УДК 004.056.5

В.С. Полетаев

НЕЧЕТКИЙ ЛОГИЧЕСКИЙ ВЫВОД О ВОЗНИКНОВЕНИИ УГРОЗ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ НА ОСНОВЕ АНАЛИЗА ДАННЫХ ХАКЕРСКИХ ФОРУМОВ

Полетаев Владислав Сергеевич, *окончил Ульяновский государственный университет, соискатель кафедры «Телекоммуникационные технологии и сети» УлГУ. Опубликовано несколько статей в области прогнозирования угроз информационной безопасности на основе метода анализа данных хакерских интернет-форумов. [e-mail: vladis173@mail.ru].*

Аннотация

В работе описывается подход к решению задачи построения прогноза возникновения новых угроз информационной безопасности путем нечеткого логического вывода, основанного на анализе потока текстовых сообщений хакерских форумов.

Рассмотрена методика определения угроз информационной безопасности. На основании анализа существующих программных платформ для реализации интернет-форумов построена модель базы данных интернет-форума. Представлена структура сообщения интернет-форума. Сформулированы эмпирические правила функционирования хакерских форумов и правила нечетких продукций системы нечеткого логического вывода о возможных угрозах безопасности информации. Приведены статистические показатели функционирования нескольких хакерских форумов и пример нечеткого логического вывода о возникновении угроз информационной безопасности.

Ключевые слова: прогнозирование угроз, нечеткий вывод, информационная безопасность, поток сообщений, модель форума.

FUZZY INFERENCE ABOUT INFORMATION SECURITY THREATS BASED ON DATA ANALYSIS FROM HACKER FORUMS

Vladislav Sergeevich Poletaev, *graduated from Ulyanovsk State University; a degree-seeking student of the Department of Telecommunication Technologies and Networks at Ulyanovsk State University; an author of several articles in the field of forecasting of information security threats based on the method of Internet Forums' data analysis. e-mail: vladis173@mail.ru.*

Abstract

This article describes an approach to solving the problem of forecasting the risk of new threats to information security by fuzzy inference based on the analysis of text message flow from hacker forums.

The method of determining the information security threats is considered. A model of the online forum database based on the analysis of the existing software platforms for the implementation of the Internet forums is created. The online-forum message structure is presented. Empirical rules for the hacker forum functioning and rules of fuzzy system outputs of the fuzzy inference about possible information security threats are formulated. The statistical performances of several hacker forums and an example of fuzzy inference about the risk of information security threats are provided.

Key words: threat forecasting, fuzzy inference, information security, message flow, model of forum.

Введение

В настоящее время глобальные информационные сети оказывают растущее влияние на все новые и новые сферы нашей жизни. Происходит стремительное развитие единого глобального информационно-телекоммуникационного пространства, формируются новые социальные группы, оказывается существенное влияние на традиционный образ жизни людей по всему миру. К сожалению, на сегодняшний день наблюдается стремительный рост не только новых технологий, удовлетворяющих информационную потребность общества, но и количества разновидностей компьютерных атак.

В данной статье речь идет о решении задач прогнозирования новых угроз информационной безопасности. Актуальность темы определяется несколькими группами факторов. С одной стороны, системы обнаружения атак на компьютерные сети уже давно применяются как одно из средств защиты информации. С другой стороны, аналитические обзоры компаний, специализирующих-

ся в сфере интернет-технологий и защиты информации, таких как Symantec, Trustware, KasperskyLabs, показывают, что за последние несколько лет количество атак на различные информационные системы продолжает расти, а средства, которыми пользуются злоумышленники, превращаются из простых хакерских инструментов в серьезное информационное оружие [1–5].

1 Методика определения угроз информационной безопасности

Вопросы обеспечения безопасности информационных систем в настоящее время рассматриваются на государственном уровне, издаются нормативные акты, регламентирующие порядок защиты информации, создаются учреждения, координирующие указанную деятельность и осуществляющие надзор за соблюдением установленных норм. В 2015 году Федеральной службой по техническому и экспортному контролю Российской Федерации (ФСТЭК России) в результате проведенных научно-исследовательских работ разработана «Методика определения угроз безопасности информации в информационных системах». Документ устанавливает единый методический подход к определению угроз безопасности информации и разработке моделей угроз безопасности информации в государственных информационных системах, защита которых обеспечивается в соответствии с приказом ФСТЭК России от 11 февраля 2013 г. № 17, утверждающим требования к защите информации, не составляющей государственную тайну, содержащейся в государственных информационных системах [6].

В соответствии с Методикой, в целях установления того, существует ли возможность нарушения конфиденциальности, целостности или доступности информации, содержащейся в информационной системе, создается модель угроз безопасности информации. Кроме того, определяется, приведет ли нарушение хотя бы одного из указанных свойств безопасности информации к наступлению неприемлемых негативных последствий (ущерба) для обладателя информации или оператора.

В ходе эксплуатации информационной системы регулярно проводится анализ изменения угроз безопасности информации, а актуальные из них подлежат периодической переоценке.

Для определения угроз безопасности информации могут использоваться опубликованные в общедоступных источниках данные об уязвимостях, компьютерных атаках, вредоносном программном обеспечении, а также результаты специально проведенных исследований по выявлению угроз безопасности информации. Одним из источников такого рода информации являются хакерские форумы (специализированные дискуссионные интернет-ресурсы, посвященные вопросам информационной безопасности). Автоматизация процесса выявления и анализа указанных сведений является составляющим

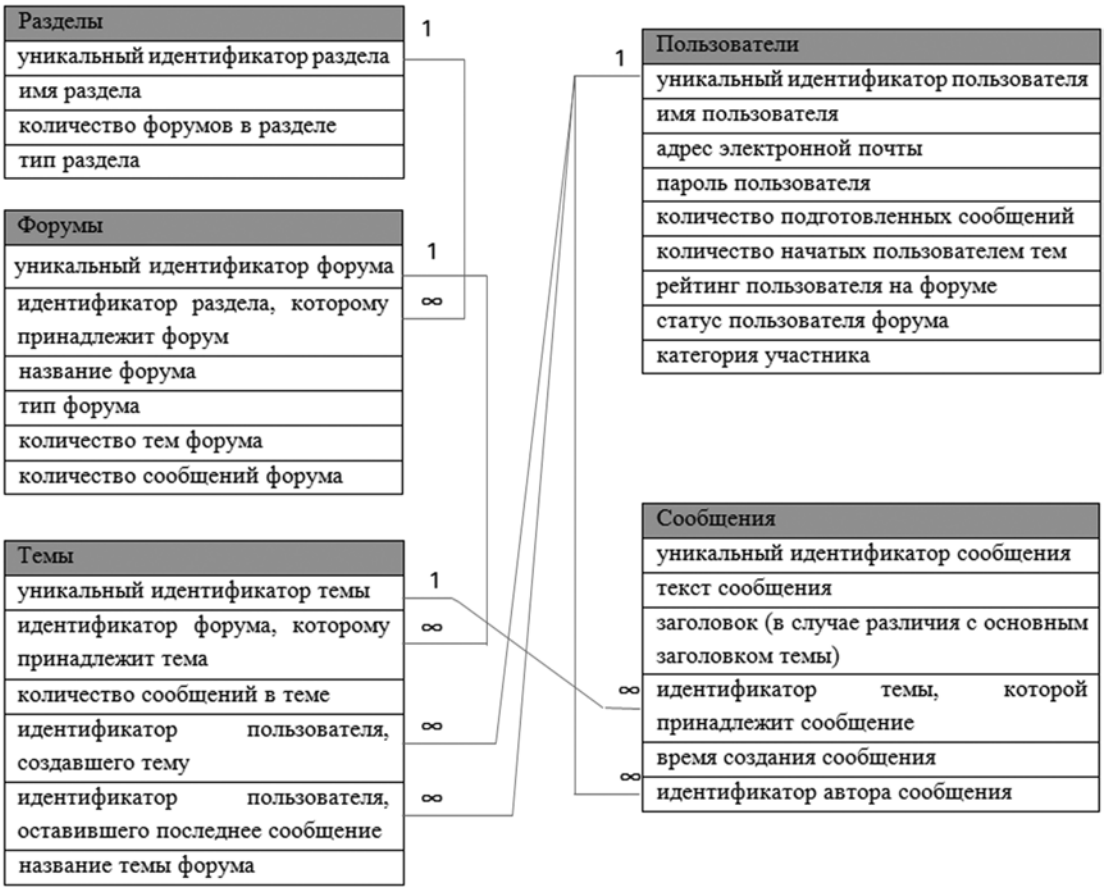

Рис. 1. Модель базы данных интернет-форумов

компонентом системы мер по защите данных в информационной системе.

В настоящей статье будет проведено исследование задачи построения прогноза возникновения новых угроз информационной безопасности путем нечеткого логического вывода, основанного на анализе потока текстовых сообщений хакерских форумов.

2 Форумы хакерской направленности

В глобальной сети Интернет в настоящее время существует значительное количество дискуссионных информационных ресурсов (далее – форумов), посвященных вопросам информационной безопасности и механизмам получения несанкционированного доступа к компьютерной информации. В части из них преобладают участники, заинтересованные в обмене сведениями о защите информации, в других – интересующиеся способами совершения компьютерных атак. Указанные форумы могут рассматриваться в качестве общедоступных источников данных об уязвимостях, компьютерных атаках, вредоносном программном обеспечении.

Наиболее популярные темы, обсуждаемые в настоящее время на хакерских форумах, соответствуют категориям актуальных угроз информационной безопасности [7–12].

При организации форумов, как правило, используются наиболее популярные программные платформы: Invision Power Board (IPB), vBulletin, PunBB, Simple Machines Forum (SMF), Vanilla, XenForo, phpBB. Перечисленные программные средства в своей реализации используют базы данных, отличающиеся по своей структуре. Вместе с тем общая структура базы данных (модель) форумов, содержащая информацию о текстовых сообщениях, представлена на рисунке 1.

3 Поток текстовых сообщений интернет-форумов

Исходя из анализа модели интернет-форумов каждое сообщение в отдельности представляет собой структуру, состоящую из связанных между собой элементов, показанных на рисунке 2.

Потоком текстовых сообщений является множество текстовых сообщений интернет-форумов, создаваемых пользователями с течением времени.

Представленная структура текстовых сообщений позволяет проводить их семантический и статистический анализ, учитывая принадлежность к конкретному форуму, теме форума, количеству сообщений темы форума, автору, рейтингу автора, а также времени создания.

Задача по сбору сообщений пользователей форумов является алгоритмически реализуемой, существуют компьютерные программы, обладающие указанными функциональными возможностями. Их применение позволяет реализовать программное средство, способное формировать поток текстовых сообщений различных хакерских форумов.

4 Системы нечеткого вывода об угрозах информационной безопасности

В связи с тем, что хакерские форумы представляют собой хранилища неформализованных данных из области информационных технологий и безопасности, содержат нечеткие понятия и знания, целесообразно применение для работы с ними нечеткой логики. Предпосылкой для применения нечетких моделей является наличие неопределенности, обусловленной неполнотой информации и сложностью предметной области [13–17].

Результаты анализа создаваемых на хакерских форумах сообщений могут быть использованы в качестве входных параметров для системы нечеткого вывода, прогнозирующей возникновение новых угроз информационной безопасности.

Получая нечеткий вывод о возникновении новой угрозы информационной безопасности, специалист по защите информации имеет возможность оценить степень угрозы для защищаемых им информационных ресурсов, пересмотреть модель угроз информационной безопасности и предпринять меры по нейтрализации возможных уязвимостей.

При построении нечеткого вывода (рис. 3) в качестве входных переменных будут использоваться частота возникновения новых сообщений и уровень рейтинга авторов сообщений.

Наполнению хакерских форумов новыми сообщениями характерны особенности, на основании которых возможно построение базы правил нечетких продукций. Так, при появлении новой угрозы информационной безопасности, участник форума, которому стало о ней известно, создает новую тему на форуме и

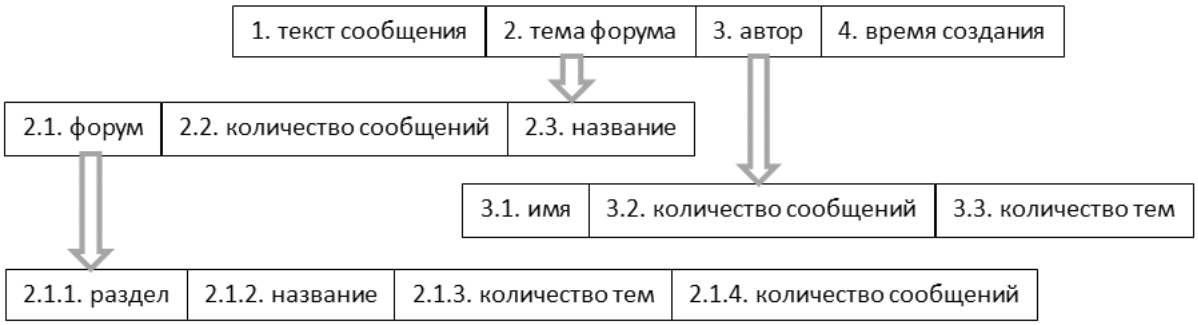

Рис. 2. Структура сообщения интернет-форума

оставляет сообщение. Другие участники форума оставляют в созданной теме сообщения, дополняющие или опровергающие предшествующие. В зависимости от степени важности информации, обсуждаемой в той или иной теме форума, различается внутренний рейтинг авторов сообщений. Как правило, при высокой значимости обсуждаемой информации в теме форума высок и рейтинг авторов сообщений. Также закономерно увеличение частоты возникновения сообщений в теме форума, где обсуждается важная информация, особенно в начальной стадии.

В этом случае эмпирические знания о рассматриваемой проблемной области могут быть представлены в форме эвристических правил (табл. 1).

Эта информация будет использоваться при построении базы правил системы нечеткого вывода. При этом,

Таблица 1

Эвристические правила функционирования хакерских форумов

№	Правило
1	Если частота появления сообщений на форуме очень высокая и уровень рейтинга авторов высокий, значит вероятность возникновения угрозы информационной безопасности очень высокая
2	Если частота появления сообщений на форуме высокая и уровень рейтинга авторов высокий, значит вероятность возникновения угрозы информационной безопасности высокая
3	Если частота появления сообщений на форуме средняя и уровень рейтинга авторов высокий, значит вероятность возникновения угрозы информационной безопасности средняя
4	Если частота появления сообщений на форуме низкая и уровень рейтинга авторов высокий, значит вероятность возникновения угрозы информационной безопасности низкая
5	Если частота появления сообщений на форуме очень низкая и уровень рейтинга авторов высокий, значит вероятность возникновения угрозы информационной безопасности очень низкая
6	Если частота появления сообщений на форуме очень высокая и уровень рейтинга авторов средний, значит вероятность возникновения угрозы информационной безопасности высокая
7	Если частота появления сообщений на форуме высокая и уровень рейтинга авторов средний, значит вероятность возникновения угрозы информационной безопасности средняя
8	Если частота появления сообщений на форуме средняя и уровень рейтинга авторов средний, значит вероятность возникновения угрозы информационной безопасности низкая
9	Если частота появления сообщений на форуме низкая и уровень рейтинга авторов средний, значит вероятность возникновения угрозы информационной безопасности очень низкая
10	Если частота появления сообщений на форуме очень низкая и уровень рейтинга авторов средний, значит вероятность возникновения угрозы информационной безопасности очень низкая
11	Если частота появления сообщений на форуме очень высокая и уровень рейтинга авторов низкий, значит вероятность возникновения угрозы информационной безопасности средняя
12	Если частота появления сообщений на форуме высокая и уровень рейтинга авторов низкий, значит вероятность возникновения угрозы информационной безопасности низкая
13	Если частота появления сообщений на форуме средняя и уровень рейтинга авторов низкий, значит вероятность возникновения угрозы информационной безопасности очень низкая
14	Если частота появления сообщений на форуме низкая и уровень рейтинга авторов низкий, значит вероятность возникновения угрозы информационной безопасности очень низкая
15	Если частота появления сообщений на форуме очень низкая и уровень рейтинга авторов низкий, значит вероятность возникновения угрозы информационной безопасности очень низкая

Рис. 3. Структура системы нечеткого вывода

в качестве одной из входных лингвистических переменных будет использоваться частота появления новых сообщений: β_1 – «частота появления сообщений», а в качестве второй входной лингвистической переменной β_2 – «уровень рейтинга авторов». В качестве выходной лингвистической переменной будет использоваться уровень угрозы информационной безопасности: β_3 – «вероятность возникновения угрозы информационной безопасности». Для сокращения записи правил будем использовать рассмотренные символические обозначения (табл. 2), при этом модификатор «очень» преобразован к значению отдельного терма.

Таблица 2

Символические обозначения термов

№ п/п	Терм	Обозначение
1	Очень высокий	ОВ
2	Высокий	В
3	Средний	С
4	Низкий	Н
5	Очень низкий	ОН

В этом случае система нечеткого вывода будет содержать 15 правил нечетких продукций следующего вида (табл. 3).

В качестве терм-множества первой лингвистической переменной будем использовать множество $T_1 =$ {«очень высокая», «высокая», «средняя», «низкая», «очень низкая»} или в символическом виде $T_1 =$ {ОВ, В, С, Н, ОН} с функциями принадлежности $\mu_{T_1}(x)$, изображенными на рисунке 4. В качестве терм-множества второй лингвистической переменной будем использовать множество $T_2 =$ {«высокий», «средний», «низкий»} или в символическом виде

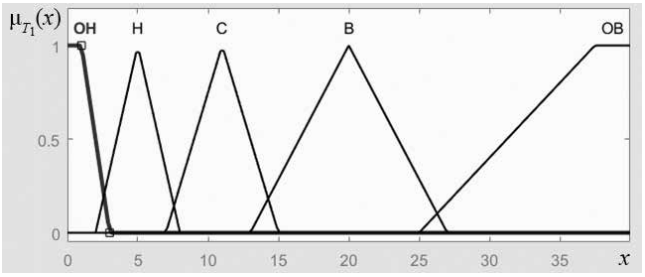

Рис. 4. График функций принадлежности термов для входной переменной «Частота появления сообщений»

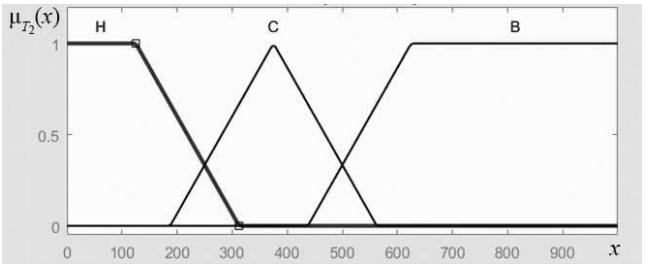

Рис. 5. График функций принадлежности термов для входной переменной «Уровень рейтинга авторов»

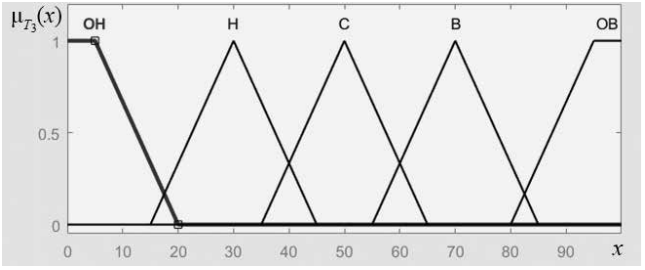

Рис. 6. График функций принадлежности термов для входной переменной «Вероятность возникновения угрозы информационной безопасности»

Таблица 3

Правила нечетких продукций

Правило_1:	ЕСЛИ «β_1 есть ОВ»	И	«β_2 есть В»	ТО	«β_3 есть ОВ»
Правило_2:	ЕСЛИ «β_1 есть В»	И	«β_2 есть В»	ТО	«β_3 есть В»
Правило_3:	ЕСЛИ «β_1 есть С»	И	«β_2 есть В»	ТО	«β_3 есть С»
Правило_4:	ЕСЛИ «β_1 есть Н»	И	«β_2 есть В»	ТО	«β_3 есть Н»
Правило_5:	ЕСЛИ «β_1 есть ОН»	И	«β_2 есть В»	ТО	«β_3 есть ОН»
Правило_6:	ЕСЛИ «β_1 есть ОВ»	И	«β_2 есть С»	ТО	«β_3 есть В»
Правило_7:	ЕСЛИ «β_1 есть В»	И	«β_2 есть С»	ТО	«β_3 есть С»
Правило_8:	ЕСЛИ «β_1 есть С»	И	«β_2 есть С»	ТО	«β_3 есть Н»
Правило_9:	ЕСЛИ «β_1 есть Н»	И	«β_2 есть С»	ТО	«β_3 есть ОН»
Правило_10:	ЕСЛИ «β_1 есть ОН»	И	«β_2 есть С»	ТО	«β_3 есть ОН»
Правило_11:	ЕСЛИ «β_1 есть ОВ»	И	«β_2 есть Н»	ТО	«β_3 есть С»
Правило_12:	ЕСЛИ «β_1 есть В»	И	«β_2 есть Н»	ТО	«β_3 есть Н»
Правило_13:	ЕСЛИ «β_1 есть С»	И	«β_2 есть Н»	ТО	«β_3 есть ОН»
Правило_14:	ЕСЛИ «β_1 есть Н»	И	«β_2 есть Н»	ТО	«β_3 есть ОН»
Правило_15:	ЕСЛИ «β_1 есть ОН»	И	«β_2 есть Н»	ТО	«β_3 есть ОН»

T_2={В, С, Н} с функциями принадлежности $\mu_{T_2}(x)$, изображенными на рисунке 5. В качестве терм-множества выходной лингвистической переменной будем использовать множество T_3= {«очень высокая», «высокая», «средняя», «низкая», «очень низкая»} или в символическом виде T_3={ОВ, В, С, Н, ОН} с функциями принадлежности $\mu_{T_3}(x)$, изображенными на рисунке 6.

При этом частота появления сообщений измеряется в единицах в сутки, средний уровень рейтинга авторов – в единицах, вероятность возникновения угрозы – в процентах.

При выборе функций принадлежности входных переменных использовались результаты анализа сообщений 10 отобранных экспертным путем хакерских форумов (rdot.org, darkmoney.cc, zloy.bz, grabberz.com, hakepok.su, nulled.io, hashcrack.in, verified.cm, hakerforum.ru, inattack.ru) в период со 2 января по 25 февраля 2018 года. В течение анализируемого периода времени создано 10491 сообщение; средняя ежесуточная частота создания сообщений: 166,5 сообщения/сутки; максимальное количество сообщений в сутки: 747 сообщений/сутки; средний рейтинг авторов сообщений: 176,8.

5 Оценка качества прогноза

Для оценки качества построенных прогнозов используются следующие показатели [18]:

- средняя абсолютная процентная ошибка прогнозирования (Mean Absolute Percent Error – MAPE):

$$MAPE = 100\% \cdot \frac{1}{h}\sum_{i=1}^{h}\left|\frac{f_{T,i} - y_{T+i}}{y_{T+i}}\right|, \quad (1)$$

где h – длина интервала (горизонт) прогнозирования;

$f_{T,i}$ – прогнозное значение временного ряда, рассчитанное в момент времени T на i шагов вперед;

y_{T+i} – истинное значение временного ряда в момент времени $T+i$;

- средняя абсолютная ошибка прогнозирования (Mean Absolute Error – MAE):

$$MAE = \frac{1}{h}\sum_{i=1}^{h}\left|f_{T,i} - y_{T+i}\right|; \quad (2)$$

- корень квадратный из средней квадратичной ошибки прогнозирования (Root Mean Squared Error – RMSE):

$$RMSE = \sqrt{\frac{1}{h}\sum_{i-1}^{h}\left(f_{T,i} - y_{T+i}\right)^2}. \quad (3)$$

Средняя абсолютная процентная ошибка является более удобным инструментом для оценки качества прогнозов, поскольку измеряется в процентах от истинного значения прогнозируемого показателя и может быть использована и как сравнительная характеристика качества прогнозов, построенных по различным моделям, и как характеристика качества прогноза конкретной модели при некотором критическом уровне ошибки прогнозирования.

Мерой точности прогноза может быть относительное число случаев к общему числу случаев, предложенное Е.М. Четыркиным [19]:

$$\eta = \frac{p}{p+q}, \quad (4)$$

где p – число прогнозов, подтвержденных фактическими данными;

q – число прогнозов, не подтвержденных фактическими данными.

Если же статистических данных о прогнозируемом показателе нет, то проблема точности рассматривается как проблема сопоставления априорных качеств или свойств, присущих альтернативным прогностическим моделям. Причем при прогнозировании статистическими методами понятия априорной точности прогноза связывают с размером доверительного интервала. В этом случае модель-прогноз считается более точной, если при одной и той же доверительной вероятности она дает более узкий доверительный интервал по сравнению с другой моделью.

Важным критерием правильности применения прогнозной модели является проверка на адекватность. Адекватными моделями считаются такие, для которых остаточная компонента имеет свойства независимости, случайности и нормальности распределения. Для проверки корреляции внутри ряда применяется критерий Дарбина-Уотсона. В соответствии с этим критерием, если величина d близка к 2, то можно считать модель регрессии достаточно адекватной:

$$d = \frac{\sum_{i=2}^{m}\left(e_i - e_{i-1}\right)^2}{\sum_{i=2}^{m} e_i^2}, \quad (5)$$

где m – длина временного ряда,

e_i – ошибка прогноза: $e_i = x_i - \bar{x}_i$.

Для оценки эффективности алгоритма прогнозирования угроз информационной безопасности на основе анализа сообщений пользователей хакерских форумов были проведены эксперименты по автоматизированному сбору сообщений 10 хакерских форумов в период со 2 января по 25 февраля 2018 года. На основе полученных данных с применением предложенного алгоритма, реализованного в информационно-аналитической системе (ИАС), произведены вычислительные эксперименты по формированию нечеткого логического вывода о возникновении угроз информационной безопасности. Полученные результаты сопоставлены с данными о выявленных угрозах, опубликованных на официальном сайте ФСТЭК России.

Статистические данные о количестве добавленных в базу данных угроз информационной безопасности ФСТЭК России в анализируемый период времени представлены на рисунке 7 и в таблице 4.

Таблица 4

Сводные данные о количестве сообщений хакерских форумов, среднем рейтинге их авторов, прогнозе ИАС возникновения угроз информационной безопасности и количестве выявленных угроз по данным ФСТЭК России в анализируемый период

№ п/п	Дата	Количество сообщений форумов	Средний рейтинг авторов сообщений	Прогноз ИАС вероятности возникновения угроз	Количество записей об угрозах в базе данных ФСТЭК России
1	02.01.2018	73	60,04110	7,56	10
2	03.01.2018	94	59,71277	7,53	19
3	04.01.2018	95	171,98950	7,42	2
4	05.01.2018	106	59,74528	8,07	0
5	06.01.2018	84	130,26430	7,86	0
6	07.01.2018	115	221,66870	8,66	1
7	08.01.2018	84	167,98100	6,88	1
8	09.01.2018	103	59,67670	7,89	16
9	10.01.2018	88	114,71480	8,36	3
10	11.01.2018	98	160,41630	7,59	1
11	12.01.2018	77	136,55710	7,68	1
12	13.01.2018	80	133,68630	7,76	0
13	14.01.2018	72	161,14860	7,06	2
14	15.01.2018	90	134,04670	7,74	2
15	16.01.2018	115	117,94960	8,66	7
16	17.01.2018	90	137,60110	7,65	0
17	18.01.2018	104	101,48170	8,80	0
18	19.01.2018	132	96,47652	19,70	4
19	20.01.2018	107	123,03080	8,14	0
20	21.01.2018	78	146,46540	7,39	0
21	22.01.2018	121	115,88180	19,50	0
22	23.01.2018	156	98,29615	20,60	0
23	24.01.2018	102	130,91180	7,86	1
24	25.01.2018	118	134,92880	14,70	3
25	26.01.2018	127	60,02520	8,86	1
26	27.01.2018	105	60,03048	8,01	1
27	28.01.2018	108	125,49810	8,20	0
28	29.01.2018	140	107,60140	24,00	1
29	30.01.2018	127	60,02520	8,86	0
30	31.01.2018	108	60,02963	8,20	0
31	01.02.2018	108	59,66019	8,20	2
32	02.02.2018	110	117,13550	8,33	2
33	03.02.2018	178	97,43933	20,20	1
34	04.02.2018	121	60,11983	9,07	0
35	05.02.2018	214	60,06776	28,30	2
36	06.02.2018	143	104,23010	22,90	2
37	07.02.2018	155	59,76323	7,53	1
38	08.02.2018	156	59,59551	7,53	0
39	09.02.2018	128	60,02500	8,80	3
40	10.02.2018	119	137,06810	16,40	1
41	11.02.2018	132	60,02424	8,53	0

Продолжение табл. 4

№ п/п	Дата	Количество сообщений форумов	Средний рейтинг авторов сообщений	Прогноз ИАС вероятности возникновения угроз	Количество записей об угрозах в базе данных ФСТЭК России
42	12.02.2018	681	59,87944	50,00	2
43	13.02.2018	145	91,27379	17,50	16
44	14.02.2018	115	106,21390	8,66	1
45	15.02.2018	168	85,92083	14,50	4
46	16.02.2018	174	87,67184	15,60	1
47	17.02.2018	146	89,62808	16,60	0
48	18.02.2018	144	98,22847	20,60	0
49	19.02.2018	226	79,18451	32,60	1
50	20.02.2018	224	60,01563	30,00	1
51	21.02.2018	225	60,05956	30,00	7
52	22.02.2018	201	60,01642	9,81	0
53	23.02.2018	140	92,82357	18,00	4
54	24.02.2018	166	86,17892	15,00	0
55	25.02.2018	193	60,01762	8,20	0

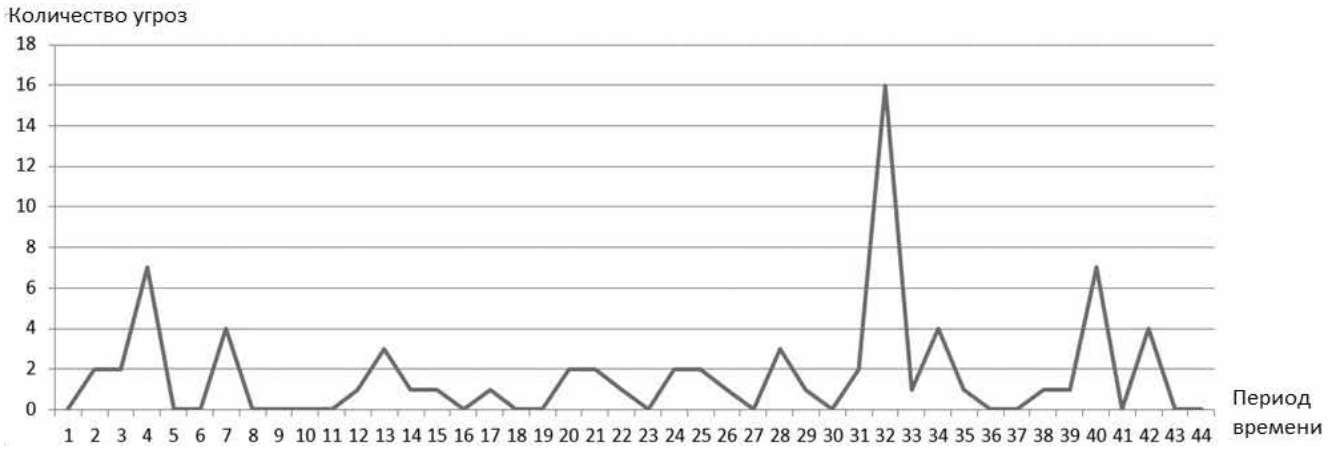

Рис. 7. Количество угроз в базе данных ФСТЭК России

Данные о количестве сообщений пользователей хакерских форумов представлены на рисунке 8 и в таблице 4.

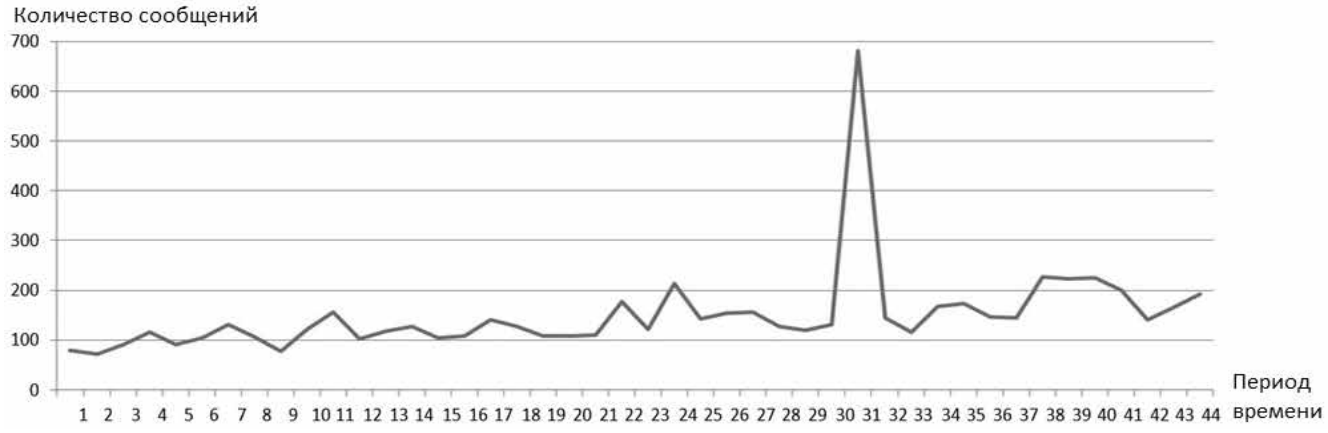

Рис. 8. Количество сообщений хакерских форумов

График среднего рейтинга авторов сообщений хакерских форумов на основе данных, указанных в таблице 4, представлен на рисунке 9.

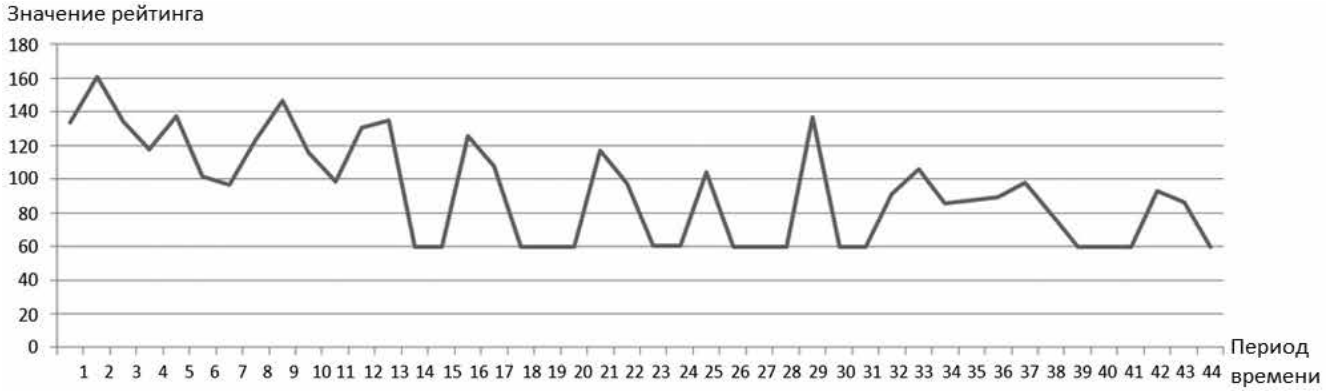

Рис. 9. Средний рейтинг авторов сообщений

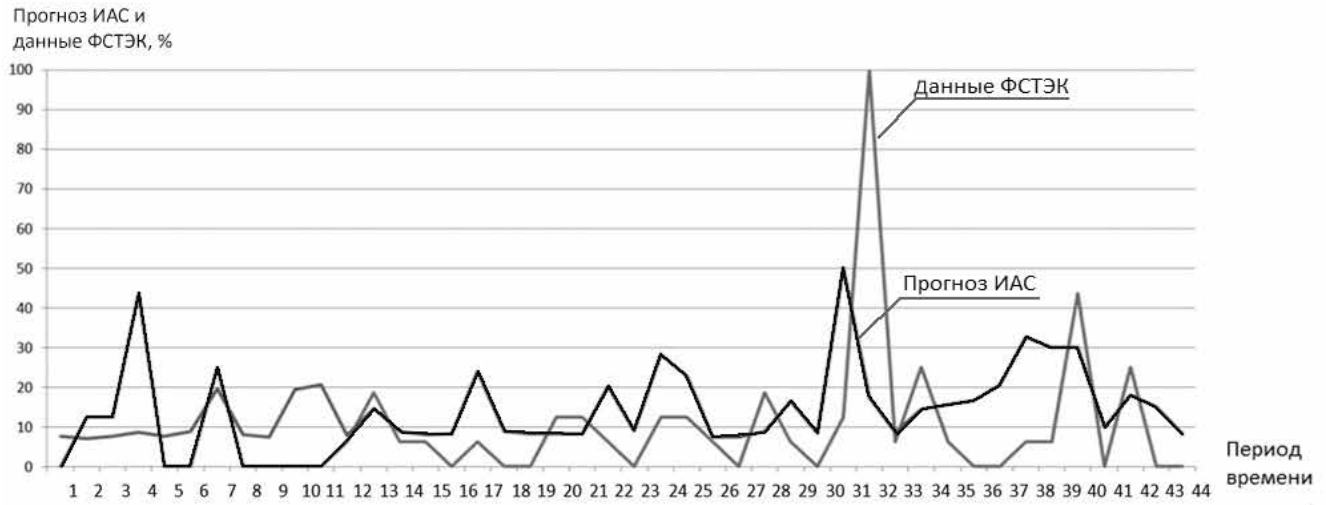

Рис. 10. Оценка соответствия прогноза ИАС данным ФСТЭК России

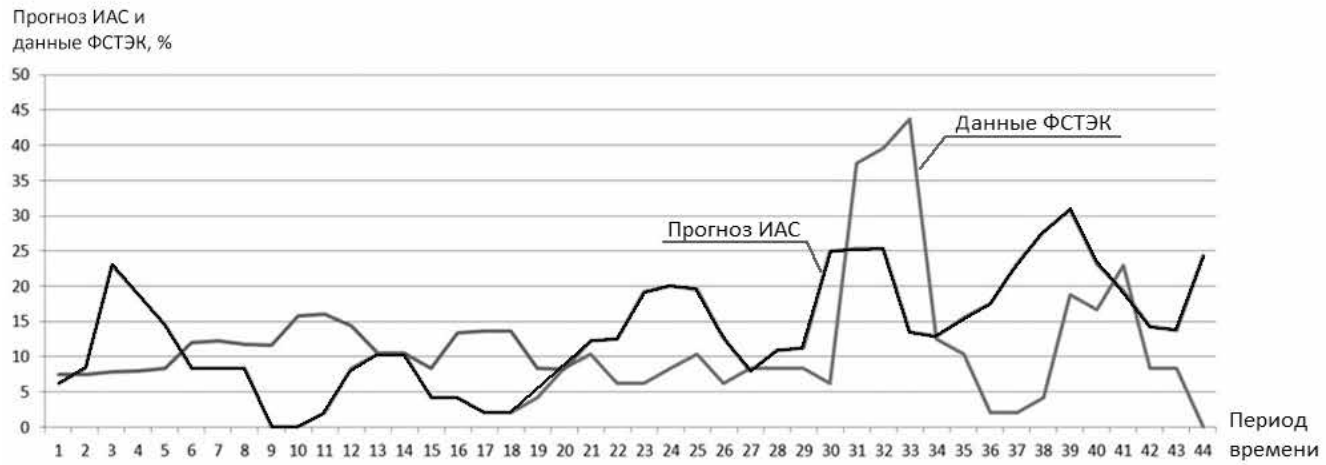

Рис. 11. Сглаживание динамического временного ряда методом скользящих средних (период скольжения 3)

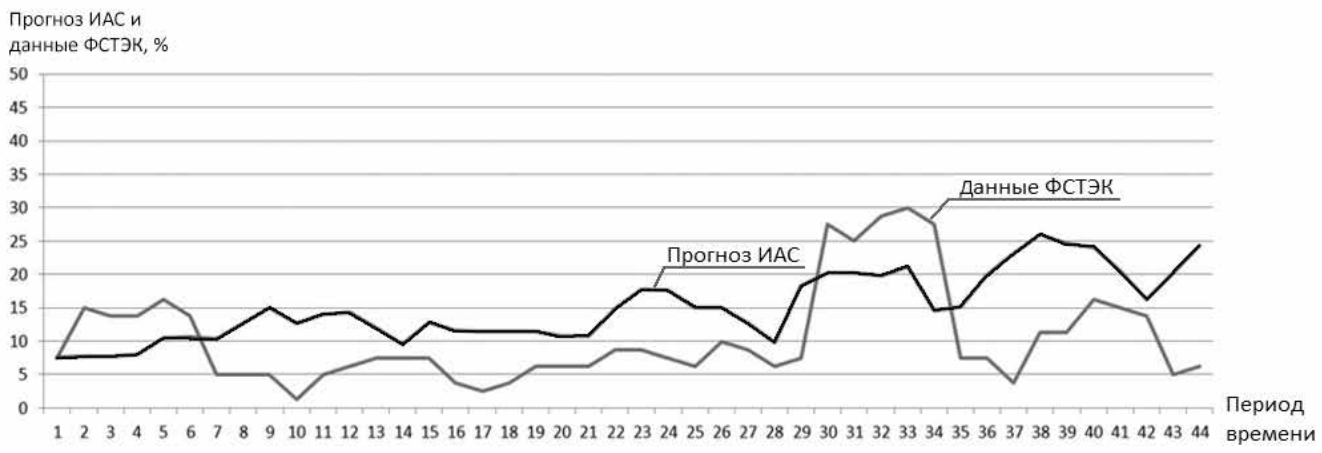

Рис. 12. Сглаживание динамического временного ряда методом скользящих средних (период скольжения 5)

На основании полученных результатов для качественного анализа прогноза в одних координатных осях построены графики реально выявленных (добавленных в базу данных ФСТЭК России) угроз информационной безопасности и прогнозных данных ИАС о возникновении угроз (рис. 10).

В связи с тем, что анализируемые значения имеют резкие колебания, для оценки изменения общего уровня угроз информационной безопасности можно применить методы определения трендов, одним из которых является сглаживание динамического временного ряда методом скользящих средних. Для этого фактические уровни ряда заменяются скользящими средними при условии, что выбран период скольжения: 3, 5, 7 или другое нечетное число. Скользящие средние представляют собой средние уровни за определенные периоды времени (3, 5, 7) путем последовательного передвижения начала периода на единицу времени.

Результаты применения метода скользящих средних к значениям прогноза ИАС возникновения угроз информационной безопасности и количеству выявленных угроз по данным ФСТЭК России в анализируемый период с периодом сглаживания 3 и 5 представлены на рисунках 11 и 12 соответственно.

Для оценки полученных результатов произведены расчеты показателей MAPE, MAE, RMSE (по формулам 1, 2, 3 соответственно) для значений прогнозов ИАС возникновения угроз информационной безопасности и количества выявленных угроз по данным ФСТЭК России в анализируемый период, а также рассчитанным на их основе сглаженным временным рядам с периодом сглаживания 3 и 5. Результаты представлены в таблице 5.

Для проверки адекватности предлагаемой модели согласно критерию Дарбина-Уотсона по формуле 5 рассчитано значение $d = 2{,}461385381$. В соответствии с этим критерием, если величина d близка к 2, то можно считать модель адекватной.

Рассчитаны значения показателя точности прогноза η, предложенного Четыркиным Е.М. (4), для доверительных интервалов 20, 15, 10 %. Результаты расчетов приведены в таблице 6.

Представленные в таблицах 5 и 6 показатели позволяют сделать вывод о том, что результаты прогнозирования информационно-аналитической системы в большинстве случаев подтверждаются данными базы угроз ФСТЭК России. Результаты экспериментов указывают на существующие временные расхождения в 1–2 дня между активизацией обсуждения вопросов информа-

Таблица 5

Показатели качества прогноза

Показатель	Экспериментальные данные	Сглаживание с периодом 3	Сглаживание с периодом 5
MAPE	93,21362	145,04050	118,737500
MAE	25,86274	15,80015	12,042810
RMSE	15,14075	11,26616	9,986906

Таблица 6

Значения показателя точности прогноза Четыркина Е.М.

Доверительный интервал	Экспериментальные данные	Сглаживание с периодом 3	Сглаживание с периодом 5
20 %	0,685185	0,692308	0,692308
15 %	0,555556	0,557692	0,480769
10 %	0,518519	0,192308	0,115385

ционной безопасности на хакерских форумах и добавлением записей о выявленных угрозах в базу данных ФСТЭК России. Наиболее показательным примером является существенное увеличение количества сообщений хакерских форумов (более чем в 4 раза превышены средние ежесуточные показатели) 12.02.2018 и последовавшее добавление 13.02.2018 в базу данных ФСТЭК 16 новых записей о выявленных угрозах безопасности, большая часть из которых имела высокий и критический уровень опасности и была связана с уязвимостями в прикладном программном обеспечении Microsoft Windows. В анализируемый период наблюдались превышения средних ежесуточных показателей активности пользователей хакерских форумов в период с 19 по 22 февраля 2018 года, связанные с появлением 7 угроз безопасности информации, добавленных в базу данных ФСТЭК 21.02.2018 и имевших критический уровень опасности.

Улучшению качества прогноза возникновения угроз информационной безопасности при помощи систем нечеткого логического вывода может способствовать применение более точных правил нечетких продукций, а также увеличение количества входных переменных, характеризующих закономерности изменения потока сообщений хакерских форумов в зависимости от возникновения новых угроз информационной безопасности. Большое значение имеет определение функций принадлежности входных и выходных переменных системы нечеткого логического вывода, при определении которых следует учитывать статистические показатели потока сообщений анализируемых хакерских форумов.

Заключение

Таким образом, в настоящей статье представлен подход к решению задачи построения прогноза возникновения новых угроз информационной безопасности путем нечеткого логического вывода, основанного на анализе потока текстовых сообщений хакерских форумов. Получаемые результаты могут быть использованы для оценки вероятности возникновения угроз информационной безопасности, принятия решения о необходимости пересмотра модели угроз и планирования мер по нейтрализации возможных уязвимостей.

СПИСОК ЛИТЕРАТУРЫ

1. Kaspersky Security Bulletin 2015 / KasperskyLab ZAO, 2015. – URL: http://www.securelist.com/.

2. Symantec Report on AttackKits and Malicious Websites-2015. – URL: http://scm.symantec.com.

3. TrustwareGlobalSecurityReport 2014 / Trustwave, 2014. – URL: http://www.trustwave.com/GSR.

4. Гмурман В.Е. Теория вероятностей и математическая статистика : учеб. пособие для вузов. – Изд. 7-е, стер. – М. : Высш. Шк., 2000. – 479 с.

5. Лукацкий А. Обнаружение атак. – СПб. : БХВ-Петербург, 2001. – 624 с.

6. Приказ ФСТЭК России от 11.02.2013 № 17 (ред. 15.02.2017) «Об утверждении требований к защите информации, не составляющей государственную тайну, содержащейся в российских информационных системах» / ФСТЭК России, 2018. – URL: http://fstec.ru/normotvorcheskaya/akty/53-prikazy/702/.

7. ГОСТ Р 50922-2006. Защита информации. Основные термины и определения. – М. : ИПК Издательство стандартов, 2006. – 12 с.

8. Крат Ю.Г., Шрамкова И.Г. Основы информационной безопасности : учеб. пособие – Хабаровск : Изд. ДВГУПС, 2008. – 112 с.

9. Блинов А.М. Информационная безопасность : учеб. пособие. – СПб. : СПБГУЭФ, 2010. – 96 с.

10. Информационная безопасность : учебник для студентов вузов. – М. : Академический Проект, Гаудеамус, 2004. – 544 с.

11. Гатчин Ю.А., Сухостат В.В. Теория информационной безопасности и методология защиты информации. – СПб. : СПбГУ ИТМО, 2010. – 98 с.

12. Макаренко С.И. Информационная безопасность : учебное пособие для студентов вузов. – Ставрополь : СФ МГГУ им. М.А. Шолохова, 2009. – 372 с.

13. Зайченко Ю.П. Нечеткие модели и методы в интеллектуальных системах : учебник для вузов. – Киев : Издательский Дом «Слово», 2008. – 344 с.

14. Усков А.А. Принципы построения систем управления с нечеткой логикой // Приборы и системы. Управление, контроль, диагностика. – 2004. – № 6. – С. 7–13.

15. Тэрано Т., Асаи К., Сугено М. Прикладные нечеткие системы. – М. : Мир, 1993. – 368 с.

16. Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. – СПб. : БХВ-Петербург, 2005. – 716 с.

17. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы / пер. с польск. И.Д. Рудинского. – М. : Горячая линия – Телеком, 2006. – 452 с.

18. Кизбикенов К.О. Прогнозирование и временные ряды. – Барнаул : АлтГПУ, 2017. – 14 с.

19. Четыркин Е.М. Статистические методы. – М. : Статистика, 1977. – 29 с.

REFERENCES

1. Kaspersky Security Bulletin 2015. KasperskyLab ZAO, 2015. Available at: http://www.securelist.com/.

2. Symantec Report on AttackKits and Malicious Websites-2015. Symantec. Available at: http://scm.symantec.com.

3. TrustwareGlobalSecurityReport 2014. Trustwave, 2014. Available at: http://www.trustwave.com/GSR.

4. Gmurman V.E. *Teoriia veroiatnostei i matematicheskaia statistika. Ucheb. posobie dlia vuzov. Izd. 7-e, ster.* [The Theory of Probabilities and Mathematical Statistics. Handbook for Higher Schools. 7th Edition]. Moscow, Vysshaia Shkola Publ., 2000. 479 p.

5. Lukatskiy A. *Obnaruzhenie atak* [Attack Detection]. St. Petersburg, BHV-Peterburg Publ., 2001. 624 p.

6. Prikaz FSTEK Rossii ot 11.02.2013 No. 17 (red. 15.02.2017) "Ob utverzhdenii trebovanii k zashchite informatsii, ne sostavliaiushchei gosudarstvennuiu tainu, soderzhashcheisia v rossiiskikh informatsionnykh sistemakh" [The Low of the Federal Service for Technology and Export Control of Russia on Information Security, dated 11.02.2013, no. 17 (revised 15.02.2017)]. FSTEK Rossii, 2018. Available at: http://fstec.ru/normotvorcheskaya/akty/53-prikazy/702/.

7. *GOST R 50922-2006. Zashchita informatsii. Osnovnye terminy i opredeleniia* [National State Standard. Protection of Information. Basic Terms and Definitions]. Moscow, IPK Izdatelstvo standartov Publ., 2006. 12 p.

8. Krat Iu.G., Shramkova I.G. *Osnovy informatsionnoi bezopasnosti. Ucheb. posobie* [Fundamentals of Information Security. Tutorial]. Khabarovsk, DVGUPS Publ., 2008. 112 p.

9. Iinov A.M. *Informatsionnaia bezopasnost. Ucheb. posobie* [Information Security. Tutorial]. St. Petersburg, SPBGUEF Publ., 2010. 96 p.

10. Informatsionnaia bezopasnost. Uchebnik dlia studentov vuzov [Information Security. Textbook for Students of Higher Schools]. Moscow, Akademicheskii Proekt, Gaudeamus Publ., 2004. 544 p.

11. Gatchin Iu.A., Sukhostat V.V. *Teoriia informatsionnoi bezopasnosti i metodologiia zashchity informatsii* [The Theory of Information Security and Methodology of Information Protection]. St. Petersburg, SPbGU ITMO Publ., 2010. 98 p.

12. Makarenko S.I. Informatsionnaia bezopasnost. Uchebnoe posobie dlia studentov vuzov [Information Security. Textbook for Students of Higher Schools]. Stavropol, SF MGGU im. M.A. Sholokhova Publ., 2009. 372 p.

13. Zaichenko Iu.P. *Nechetkie modeli i metody v intellektualnykh sistemakh. Uchebnik dlia vuzov* [Fuzzy Models and Methods in Smart Systems. Textbook for Higher Schools]. Kiev, Izdatelskii Dom Slovo Publ., 2008. 344 p.

14. Uskov A.A. Printsipy postroeniia sistem upravleniia s nechetkoi logikoi [Principles of Construction of Control Systems with Fuzzy Logic]. *Pribory i sistemy: Upravlenie, kontrol, diagnostika* [Instruments and Systems: Monitoring, Control and Diagnostics], 2004, no. 6, pp. 7–13.

15. Terano T., Asai K., Sugeno M. *Prikladnye nechetkie sistemy* [Applied Fuzzy Systems]. Moscow, Mir Publ., 1993. 368 p.

16. Leonenkov A.V. *Nechetkoe modelirovanie v srede MATLAB i fuzzyTECH* [Fuzzy Modelling in MATLAB and FuzzyTECH]. St. Petersburg, BHV-Peterburg Publ., 2005. 716 p.

17. Rutkovskaia D., Pilinskii M., Rutkovskii L. *Neironnye seti, geneticheskie algoritmy i nechetkie sistemy. Per. s polsk. I.D. Rudinskogo* [Neural Networks, Genetic Algorithms and Fuzzy Systems. Translated from Polish by I.D. Rudinskii]. Moscow, Goriachaia liniia – Telekom Publ., 2006. 452 p.

18. Kizbikenov K.O. *Prognozirovanie i vremennye riady* [Forecasting and Time Series]. Barnaul, AltGPU Publ., 2017. 14 p.

19. Chetyrkin E.M. *Statisticheskie metody* [Statistical Methods]. Moscow, Statistika Publ., 1977. 29 p.