УДК 004.891: 004.78

Лускатов И.В., Пилькевич С.В.

Санкт-Петербург, Военно-Космическая академия имени А.Ф. Можайского

МОДЕЛЬ ВЫЯВЛЕНИЯ КИБЕРУГРОЗ НА ИНФОРМАЦИОННЫЕ РЕСУРСЫ СЕТИ ИНТЕРНЕТ

Аннотация. В работе рассматривается построение модели защиты информационных ресурсов на основе новых подходов для активного поиска, унифицированного расследования и реагирования на киберугрозы. В основу процесса выявления киберугроз положены методы поискового прогнозирования, взаимоувязанные с циклическим характером поведенческой активности пользователей сетевых ресурсов.

Ключевые слова: киберугрозы, аномалии, прогнозирование, поведение пользователей, интеллектуальная система.

Lyskatov I.V., Pilkevich S.V.

Saint-Petersburg, Mozhaisky Military Space Academy

DETECTING CYBER THREATS MODEL IN ONLINE MEDIA SOURCES

Abstract. The creation of media source protection model based on new methods of active search, unified investigation and cyber threat response. The process of detecting cyber threats based on methods of search forecasting, interlinked which the cyclic nature of the behavioral activity network resources users.

Keywords: cyber threats, anomaly, forecast, user behavior, intelligence system.

Введение

В современном мире лавинообразный рост разного рода киберугроз делает задачу обеспечения информационной безопасности чрезвычайно актуальной. Существуют многочисленные образчики прикладного программного обеспечения, позволяющего решать те или иные задачи защиты информации универсальными методами. Тем не менее, наиболее органичным представляется наделение соответствующим функционалом самих информационных ресурсов, что обеспечит их защиту с учетом специфики конкретных условий эксплуатации. Решение обозначенной задачи сопряжено с рассмотрением киберугроз, которым должна противостоять система защиты и особенностями защищаемого информационного ресурса.

В рамках настоящей статьи под информационными ресурсами сети Интернет будут пониматься онлайновые информационные площадки (ОИП) – это сайты, сервисы которых имеют выраженную социальную направленность, т.е. характеризуются тесным взаимодействием с пользователями: предоставляют им широкие возможности для опубликования разнородного контента, обмена им, а также хранят персонифицированные сведения о своих пользователях. Совокупность перечисленных свойств переводит ОИП в разряд объектов, предпочтительных для атак нарушителей информационной безопасности, маскирующихся, как правило, под легальных пользователей ОИП и получающих прибыль, от рассылки спама, хищения персональных данных и конфиденциальной информации [1].

1. Подсистема выявления киберугроз онлайновым информационным площадкам

Представляется целесообразным наделить ОИП функциями выявления и устранения киберугроз, реализуемых в рамках одноименной подсистемы ОИП (см. рисунок 1).

Целевое предназначение рассматриваемой подсистемы состоит в выявлении аномалий сетевой активности пользователей информационного ресурса, их оценивании и принятии решения о последующей блокировке

пользователей с подозрительным поведением, либо временном ограничении их прав доступа к сервисам сетевого ресурса.

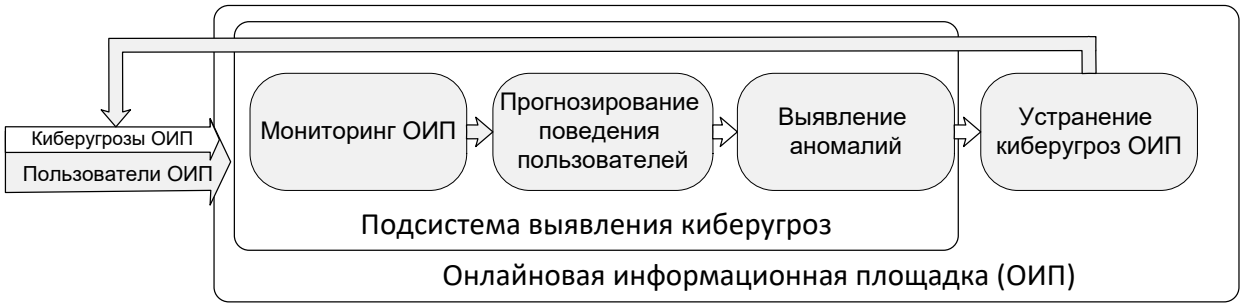

Рисунок 1 - Функциональная схема взаимодействия основных элементов системы защиты онлайновой информационной площадки

В рамках настоящего исследования под аномалией понимается любое отклонение от модели нормального поведения пользователей, например, распространение спама, автоматизированная накрутка лайков и т.п. Основными контролируемыми параметрами ОИП, имеющими потенциал для выявления аномалий путём анализа сетевой активности пользователей, являются: среднее количество пользователей, интенсивность обмена сообщениями, время появления репостов, лайков и т.п.

Анализ современного состояния исследований в области реализации методов выявления (обнаружения) киберугроз информационным ресурсам (см. таблицу 1) показал, что по совокупности таких параметров как достоверность, оперативность и надежность приоритетным является метод на основе прогностической оценки поведения пользователей.

Таблица 1. Качественное сравнение характеристик методов выявления киберугроз

Наименование метода	**Оцениваемый параметр**		
	достоверность	**оперативность**	**надежность**
экспертных оценок	средняя	низкая	средняя
критериальный	средняя	низкая	средняя
статистическо-экспертный	высокая	низкая	высокая
прогностической оценки	высокая	высокая	высокая

2. Интеллектуальная система выявления и реагирования на аномалии онлайновых информационных площадок

Реализация метода, носящего прогностический характер, в рамках подсистемы выявления киберугроз информационного ресурса приводит к необходимости создания интеллектуальной системы, способной к накоплению и обобщению данных мониторинга с последующим автоматическим построением правил реагирования на выявленные киберугрозы, обобщенная схема данной системы приведена на рисунке 2.

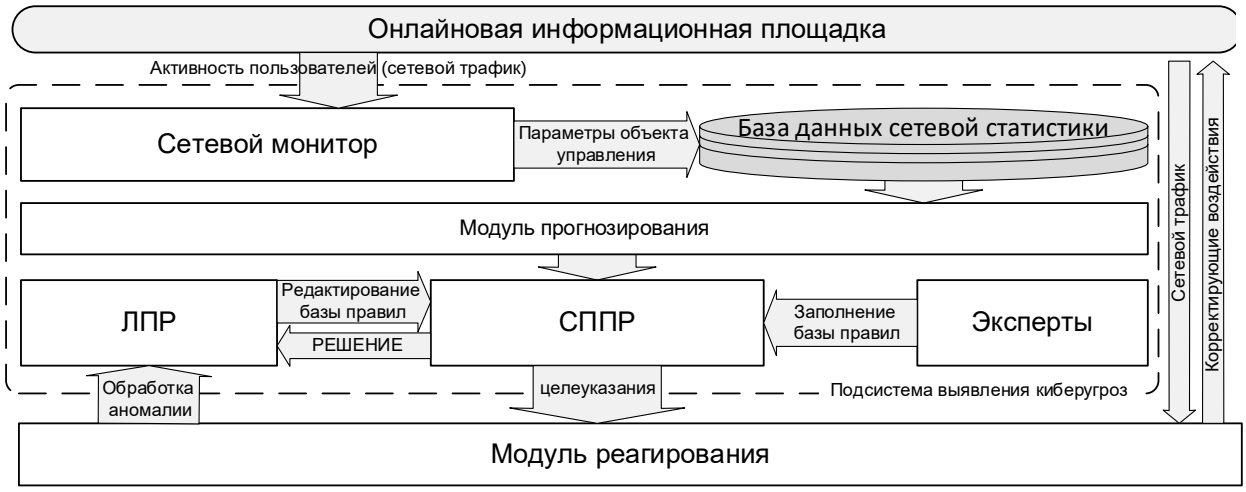

Рисунок 2 - Обобщенная схема интеллектуальной системы выявления и реагирования на аномалии ОИП

В качестве исходных данных выступают сведения о сетевой активности пользователей ОИП, извлекаемые из сетевого трафика, а также информация о современных угрозах ОИП, получаемая из сторонних источников.

Сетевой монитор собирает статистическую информацию о поведении пользователей ОИП. Накопленные сведения сохраняются в базе данных сетевой статистики, на ее основе формируется прогноз сетевой активности пользователей. В системе поддержки принятия решений (СППР) происходит сравнение реальной активности с прогнозируемой и оценивается величина расхождения, в случае превышения значения данного параметра сверх установленного предела, принимается решение о наличии аномалии и производится поиск её источников (причастных пользователей). Заполнение

базы правил СППР производится экспертом, а редактирование - лицом, принимающим решения (ЛПР). По результатам принятого решения формируются своеобразные целеуказания (инструкции), которые передаются в модуль реагирования. Далее осуществляются корректирующие воздействия, состоящие в предупреждении пользователей, блокировке аккаунтов, удалении сообщений и т. п.

Рассмотрим последовательность операций, применяемых к исходным данным и позволяющим, в конечном счете, выявлять и оценивать аномалии.

2.1 Сетевой мониторинг

На первом этапе необходимо провести *сбор статистической информации о поведении пользователей ОИП.* Поскольку цель прогнозирования, на основе имеющихся данных о полезной нагрузке ОИП, получить значения параметров активности на определенный период времени в будущее, из содержания каждого сообщения необходимо выделять информацию о ключевых словах, а также сохранять дату и время его появления на сайте. Для реагирования на выявленную аномалию необходима информация о ее источнике (аккаунте соответствующего пользователя ОИП) и месте публикации. Таким образом, для прогнозирования параметров пользовательской активности из метаданных сообщения извлекаются: автор, ключевые слова (тэги), страница назначения, дата публикации.

2.2 Прогнозирование поведенческой активности пользователей онлайновых информационных площадок

Следующий этап состоит в *построении прогноза усредненных значений интегрального параметра сетевой активности пользователей ОИП*. Прогнозирование реализуется методом циклического анализа и состоит из представленных ниже этапов. За основу взята идея о том, что активность пользователей, характеризуется некоторой периодичностью [2]. Так, например, глобальный анализ Twitter показал, что график интенсивности общения в нем в течение дня сходен с кардиограммой (см. рисунок 3) [3].

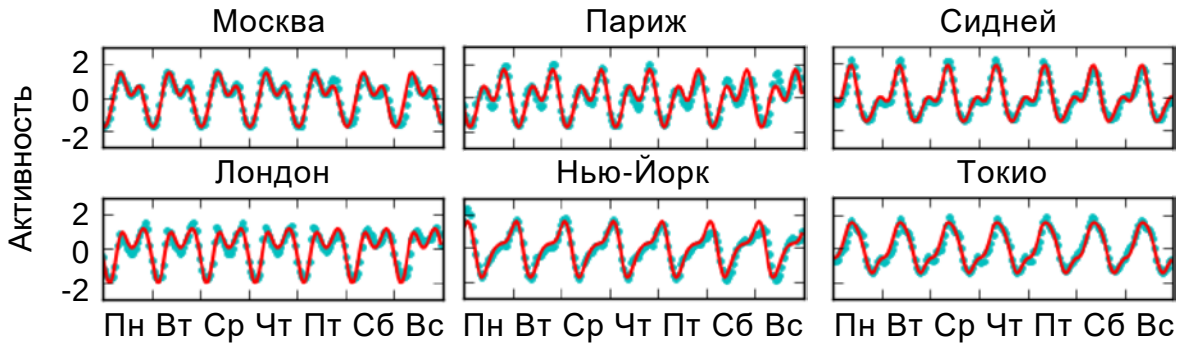

Рисунок 3 - Активность пользователей Twitter
и характерная для неё периодичность

Параметры активности пользователей других социальных сетей также характеризуются способностью образовывать своеобразные «поведенческие паттерны», повторяющиеся во времени. Основываясь на этой особенности, оказалось возможным формировать прогнозные значения активности пользователей ОИП и использовать их для последующего выявления аномалий.

Этап *отбора данных* характерен тем, что собранную статистику необходимо представить в виде ряда данных. Пусть имеется статистика по активности пользователей ОИП, собранная за период времени T. Чтобы получить ряд данных, разделим период времени T на Q равных интервалов Δt: $Q = T/\Delta t$. Значения T и Δt подбираются таким образом, чтобы $Q \in \mathbb{N}$.

Введем интегральный параметр сетевой активности пользователей ОИП – своеобразный «объём» j-ого сообщения $\left(V_j\right)$, вычисляемый в соответствии со следующим выражением:

$$V_j = k_l N_l + k_r N_r + k_u N_u + k_c N_c,$$

где N_l – число лайков, N_r – репостов, N_c – комментариев, N_u – ссылок на СМИ, k_l, k_r, k_u, k_c – коэффициенты нормирования.

Далее для каждого интервала Δt складываем «объемы» R сообщений, попавших в данный интервал времени: $X_q\left(\Delta t\right) = \sum_{j=1}^{R} V_j$,

где R – количество сообщений, попавших в интервал Δt,

q – номер интервала, $q \in [1,...,Q]$,

X_q – ряд данных, описывающий изменения «объема» сообщений во времени, с частотой дискретизации Δt.

На этапе *сглаживания результатов мониторинга сетевой активности пользователей ОИП* с помощью метода центрированной скользящей средней выполняется сглаживание случайных колебаний в анализируемой выборке. Количество точек для сглаживания данных возьмем равным L (на каждую выполняемую итерацию необходимо выбирать нечетное количество точек, $L \geq 3$). При вычислении скользящей средней по L точкам, из первоначального ряда данных будет отброшено L-1 точек: $(L-1)/2$ – в начале и столько же в конце ряда. Пример применения описанного подхода представлен на рисунке 4.

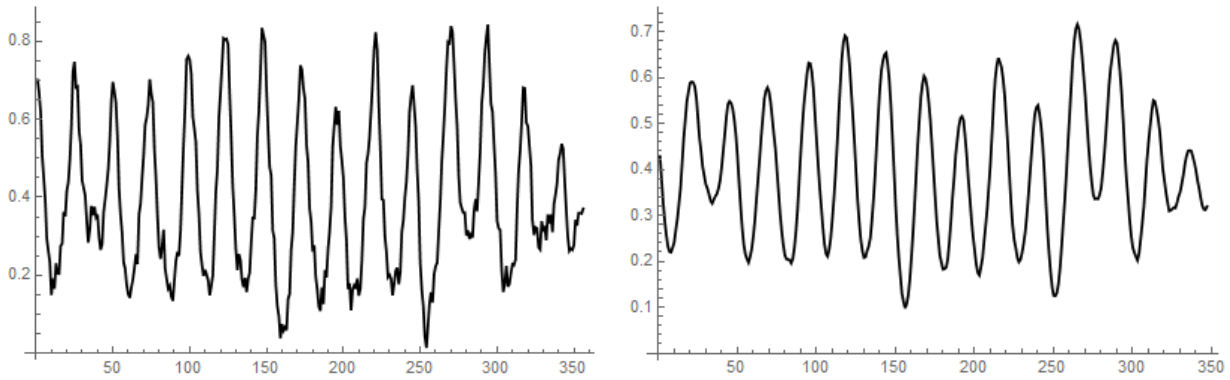

Рисунок 4 – Исходные данные о пользовательской активности, содержащие случайные колебания и результат их сглаживания

Периодический характер анализируемых данных проявляется с определенной цикличностью. Данное обстоятельство предопределяет наличие следующего этапа обработки - *поиска возможных циклов*.

Для определения частотных составляющих рассматриваемого ряда, используем метод спектрального анализа, математической основой которого является преобразование Фурье.

С помощью прямого дискретного преобразования Фурье найдем комплексные амплитуды ряда данных X_k, полученного в результате сглаживания X_q:

$$Y_n = \sum_{k=1}^{N} X_k e^{\left(-\frac{2\pi i}{N} nk\right)},$$

где N – количество значений сигнала (длина исследуемого фрагмента данных), измеренных за период, а также количество компонент разложения;

$k = 1,\ldots, N$ – номера дискретных временных точек, в которых измерялись значения X_k;

i – мнимая единица;

$n = 1,\ldots, N$ – индекс частоты.

На основе значений комплексных амплитуд Y_n вычисляется спектр мощности $\mathrm{R}_n = |Y_n|^2 = \mathrm{Re}^2(Y_n) + \mathrm{Im}^2(Y_n)$,

На рисунке 5 видно, что локальные максимумы анализируемой функции могут быть найдены как эмпирически, так и численными методами. Полученные таким образом значения локальных максимумов свидетельствуют о возможном наличии циклов. Значением частоты цикла будет являться индекс *n*, при котором наблюдается высокое значение спектра мощности R_n.

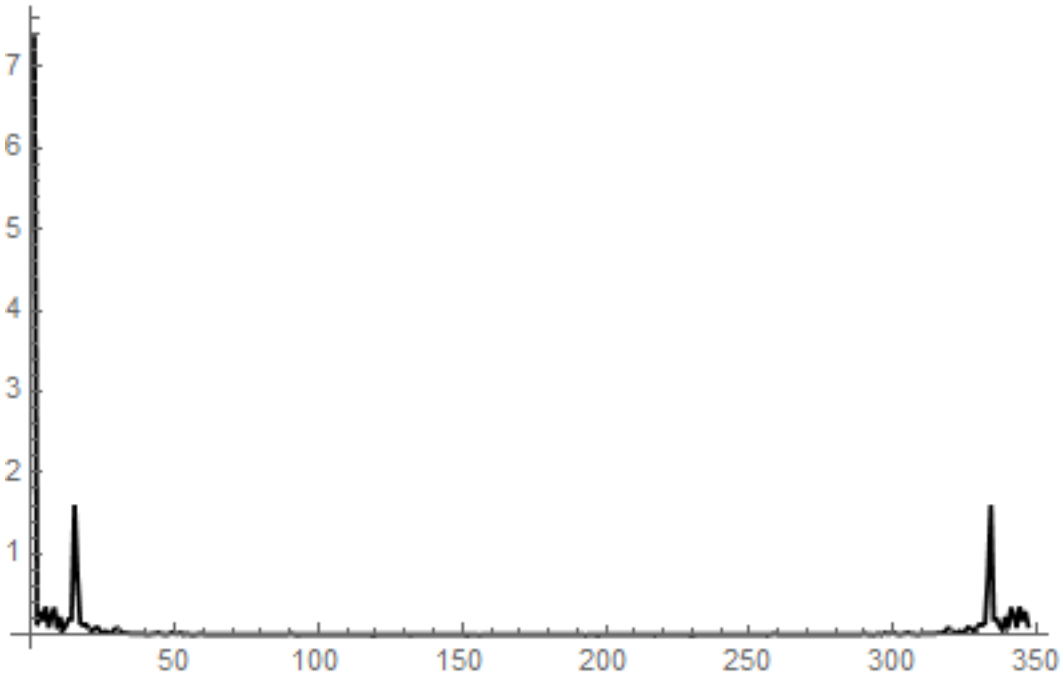

Рисунок 5 – Результат применения дискретного преобразования Фурье (Спектр мощности R_n)

Определив возможные циклы и их частоты, рассчитаем вещественную амплитуду A и фазу φ. Пусть найдено b возможных циклов, частоты которых составляют множество S. Тогда амплитуды и фазы найденных циклов могут быть вычислены по формулам:

$$A_h = \frac{|S_h|}{N} = \frac{1}{N}\sqrt{\mathrm{Re}^2(S_h) + \mathrm{Im}^2(S_h)},\ \varphi_h = Arg(S_h) = arctg\left(\frac{\mathrm{Im}(S_h)}{\mathrm{Re}(S_h)}\right),$$

где $h = 1,...,b,\ Arg(S_h)$ – функция комплексного числа, угол комплексного числа (в радианах), соответствующий S_h.

В таком случае функция, описывающая цикл, выглядит следующим образом: $f_h(t) = A_h \cos(S_h t + \varphi_h)$.

Удаление трендовых компонентов в анализируемых данных. Качество проверки циклов на статистическую надежность сильно зависит от существования направленности в данных. Чтобы удалить тренд в данных необходимо для каждой найденной частоты рассчитать скользящую среднюю $\bar{X}'_k$ для ряда данных X_k с количеством точек сглаживания $L = \begin{cases} S_h, & \text{если } L = 2p+1 \\ S_h + 1, & \text{если } L = 2p \end{cases}$, $p \in \mathbb{N}$, в таком случае получим $\bar{X}'_k = \frac{1}{L}\sum_{j=k}^{k+L-1} X_j$.

Далее вычитаем из исходного ряда данных X_k полученную скользящую среднюю $\bar{X}'_k : X''_k = X_k - \bar{X}'_k$. Сгладив таким образом краткосрочные колебания исходных данных можно приступать к проверке найденных возможных циклов на статистическую значимость.

Проверка циклов с точки зрения статистической значимости. Для оценки циклов используются критерий Фишера и χ^2. Первый измеряет надежность амплитуды цикла (его формы), второй - надежность фазы цикла (его времени).

Комбинирование и проецирование циклов в будущее. Допустим, что проверочные тесты предыдущего этапа прошли D циклов. Подтвердившиеся циклы позволяют объединить соответствующие функции $f(t)$ (полученные

на этапе поиска возможных циклов) в общую кривую, описывающую периодичность, выявленную в ряде данных: $\overline{V}(t) = \sum_{j=1}^{D} f_j(t)$ (см. рисунок 6).

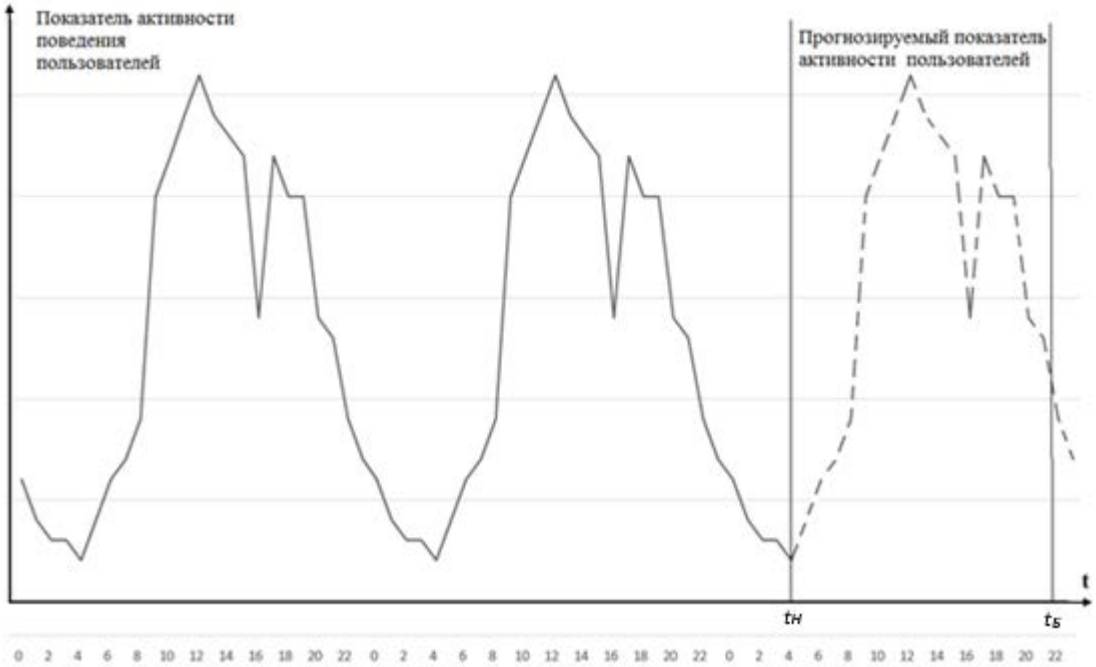

Рисунок 6 - Прогнозирование активности пользователей.

$[0, t_H]$ – период мониторинга, $(t_H, t_Б]$ – прогнозируемый период

Полученная функция $\overline{V}(t)$ может быть экстраполирована и позволяет получить прогнозируемое значение активности пользователей ОИП на предстоящий период времени $t' \in (t_H, t_Б]$.

2.3 Выявление аномалий

На основе данных о прогнозируемой и актуальной активности пользователей ОИП, а также совокупности правил принятия решений, возможно реализовать процедуры выявления и оценивания потенциальной опасности аномалий по отношению к безопасности ОИП.

Поиск аномалий происходит на основе сравнения введенного параметра активности пользователей (значения которого рассчитываются в реальном режиме времени на основе результатов мониторинга сетевой пользовательской активности согласно изложенному выше методу) с прогнозируемым значением. Для этого в единицу времени *t* сравниваются два

значения $V_{реал}$ – текущая величина параметра сетевой активности и $V_{прогноз}$ – прогнозируемое значение активности. Аномальным считается отклонение параметра, превышающее или равное заданной пороговой величине α: $\left|V_{прогноз} - V_{реал}\right| \geq \alpha$. Если аномалия была обнаружена, осуществляется поиск её источников. Определение источников (персонального и/или коллективного автора сообщения, сетевого информационного ресурса) производится на основе информации, извлекаемой из метаданных текущего трафика ОИП.

Далее оценивается величина аномалии и принимается решение по ее обработке. Рассматриваются два основных способа обработки аномалий: предупреждение – фильтрация аномальной активности и блокировка – отключение вредоносных аккаунтов.

Оценивание величины аномалии происходит на основе продукционной базы правил. При этом разнообразие ОИП, интенсивности обмена сообщениями различной тематической направленности, число лайков и репостов, зависящее не только от персоны пользователя ОИП, но и от наличия/отсутствия соответствующего информационного повода и состояния информационного пространства в целом чрезвычайно затрудняют деятельность экспертов, оперирующих, как правило, качественными оценками и испытывающими затруднения при указании точных диапазонов количественных параметров, описывающих аномалию. Изложенные обстоятельства обусловили необходимость использования теории нечеткости Л.Заде [4].

Рассмотрим структуру базы правил (БП) более подробно. Для оценки величины аномалии используются лингвистические переменные с соответствующими терм-множествами:

- $V \in$ {низкая, ниже среднего, средняя, выше среднего, высокая} – величина отклонения;
- $M \in$ {низкая, ниже среднего, средняя, выше среднего, постоянная} – частота появления аномалии;
- $I \in$ {незначительное, ниже среднего, среднее, выше среднего,

большое} – количество источников аномалий;

– $W \in$ {незначительный, ниже среднего, средний, выше среднего, высокий} – объем трафика от одного источника.

Выходной параметр $E \in$ {незначительная, ниже среднего, средняя, выше среднего, высокая} – величина аномалии.

Хранение правил организовано с использованием методологии экспертных систем. Приведем ряд примеров используемых правил:

– ЕСЛИ $V \in$ {ОТ низкая ДО ниже среднего} и $M \in$ {ОТ низкая ДО средняя} и $I \in$ {ОТ незначительное ДО среднее} и $W \in$ {ОТ незначительный ДО средний} ТО E = «незначительная»;

– ЕСЛИ $V \in$ {ОТ выше среднего ДО высокая} и $M \in$ {ОТ средняя ДО средняя } и $I \in$ {ОТ выше среднего ДО большое} и $W \in$ {ОТ выше среднего ДО выше среднего} ТО E = «высокая».

Введение в перечень используемого научно-методического аппарата нечетких множеств требует указания границ значений для соответствующих функций принадлежности (μ_V, μ_M, μ_I, μ_W и μ_E). Ввиду того обстоятельства, что для каждого ОИП рассматриваемые значения будут индивидуальными, укажем их общие характерные черты.

В работе использованы колоколообразные функции принадлежности, т.е. функции вида $\mu(x,a,b,c) = \dfrac{1}{1+\left(\dfrac{x-c}{a}\right)^{2b}}$. Получение наиболее характерных значений $(a, b \text{ и } c)$ данных функций принадлежности осуществлялось прямыми методами (как одиночными, так и групповыми).

2.4 Реагирование на выявленные аномалии онлайновых информационных площадок

Заключительные операции, реализуемые СППР состоят в формировании наборов инструкций, передаваемых модулю реагирования для нейтрализации обнаруженных кибератак.

Правила, задающие параметры фильтрации скомпрометированного аккаунта (источника обнаруженной аномалии), выбираются в зависимости от истории его существования. Если пользователь не был ранее скомпрометирован и характеризуется длительным временем своего существования на ресурсе, то он временно блокируется до подтверждения регистрационных данных, в противном случае – удаляется.

Общая модель формирования правил управления корректирующими воздействиями может быть представлена следующим образом:

$$G = Func(e, Z, Flt, U),$$

где *Func* - функция формирования правил управления;

e – прогнозные значения активности пользователей;

Z – аккаунты источников аномалии;

Flt – параметр, характеризующий ответную реакцию (например, время фильтрации), определяется в зависимости от величины *e*;

U – список исключений.

3. Модель выявления и оценки аномалий

Опираясь на рассмотренную схему поиска и оценки аномалий, построим модель выявления и оценки аномалий на основе прогнозирования активности пользователей:

$$e = \begin{cases} Prod(V, M, I, W, E), & если \left| V_{прогноз} - V_{реал} \right| \geq \alpha \\ 0, & иначе \end{cases},$$

где *Prod* – процедура использования продукционных правил;

α – пороговое значение, на основе сравнения с которым принимается решение о наличии или отсутствии аномалии.

Необходимость принятия мер для устранения аномалии определяется на основании информации об аномалии, а также за счет настроек ЛПР, на основе которых формируются исключения $\{Z\}$. Далее происходит непосредственное управление корректирующими воздействиями и подготовка отчета об аномалии.

Рассматриваемая интеллектуальная система может быть отнесена к классу распознающих, отсюда проблема выбора значения α порога решающего правила представляет собой оптимизационную задачу поиска соотношения между ошибками первого и второго рода. Возможные комбинации результатов классификации и исходных данных приведены в таблице контингентности (см. таблицу 2).

Таблица 2. Соотнесение результатов распознавания и истинного положения.

		Результат классификации аномалии	
		Наличие атаки	**Штатный режим**
Истинное положение	**Наличие атаки**	Верно (*A*)	Ошибка 1 рода (пропуск атаки, *C*)
	Штатный режим	Ошибка 2 рода (ложное срабатывание, *B*)	Верно (*D*)

При этом различают [5] ряд показателей, среди которых выделим следующие:

– *TPR* – чувствительность алгоритма классификации (доля ложно классифицированных ситуаций) $TPR = A/(A+C)$;

– *PPV* – точность алгоритма классификации (доля верно классифицированных ситуаций) $PPV = A/(A+B)$.

В качестве обобщающей характеристики перечисленных показателей выбрана сбалансированная *F*-мера: $F = 2\dfrac{TPR \cdot PPV}{TPR + PPV}$. Анализ графика *F*-меры позволяет принимать обоснованное решение по выбору значения α.

Заключение

Таким образом, представленная модель выявления киберугроз является центральным элементом системы защиты информационных ресурсов сети Интернет и базируется на интеллектуальной системе выявления и реагирования на аномалии. Предложенная модель, применительно к защите онлайновых информационных площадок, позволяет формализовать механизм прогнозирования поведенческой активности пользователей, выявления и

реагирования на аномалии онлайновых информационных площадок. Приведенные формализмы, реализованные в функциональных модулях подсистемы защиты информационной системы, позволяют повысить эффективность защиты от спам-атак, осуществить блокировку скомпрометированных и удаление «зловредных» аккаунтов для обеспечения комфортного пребывания пользователей на веб-ресурсе.

Библиографический список:

1. **Кириченко, Л.** Обнаружение киберугроз с помощью анализа социальных сетей / Л. Кириченко, Т. Радивилова, А. Барановский // International Journal "Information Technologies & Knowledge". – 2017. – Vol. 11. – № 1. – P. 23–48.

2. **Yang, C.C.** Analysis of terrorist social networks with fractal views / C.C. Yang, M. Sageman // Journal of Information Science. – 2009. – Vol. 35. – № 3. – P. 299–320.

3. **Morales, A.J.** Global Patterns of Synchronization in Human Communications / A.J. Morales, V. Vavilala, R.M. Benito, Y. Bar-Yam // Journal of the Royal Society Interface. – 2017. – Vol. 14(128) [Электронный ресурс]. URL: figshare.com/collections/Supplementary_material_from_Global_patterns_of_synchronization_in_human_communications_/3694468 (дата обращения: 29.01.2018).

4. **Zadeh, L.A.** Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic / L.A. Zadeh // Fuzzy Sets Syst. – 1997. – № 4. – P. 103–111.

5. One ROC Curve and Cutoff Analysis // NCSS Statistical Software [Электронный ресурс]. URL: www.NCSS.com/wp-content/themes/ncss/pdf/Procedures/ One_ROC_Curve_and_Cutoff_Analysis.pdf (дата обращения: 07.10.2017).