования ошибок предложенного метода с применением математической статистики. Данные исследования проведены на примере прогнозирования ряда технических и экономических показателей: параметров вычислительной техники, биржевого курса валют и драгоценных металлов. Проведенные исследования показали, что во всех рассмотренных



случаях значения относительных ошибок прогнозирования не превышали 5,85%.

По мнению автора, проведенные исследования внесут свой вклад в решение актуальной задачи разработки методов оценки параметров, характеризующих эффективность процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях.

## Основные результаты и выводы

1. Разработаны вероятностные модели и методы исследования компьютерных событий, возникающих в процессе обработки данных в вычислительных машинах, комплексах и компьютерных сетях и получены оценки возможности их реализации. Специфика разработанных методов позволяет рассчитать введенный в данной работе важный параметр эффективности обработки данных в вычислительных машинах, комплексах и компьютерных сетях – время, в течение которого в процессе обработки данных обеспечивается целостность информации и возможность доступа к ней.

2. Разработан метод прогнозирования состояния динамического процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях, основанный на построении непрерывных аппроксимирующих функций, позволяющий делать обоснованные прогнозы на более длительные промежутки времени и снижающий ошибки прогнозирования более чем в 2 раза по сравнению с известными методами. Указанный метод является универсальным и, в частности, не зависит от природы входных данных.

3. Разработан оригинальный алгоритм построения непрерывных функций прогнозирования и его программная реализация, позволяющая эффективно строить данные функции в различных режимах.

4. Проведены экспериментальные исследования точности прогнозирования интенсивности появления компьютерных событий, которые исследовались на отрезке экстраполяции, и подтвердившие относительно небольшие значения ошибок прогноза.

5. Для значений ошибок прогнозирования впервые получены теоретические оценки вероятности превышения ими заданной величины.

6. Разработана методика оценки эффективности процесса обработки данных в вычислительных машинах, комплексах и компьютерных сетях, основанная на построении непрерывной временной функции рисков. Преимуществом разработанной методики является ее независимость от природы

риска. Тем самым, данный подход может применяться для прогнозирования различных событий в вычислительных машинах, комплексах и компьютерных сетях.

## ГЛОССАРИЙ

**Информация** – сведения (сообщения, данные), независимо от формы их представления.

**Информационная система (ИС)** – совокупность содержащейся в базах данных информации и обеспечивающих ее обработку информационных технологий и технических средств.

**Под компьютерной системой (КС)** понимается любое устройство или группа взаимосвязанных, или смежных устройств, из которых одно или более, действуя в соответствии с программой, осуществляет автоматизированную обработку данных.

**Под событием** (компьютерным событием) понимается любое явление, выходящее за рамки штатной работы вычислительных машин, комплексов и компьютерных сетей, прямо, косвенно или потенциально, ведущее к остановке процессов системы или негативно отражающееся на качестве ее функционирования.

**Под активным событием** понимается взаимодействие программных систем, при котором одна из них осуществляет направленное воздействие на другую.

**Под эффективностью процесса обработки (передачи, хранения) данных** понимается свойство указанного процесса, при котором обеспечивается целостность информации и возможность доступа к ней.

**Несанкционированный доступ (НСД) к информации** – доступ к информации или действия с информацией, нарушающие правила разграничения доступа с использованием штатных средств, предоставляемых средствами вычислительной техники или АС; преднамеренное обращение субъекта к компьютерной информации, доступ к которой ему не разрешен, независимо от цели обращения.

**Политика управления событиями** – набор законов, правил и норм поведения, определяющих, как организация обрабатывает и распространяет информацию.

**Информационные ресурсы** – отдельные документы и отдельные массивы документов, документы и массивы документов в информационных системах (библиотеках, архивах, фондах, банках данных, других информационных системах).

**Собственник информационных ресурсов, информационных систем, технологий и средств их обеспечения** – субъект, в полном объеме реализующий полномочия владения, пользования, распоряжения указанными объектами.

**Владелец информационных ресурсов, информационных систем, технологий и средств их обеспечения** – субъект, осуществляющий владение и пользование указанными объектами и реализующий полномочия распоряжения в пределах, установленных законом.

**Пользователь (потребитель) информации** – субъект, обращающийся к информационной системе или посреднику за получением необходимой ему информации и пользующийся ею.

**Угроза** – это потенциальная возможность определенным образом нарушить эффективность процесса обработки данных.

**Метод наименьших квадратов (МНК)** — математический метод, применяемый для решения различных задач, основанный на минимизации суммы квадратов отклонений некоторых функций от искомым переменных.

## ЛИТЕРАТУРА

1. ГОСТ Р ИСО/МЭК 13335-1:2006. Информационные технологии. Методы и средства обеспечения безопасности. Концепция и модели менеджмента безопасности информационных и телекоммуникационных технологий.
2. ГОСТ Р ИСО/МЭК 13335-3:2007. Информационные технологии. Методы и средства обеспечения безопасности. Методы менеджмента безопасности информационных технологий.
3. ГОСТ Р ИСО/МЭК 13335-4:2007 Информационные технологии. Методы и средства обеспечения безопасности. Выбор защитных мер.
4. ГОСТ Р ИСО/МЭК 13335-5:2007. Информационные технологии. Методы и средства обеспечения безопасности. Руководство по менеджменту безопасности сети.
5. ГОСТ Р ИСО/МЭК 27000-2012 Информационная технология (ИТ). Методы и средства обеспечения безопасности. Системы менеджмента информационной безопасности.
6. ГОСТ Р ИСО/МЭК 17799-2005. Информационная технология. Практические правила управления информационной безопасностью.
7. ГОСТ Р 58771-2019. Национальный стандарт Российской Федерации. Менеджмент риска. Технологии оценки риска.
8. Астахов А.М. Искусство управления информационными рисками. – М.– ДМК Пресс. – 2010. – 312 с.
9. Айвазян С.А. Мхитарян В.С. Прикладная статистика и основы эконометрики. М.: ИНИТИ, 1998.
10. Боровков А.А. Математическая статистика. – М. – Наука. – 1984. – 471с.
11. Болотова Л.Р., Математические методы статистики и нелинейной динамики для оценки валютных рисков на базе предпрогнозного анализа. – Дис. к. э. н. – 2005. – Черкесск. – 193 с.

12. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление: Пер. с англ. М.:МИР, 1974, Вып. 1,2.
13. Боровиков В.П., Ивченко Г.И. Прогнозирование в системе STATISTICA в среде Windows. Основы теории и интенсивная практика на компьютере. М.: Финансы и статистика, 2006.
14. Боровков А.А. Теория вероятностей. М: Наука, 1986. 431 с.
15. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1965. – 35 с.
16. Боярский А.Я., и др. Математическая статистика для экономистов. – М.: Статистика, 1979. – 253 с.
17. Батаронов И.Л. Оценка и регулирование рисков обнаружение и предупреждение компьютерных атак на инновационные проекты [Текст] / И.Л. Батаронов, А.В. Паринов, К.В. Симонов // Информация и безопасность. – 2013. – Т. 16. – Вып. 2. – С. 243–246.
18. Борисов В.И. Вероятностные аналитические модели сетевой атаки с внедрением вредоносного программного обеспечения [Текст] / В.И. Борисов, Н.М. Радько, А.А. Голозубов, И.Л. Батаронов, Е.В. Ермилов // Информация и безопасность. – 2013. – Т. 16. – Вып. 1. – С. 5–30.
19. Борисов В.И. Разработка методологии оценки эффективности средств защиты беспроводных сетей группы стандартов IEEE 802.11 [Текст] / В.И. Борисов, В.Б. Щербаков, С.А. Ермаков, И.Л. Батаронов // Информация и безопасность. – 2011. – Т. 14. – Вып. 3. – С. 317–336.
20. Бурса М.В. DDOS–атаки на информационно–телекоммуникационные системы: управление рисками [Текст] / М.В. Бурса, Ю.Г. Пастернак // Информация и безопасность. – 2013. – Т. 16. – Вып. 2. – С. 255–256.
21. Бутузов В.В. К вопросу обоснования функции ущерба атакуемых систем / В.В. Бутузов, А.В. Заряев // Информация и безопасность. – 2013. – Т. 16. – Вып. 1. – С. 47–54.

22. Венцель Е.С. Теория вероятностей: Учебн. для вузов, М.: Высшая школа, 1999 г.
23. Васильев А.Н. Mathematica. – «Корона век». – СПб. –2008. – 447 с.
24. Вербик Марно. Путеводитель по современной эконометрике. Пер. с англ. М.: Научная книга, 2008.
25. Горелик А.Н., Френкель А.А. Статистические проблемы экономического прогнозирования // Статистические методы анализа экономической динамики. М.: Наука, 1983.
26. Голованова Н.Б., Кривов Ю.Г. Методические вопросы использования межотраслевого баланса в прогнозных расчетах//Взаимосвязи НТП и экономического развития: Сб.науч.тр./АН СССР. СО, ИЭиОПП. – Новосибирск, 1987. –С. 62–77.
27. Гренджер К., Хатанака М. Спектральный анализ временных рядов в экономике. Пер.с англ. – М.: Статистика., 1972. – 312 с.
28. Ермакова А.Ю., Построение модели ущерба активам организации при возникновении инцидентов проводящих к нарушению информационной безопасности /А.Ю. Ермакова, А.Б. Лось // Компьютерные науки и информационные технологии: сборник статей IX Международной научно-практической конференция «Математическое и компьютерное моделирование в экономике, страховании и управлении рисками», Саратов, 2020. – С.144–149.
29. Ермакова А.Ю. Модель компьютерной атаки в условиях ограниченных возможностей защиты и построение прогнозных моделей компьютерных инцидентов / А.Ю. Ермакова //Промышленные АСУ и контроллеры. – 2020. – №6. – С.50–57.
30. Ермакова А.Ю. Модели DDoS атак и исследование защищенности информационной системы от данного типа угроз / А.Ю. Ермакова //Промышленные АСУ и контроллеры. – 2019. – №12. – С.54–59.



31. Ермакова А.Ю. Оценка качества прогнозирования динамики изменения валютных курсов на основе построения аппроксимирующих функций / А.Ю. Ермакова // Качество. Инновации. Образование. – 2013. – №2 (93). – С.71-79.
32. Ермакова А.Ю. Исследование качества прогнозирования биржевых курсов драгоценных металлов / А.Ю. Ермакова // Качество. Инновации. Образование. – 2014. – №1 (104). – С. 49-56.
33. Ермакова А.Ю. Построение прогнозной модели динамики изменения цен на древесину / А.Ю. Ермакова // Лесной Вестник. – 2016. – № 6. – С.88–97.
34. Ермакова А.Ю. Разработка методов прогнозирования на примере анализа средств вычислительной техники / А.Ю. Ермакова // Промышленные АСУ и контроллеры. – 2017. – №1. – С.28–34.
35. Ермакова А.Ю. Об оценке точности прогнозирования состояния динамической системы методом построения аппроксимирующих функций / А.Ю. Ермакова // Промышленные АСУ и контроллеры. – 2018. – №5. – С.36–42.
36. Ермакова А.Ю. Исследование прогнозных моделей динамической системы на примере прогноза инцидентов информационной безопасности / А.Ю. Ермакова, А.Б. Лось // Компьютерные науки и информационные технологии: сборник статей Международной научно-практической конференции, Саратов, 2018. – С.144–149.
37. Ермакова А.Ю. Об одном подходе к оценке защищенности информационной системы на основе анализа инцидентов / А.Ю. Ермакова // Системы высокой доступности. – 2018. – №4. – С.32–35.
38. Ермакова А.Ю. Построение прогнозной модели изменения показателей на примере прогнозирования мировых цен на древесину // Традиционная и инновационная наука: история, современное состояние, перспективы: сборник статей Международной научно-практической конференции. (г. Пермь, 10 января 2018г.). - С.37–42.

39. Ермакова А.Ю. Построение прогнозных моделей динамики изменения экономических показателей/А.Ю. Ермакова, А.Б. Лось// Актуальные проблемы социально-экономических исследований. – Материалы III Международной научно-практической конференции 30 апреля 2013 г. Актуальные проблемы социально-экономических исследований. – С.66–72.
40. Ершов Э.Б. Об одном методе объединения частных прогнозов // Статистический анализ экономических временных рядов и прогнозирование : Уч. зап. по статистике. Т. XXII. М.: Наука, 1973.
41. Ермилов Е.В. Риск-анализ распределенных систем на основе параметров рисков их компонентов / Е.В. Ермилов, Е.А. Попов, М.М. Жуков, О.Н. Чопоров // Информация и безопасность. 2013. Т. 16. № 1. С. 123-126.
42. Жуков М.М. Построение динамической риск-модели для компонент распределенной системы на основе заданного закона распределения ущерба / М.М. Жуков, Е.В. Ермилов, О.Н. Чопоров, А.В. Бабурин // Информация и безопасность. 2012. Т. 15. № 4. С. 449-460.
43. Зайкин В.С. Применение простых цепей Маркова для прогнозирования расходов населения//Проблемы моделирования народного хозяйства, 4IV. Новосибирск, 1973. – С. 45 –47.
44. Занг В.Б. Синергитическая экономика. Время и перемены в нелинейной экономической теории. – М.: Мир, 1999. – 216с.
45. Ивахненко А.Г., Степаненко В.С. Особенности применение метода группового учета аргументов в задачах прогнозирования случайных процессов//Автоматика. –1986. – №5. – С. 3-14.
46. Ивахненко А.Г., Юрачков Ю.П. Моделирование сложных систем по экспертным данным. – М.: Радио и связь, 1987. – 119 с.
47. Иванов В.П. Математическая оценка защищенности информации от несанкционированного доступа / В.П. Иванов // Специальная техника. 2004. № 1. С. 58—64.

48. Калашников А.О. Модели и методы организационного управления информационными рисками корпораций / А.О. Калашников. М.: Эгвес, 2011. 312 с.
49. Кельтон В. Имитационное моделирование. Классика CS. 3-е изд. / В. Кельтон А. Лоу. Издательство: Питер, 2004. 848 с.
50. Корн Г. Справочник по математике / Г. Корн, Т. Корн. М: Наука, 1973. 832 с.
51. Корниенко А.А. Метод построения модели оценки рисков информационной безопасности в автоматизированной системе. Проблемы информационной безопасности. / А.А. Корниенко, А.Г. Котенко // Компьютерные системы. №4, 2003.
52. Капица С. П., Курдюмов С. П., Малинецкий Г. Г. Синергетика и прогнозы будущего. М.: Наука, 1997. – 236с.
53. Кейн Э. Экономическая статистика и эконометрия. М.: Наука, 1977, Вып. 1, 2.
54. Кемени Дж., Снелл Дж. Конечные цепи Маркова. М.: Наука, 1970. – 136 с.
55. Кендэл М. Временные ряды. Пер. с англ. Ю.П. Лукашина. –М.: «Финансы и статистика», 1979. – 198 с.
56. Кендэл М. Временные ряды. М.: Финансы и статистика, 1981.
57. Кильдинов Г.С., Френкель А.А. Анализ временных рядов и прогнозирование. М. Статистика. 1973. – 432 с.
58. Клеопатров Д.И., Френкель А.А. Прогнозирование экономических показателей с помощью метода простого экспоненциального сглаживания. – Статистический анализ экономических временных рядов и прогнозирование. – М.: Наука, 1973. –298с.
59. Кобринский Н.Е., и др. Экономическая кибернетика: Учебник для студентов вузов и фак., обучающихся по спец. «Экономическая кибернетика». – М.: Экономика, 1982. – 408 с.

60. Калашников А.И. Обзор несанкционированных переводов денежных средств за 2017 год. Электронный ресурс (режим доступа – свободный): <https://ural.ib-bank.ru/files/files/aterials2018/45%20Kalashnikov.pdf>
61. Колмогоров А.Н. Об энтропии на единицу времени как метрическом инварианте автоморфизмов. ДАН СССР, 1959. – Т.124– С.754-755
62. Кондратьев А.И. Теоретико-игровые распознающие алгоритмы. – М.: Наука, 1990. – 272 с. Кульбак С. Теория информации и статистика. – М.: Наука, 1967. – 408с.
63. Котенко Д.А. Метод оценки риска информационной безопасности на основе сценарного логико-вероятностного моделирования / Д.А. Котенко. Санкт-Петербург, 2010. С.116.
64. Куканова Н. Методы и средства анализа рисков и управление ими в ИС / Н. Куканова // Byte. Россия, 2005. № 12. С. 69—73.
65. Кулаков В.Г. Риск-анализ информационных систем / В.Г. Кулаков, Д.О. Карпеев, А.Г. Остапенко // Информация и безопасность. 2008. Вып. 1. С. 485-531.
66. Куликов С.С. Метод риск-анализа информационно-телекоммуникационных систем при атаках на их ресурсы / С.С. Куликов, В.И. Белоножкин // Информация и безопасность. 2013. Т. 16. № 1. С.143-144.
67. Куликов С.С. Параметры и характеристики риск-модели для распределения максимальных значений переменных состояния защищенности системы / С.С. Куликов, И.Д. Петров, Ю.Г. Пастернак // Информация и безопасность. 2013. Т. 16. № 1. С. 127-130.
68. Костогрызов А.И., Лазарев В.М., Любимов А.Е., Прогнозирование рисков для обеспечения эффективности систем информационной безопасности в их жизненном цикле // Правовая информатика. – 2013. – С.4-16.
69. Лизер С. Эконометрические методы и задачи. – М.: Статистика, 1971. – 141с.

70. Лось. А.Б. О предельном распределении максимума процесса скользящего суммирования (Частичных сумм Эрдеша – Реньи), Вестник московского государственного университета леса, №3(79), 2011, С.185-189.
71. Лось А.Б. Временная модель оценки риска нарушения информационной безопасности/ А.Б. Лось, А.С. Кабанов, В.И. Трунцев // Доклады ТУСУР. – Томск. – 2012. – № 1. – ч.2. – С. 87-91.
72. Лось В.П., Щербинин В.Н. Методы и алгоритмы обработки информации в информационно-аналитических системах для анализа развития событий. // «Доклады ТУСУР» №3 т. 20 2017 г.
73. Лукацкий А.В. Безопасность Commerce / А.В. Лукацкий // Системы безопасности, связи и телекоммуникаций. 2000. №4. С. 24-26.
74. Лукацкий А.В. Комплексный подход к обеспечению информационной безопасности / А.В. Лукацкий // Системы безопасности, связи и телекоммуникаций. 1998. №1. 50 с.
75. Лугачев М.И., Ляпунцов Ю.П. Методы социального прогнозирования. М.: МГУ, ТЕИС, 1999.
76. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов. М.: Финансы и статистика, 2003.
77. Макаров О.Ю. К вопросу построения модели риск-анализа выживаемости распределенных автоматизированных информационных систем / О.Ю. Макаров, Д.Г. Плотников, А.С. Рогозина // Информация и безопасность. 2013. Т. 16. № 2. С. 265-266.
78. Максимов В.И. Моделирование риска и рисков ситуаций: Учебное пособие / В.И. Максимов, О.И. Никонов. Екатеринбург: ГОУ ВПО УГТУ – УПИ, 2004.
79. Малишевский А.В., Качественные модели в теории сложных систем / А.В. Малишевский. М.: Наука, 1998. 432с.
80. Мельников В.В. Безопасность информации в автоматизированных системах / В.В. Мельников. Издательство: Финансы и статистика, 2003. 368 с.

81. Морозова Т.Г., Пикулькин А.В., Тихонов В.Ф., и др. Прогнозирование и планирование в условиях рынка. Учеб. Пособие для вузов. Под ред. Т.Г. Морозовой, А.В. Пикулькина. – М.:ЮНИТИ-ДАНА, 1999. – 318 с.
82. Менжулин Р.В. Распределенные платежные системы на основе банковских карт с магнитной полосой: моделирование и регулирование рисков несанкционированного доступа к информации, Автореферат диссертации, Воронеж, 2011 г
83. Мхитарян В.С., Л.И. Трошин Методология статистического исследования социально-экономических процессов. Исследование влияния номинальных факторов на результативный признак методами дисперсионного анализа М.:ИНТИ, 2012.
84. Остапенко А.Г. К вопросу об оценке выживаемости информационных систем инновационного характера/ А.Г. Остапенко, Д.Г. Плотников //Информация и безопасность: Регион. науч.-техн. журнал. – Воронеж. –2012. – № 3. – С. 413-417.
85. Остапенко А.Г. К вопросу об оценке ущерба в жизнестойкости атакуемых распределенных информационных систем: Развитие методического обеспечения / А.Г. Остапенко, Н.М. Радько, Д.Г. Плотников //Информация и безопасность. – 2012. – №4. – С. 583 – 584.
86. Остапенко А.Г. Жизнестойкость атакуемых распределенных систем: оценка рисков фатальных отказов компонентов / А.Г. Остапенко, Д.Г. Плотников О.Ю. Макаров, Н.М. Тихомиров, В. Г. Юрасов; под ред. чл.-корр. РАН Д.А. Новикова //Монография. – Воронеж: Научная книга. – 2013. – 160 с.
87. Остапенко А.Г. Информационные риски в социальных сетях / А.Г. Остапенко, Л.В. Парина, В.И. Белоножкин, И.Л. Батаронов, К.В. Симонов // Монография. под ред. чл.-корр. РАН Д.А. Новикова. – Воронеж: Научная книга, 2013. – 161.
88. Остапенко А.Г. Методология риск-анализа и моделирования кибернетических систем, атакуемых вредоносным программным обеспечением: учеб.

пособие / А.Г. Остапенко, Д.Г. Плотников, С.В. Машин. Воронеж: ФГБОУ ВПО «Воронежский государственный технический университет», 2012. 189 с.

89. Остапенко А.Г. Перспективы развития методологии риск-анализа / А.Г. Остапенко, Д.О. Карпеев, Д.Г. Плотников // Информация и безопасность. 2009. №3. С. 419-424.

90. Остапенко А.Г. К вопросу об управлении рисками распределенных информационных систем / А.Г. Остапенко, Д.Г. Плотников, А.Ю. Дуплищева // Информация и безопасность. 2010. №2. С. 259-260.

91. Остапенко А.Г. Модели выживаемости атакуемой распределенной информационной системы: риск-формализация с учетом возможного ущерба / А.Г. Остапенко, Д.Г. Плотников // Информация и безопасность. 2012. №4.

92. Остапенко Г.А. Оценка рисков и регулирование защищенности распределенной платежной системы на основе банковских карт / Г.А. Остапенко, Р.В. Менжулин, Л.В. Парина // Информация и безопасность. 2011. №3. С. 359 – 381.

93. Остапенко Г.А. Основы оценки рисков и защищенности компьютерно атакуемых информационных систем и технологий: учеб. пособие / Г.А. Остапенко, Д.Г. Плотников, О.А. Остапенко. Воронеж: ФГБОУ ВПО «Воронежский государственный технический университет», 2013. 143 с.

94. Остапенко Г.А. Модели выживаемости атакуемой распределенной информационной системы: риск-формализация с учетом возможного ущерба / Г.А. Остапенко, Д.Г. Плотников, Н.Ю. Щербакова, Н.И. Баранников // Информация и безопасность. 2013. Т. 16. № 1. С.63-68.

95. Остапенко Г.А. К вопросу об оценке ущерба и атакуемых распределенных информационных систем: развитие методического обеспечения / Г.А. Остапенко, Д.Г. Плотников, Н.Ю. Щербакова, В.С. Зарубин // Информация и безопасность. 2013. Т. 16. № 1. С.141-142.

96. Остапенко Г.А. Методика риск-анализа систем, атаки на которые предусматривают внедрение вредоносного программного обеспечения: экспоненциальные модели / Г.А. Остапенко, Н.М. Радько, Д.Г. Плотников, А.А. Голозубов, А.Н. Шершень // Информация и безопасность. 2013. Т. 16. № 1. С. 99-102.
97. Остапенко Г.А. Построение функций ущерба и риска для компьютерных атак, приводящих к нарушению доступности к информации / Г.А. Остапенко, Е.В. Ермилов, А.О. Калашников // Информация и безопасность. 2013. Т. 16. № 2. С. 207-210.
98. Остапенко Г.А. Концептуальный подход к расчету и регулированию рисков нарушения актуальности информации в элементах критической информационной структуры / Г.А. Остапенко, А.Н. Шершень, А.О. Калашников // Информация и безопасность. 2013. Т. 16. № 2. С. 239-242.
99. Попов Е.А. Риск-анализ атакуемых информационно-телекоммуникационных систем с использованием нерегулярного распределения / Е.А. Попов, Н.Ю. Щербакова, Н.М. Тихомиров, А.Н. Шершень // Информация и безопасность. 2013. Т. 16. № 1. С. 39-46.
100. Радько Н.М. Риск-модели информационно-телекоммуникационных систем при реализации угроз удаленного и непосредственного доступа / Н.М. Радько, И.О. Скобелев. М: Радио Софт, 2010. 232с.
101. Радько Н.М. Концептуальные основы риск-анализа и оценки эффективности защиты информационно-телекоммуникационных систем от атак несанкционированного доступа / Н.М. Радько, И.О. Скобелев // Информация и безопасность. 2011. Т. 11. № 2. – с. 239-245.
102. Рябков В.Е. О применении методов визуального анализа многомерных данных в области защиты информации / В.Е. Рябков, А.П. Пахомов, Н.И. Баранников // Информация и безопасность. 2013. Т. 16. № 2. С. 259-260.



103. Оценка стоимости информационных активов. Электронный ресурс (режим доступа – свободный): [https://www.anti-malware.ru/analytics/Technology\\_Analysis/informational assessment](https://www.anti-malware.ru/analytics/Technology_Analysis/informational%20assessment)
104. Методика определения угроз безопасности информации в информационных системах. Электронный ресурс (режим доступа – свободный): [https://fstec.ru/component/ attachments/ download/812](https://fstec.ru/component/attachments/download/812)
105. Половников В.А., и др. Оценивание точности и адекватности моделей экономического прогнозирования // Математическое моделирование экономических процессов: Сб. науч. тр./МЭСИ– М., 1986. – С. 37–47.
106. Рожков Л.Н., Френкель А.А. Выбор оптимального параметра сглаживания в методе экспоненциального сглаживания. – Основные проблемы и задачи научного прогнозирования. – М.:Наука, 1972.- 154 с.
107. Рыбников К.К. Введение в дискретную математику и теорию решения экстремальных задач на конечных множествах. – М: Гелиос АРВ. – 2010. – 318 с.
108. Сачков В.Н. Комбинаторные методы дискретной математики – М.: Наука, 1977. - 320 с.
109. Язов Ю. К. Организация защиты информации в информационных системах от несанкционированного доступа : монография [Текст] / Ю. К. Язов, С.В. Соловьев. – Воронеж : Кварта, 2019. – 588 с.
110. Язов Ю.К. Моделирование динамики реализации угроз безопасности информации с использованием аппарата сетей Петри–Маркова [Текст] / Ю.К. Язов В.В Текунов // Информация и безопасность : научный журнал. – Воронеж : ВГТУ, 2018. – Т. 21. – №1. – С. 38–47.
111. Наиболее мощные компьютеры мира. Топ 500, Википедия, электронный ресурс, режим доступа: [https://ru.wikipedia.org/wiki/TOP 500](https://ru.wikipedia.org/wiki/TOP_500)
112. «Ландшафт угроз для систем промышленной автоматизации, Первое полугодие 2018, Kaspersky Lab ICS CERT», Электронный ресурс (режим доступа – свободный): [https:// securelist.ru/](https://securelist.ru/)

113. Сайт Лаборатории Касперского. Электронный ресурс (режим доступа – свободный): <https://securelist.ru/>
114. Cisco 2013 Annual Security Report [[http://www.cisco.com/en/US/prod/vpndevc/annual\\_security\\_report.html](http://www.cisco.com/en/US/prod/vpndevc/annual_security_report.html)] (дата обращения 20.04.2018)
115. Глобальный отчет по безопасности сетей и приложений [<http://radware.com>] (дата обращения 01.02.2019)
116. Экспертиза и аудит информационной безопасности. [Электронный ресурс]. Режим доступа: [sudexpra.ru/expertises/ekspertiza-i-audit-informatcionnoi-bezopasnosti/](http://sudexpra.ru/expertises/ekspertiza-i-audit-informatcionnoi-bezopasnosti/) (дата обращения 17.02.2018)
117. Аудит информационных систем. Регола-мониторинг. [Электронный ресурс]. Режим доступа: [spb.systematic.ru/about/news/regola-monitoring.htm](http://spb.systematic.ru/about/news/regola-monitoring.htm) (дата обращения 20.02.2018)
118. Обзор рынка SIEM-систем. [Электронный ресурс]. Режим доступа: [www.antimalware.ru/node/11637](http://www.antimalware.ru/node/11637). (дата обращения 15.03.2018)
119. Математическая статистика Учебное пособие Ивченко Г.В., Медведев Ю.И. М.: Высшая школа, 1984-248.
120. Практикум по эконометрике / Под ред. И.И. Елисейевой. М.: Финансы и статистика, 2006.
121. Френкель А.А. Прогнозирование производительности труда: методы и модели. М.: Эконометрика, 2007.
122. Эконометрика / Под ред. В.С. Мхитаряна. М.: Проспект, 2010.
123. Bates J.M., Granger C.W. J. The Combination of Forecasts. Oper. Reser. Quart. 1969. V. 20. №4.
124. Greene W.H. Econometric Analysis. 4<sup>th</sup> ed. Prentice Hall, 2000.
125. Newbold P. Statistics for Business and Economics. London, Prentice Hall, 1995.
126. Pindyck R.S., Rubinfeld D.L. Econometric Models. Economic Forecasts. 4<sup>th</sup> ed. McGraw-Hill, 1998.

127. Tukey J.W. *Exploratory Data Analysis*. Addison-Wesley. 1977.
128. Bineham J. A. *Historical Account of the Hypodermic Model in Mass Communication* / J. A. Bineham . – *Communication Monographs*, 1988. – P. 230–246.
129. Brown R.G. (1962) “Smoothing, Forecasting and Prediction of Discrete Time-Series”. Prentice-Hall, New Jersey.
130. Brown R.G. (1963) “Smoothing, Forecasting and Prediction”. Prentice-Hall, Englewood Cliffs, N.Y.
131. Cheung Y.-W., K.S.Lay (1995) “Lag Order and Critical Values of a Modified Dickey-Fuller Test”, *Oxford Bulletin of Economics and Statistics*, 57, №3, 411-419.
132. Dolado H., T. Jenkinson, S. Sosvilla-Rivero (1990) “Cointegration and Unit Roots”, *Journal of Economic Surveys*, 4, 243-273.
133. Dutt S.D., D. Ghosh (1999) “An Empirical Examination of Exchange Market Efficiency”, *Applied Economic Letters*, 6, №2, 89-91.
134. Dweyer G.P., Wallace M.S. (1992) “Cointegration and Market Efficiency”, *Journal of International Money and Finance*, 11 318-327. Elliott G., T.J. Rothenberg, J.H. Stock (1996) “Efficient Tests for an Autoregressive Unit Root”, *Econometrica*, 64, 813-836.
135. Enders W. (1995) “*Applied Econometric Time Series*”, Wiley, New York
136. Fuller W.A. (1996) *Introduction to Statistical Time Series*, 2nd Ed, Wiley, New York
137. Green W.H. (1997) “*Econometric Analysis*”. 3rd edition, Prentice-Hall.
138. Saul L.K., Jordan M.I. *Mixed Memory Markov Models: Decomposing Complex Stochastic Processes as Mixtures of Simpler Ones / Machine Learning*. –1999. –V. 37, –№ 1. P. 75–87.
139. Schiff J.L. *Cellular Automata: A Discrete View of the World*.—New York: Wiley, 2007. – 119 p.
140. Pickands, J. *Statistical inference using extreme order statistics*. *Ann. Statist.* 3, 1975. 119-131

141. Pickands, J. Multivariate extreme value distributions. Bull. I.S.I. XLIX (Book 2), 1981. 859-878
142. Davison, A.C. and Smith, R.L. Models for exceedances over high thresholds (with discussion). J.R. Statist. Soc. B 52, 1990. 393-442
143. Smith, R.L. Extreme value theory based on the  $r$  largest annual events. J. Hydrology 86, 1986. 27-43
144. Smith, R.L. Multivariate threshold methods. In Extreme Value Theory and Applications, eds. J. Galambos, J. Lechner and E. Simiu. Kluwer Academic Publishers, Dordrecht, 1994. pp. 225-248
145. Smith, R.L. and Goodman, D.J. Bayesian risk assessment. To be published in Extremes and Integrated Risk Management, edited by P. Embrechts. Risk Books, London, 2000
146. Coles, S.G. and Tawn, J.A. Modelling extreme multivariate events. J. R. Statist. Soc. B 53, 1991. 377-392

РОССИЙСКАЯ ФЕДЕРАЦИЯ



**СВИДЕТЕЛЬСТВО**  
о государственной регистрации программы для ЭВМ

**№ 2019667604**

«Программа построения прогнозных моделей динамических систем»

Правообладатель: *Ермакова Алла Юрьевна (RU)*

Автор: *Ермакова Алла Юрьевна (RU)*

Заявка № **2019666792**  
Дата поступления **11 декабря 2019 г.**  
Дата государственной регистрации  
в Реестре программ для ЭВМ **25 декабря 2019 г.**

Руководитель Федеральной службы  
по интеллектуальной собственности



 *Г.П. Ивлиев*



ООО «Смарт-Ком» ■ ИНН 5045044350 ■ КПП 504501001 ■ ОГРН 1095045000533

**Акт**  
**внедрения результатов**

диссертационной работы Ермаковой Аллы Юрьевны  
«Разработка методов прогнозирования негативных событий при компьютерной  
обработке данных»

Настоящим актом подтверждается, что основные научные и практические результаты диссертационного исследования используются российской компанией «Смарт-ком» при проектировании, разработки, внедрении и сопровождении программного обеспечения для проведения оценки эффективности процесса обработки, хранения и передачи данных в вычислительных машинах, комплексах и компьютерных сетях.

«14» мая 2021 г.

Технический директор  
ООО «Смарт-ком»  
Волович М.Е.



Национальный исследовательский университет  
«Высшая школа экономики»

Московский институт электроники и математики им. А.Н. Тихонова

Кафедра Компьютерной безопасности

### Акт

внедрения результатов  
диссертационной работы Ермаковой Аллы Юрьевны  
«Разработка методов прогнозирования негативных событий при компьютерной  
обработке данных»

Настоящим актом подтверждается, что научные и практические результаты диссертационного исследования А.Ю. Ермаковой применялись на кафедре Компьютерной безопасности МИЭМ НИУ ВШЭ при выполнении научно-исследовательской работы «Исследование возможности применения риск - ориентированного подхода для оценки качества преобразований, реализуемых криптографическими алгоритмами», реализуемой в интересах Академии криптографии Российской Федерации. В частности, указанные результаты применялись для построения активных событий, возникающих в процессе обработки, хранения и передачи данных в вычислительных машинах, комплексах и компьютерных сетях.

Заведующий кафедрой  
Компьютерной безопасности  
МИЭМ НИУ ВШЭ  
К.т.н. доцент

« 29 » апреля 2021 г.

А.Б. Лось



Подпись заверено

Специал.: по созданию и развитию  
отдела по кадровому обеспечению  
Управления персоналом  
Стоянова Т. И.





МИНОБРАЗОВАНИЯ РОССИИ  
 Федеральное государственное бюджетное образовательное учреждение  
 высшего образования  
 «МИРЭА – Российский технологический университет»  
 РТУ МИРЭА

Институт комплексной безопасности и специального приборостроения  
 Кафедра КБ-2 «Прикладные информационные технологии»

«Утверждаю»  
 Директор Института комплексной  
 безопасности и специального  
 приборостроения **Батасев А.А.**  
 « 17 » \_\_\_\_\_ 2021 г.

**Акт**

внедрения в учебный процесс материалов  
 диссертационной работы Ермаковой Аллы Юрьевны  
 «Разработка методов прогнозирования негативных событий при  
 компьютерной обработке данных»

Настоящим актом подтверждается, что основные положения и результаты диссертационного исследования внедрены в учебный процесс кафедры «Прикладные информационные технологии» МИРЭА – Российский технологический университет.

Результаты исследований, полученные Ермаковой А.Ю., используются на лабораторных (практических) занятиях по дисциплинам «Методы оценки целостности информации» и «Оценка качества программного обеспечения» для оценки эффективности процесса обработки данных при проектировании и разработки информационных систем. Указанные дисциплины читаются студентам по направлению «Информационные системы и технологии».

Заведующий кафедрой «Прикладные информационные технологии»  
 МИРЭА

О.В. Трубиенко

« 17 » \_\_\_\_\_ 2021 г.



ФЕДЕРАЛЬНОЕ АГЕНТСТВО СВЯЗИ  
 Ордена Трудового Красного Знамени федеральное государственное  
 бюджетное образовательное учреждение высшего образования  
 «МОСКОВСКИЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ  
 СВЯЗИ И ИНФОРМАТИКИ»  
 (МТУСИ)



«Утверждаю»  
 Проректор МТУСИ по научной работе  
 д.т.н., профессор

Леохин Ю.Л.

**Акт**

внедрения в учебный процесс материалов  
 диссертационной работы Ермаковой Аллы Юрьевны  
 «Разработка методов прогнозирования негативных событий  
 при компьютерной обработке данных»

Настоящим актом подтверждается, что основные положения и результаты диссертационного исследования внедрены в учебный процесс кафедры «Информатика» Московского технического университета связи и информатики (МТУСИ).

Результаты, полученные Ермаковой А.Ю., используются на лабораторных (практических) занятиях по дисциплинам «Численные методы» и «Вычислительные модели» для исследования методов построения аппроксимирующих функций при изучении метода наименьших квадратов в задачах аппроксимации функций, а также на практических занятиях по дисциплине «Моделирование вычислительных систем» для построения прогнозных моделей динамических процессов и оценки вероятности превышения заданного уровня ошибок прогнозирования при изучении средств отображения и обработки результатов моделирования. Указанные дисциплины читаются студентам технических направлений подготовки.

Заведующий кафедрой «Информатика»  
 МТУСИ, к.т.н., доцент

А.И. Волков

« 19 » мал 2021 г.