УДК 519.688

Ю.М. Краковский, Б.В. Курчинский, А.Н. Лузгин

Интервальное прогнозирование интенсивности кибератак на объекты критической информационной инфраструктуры

В современном мире вопросы кибербезопасности занимают одну из ключевых и чрезвычайно значимых стратегических ниш в системе государственного планирования и управления. Кибератаки происходят ежедневно, а их число растет экспоненциально. В таких условиях особую актуальность приобретают вопросы кибербезопасности в отношении объектов критической информационной инфраструктуры государства. Актуальность проведения исследований в направлении создания и усовершенствования технологий защиты от кибератак на соответствующие объекты не вызывает сомнений. В данной работе приводятся результаты применения интервального прогнозирования интенсивности кибератак посредством интеллектуального алгоритма, основанного на вероятностной нейронной сети с динамическим обновлением параметра сглаживания. В качестве исходных данных в работе используются данные о почасовой интенсивности кибератак, полученных с помощью хонипотов (honeypots) с марта по сентябрь 2013 г. Полученные результаты свидетельствуют о высокой точности прогнозирования по предложенному алгоритму. По результатам исследования авторы дают необходимые практические рекомендации о применении результатов интервального прогнозирования при противодействии кибератакам на объекты критической информационной инфраструктуры.

Ключевые слова: интервальное прогнозирование, кибератаки, вероятностная нейронная сеть, критическая информационная инфраструктура.

doi: 10.21293/1818-0442-2018-21-1-71-79

Появление новых информационных технологий приводит к появлению новых уязвимостей, которые активно используются не только для добывания конфиденциальной информации и деструктивного воздействия на неё, но и для нарушения работоспособности различных технических средств и систем [1]. Подобные действия называются кибератаками (или компьютерными атаками). Фактически, кибератаки (такие, как отключение камер, отключение подсветки высотных зданий, нарушение работоспособности беспилотных аппаратов) стали подменять собой атаки физические [2]. Понятия «информационная война», «кибервойна» или «кибертерроризм» стали восприниматься не абстрактно, а как объективная реальность [3].

Кибератаки происходят ежедневно, а их число растет экспоненциально. В таких условиях особую актуальность приобретают вопросы кибербезопасности в отношении объектов критической информационной инфраструктуры (ОКИИ) государства. В результате нарушения работоспособности ОКИИ может сложиться чрезвычайная ситуация, связанная с гибелью людей, экологическими катастрофами, нанесением крупного материально-финансового, экономического ущерба или крупномасштабными нарушениями жизнедеятельности городов и населенных пунктов. Например, по данным АО «Лаборатория Касперского» [4], в 2017 г. наиболее значительной угрозой для ОКИИ стали кибератаки с помощью программ «шифровальщиков-вымогателей», которым подверглись ОКИИ в 63 странах мира.

В принятой в 2016 г. Доктрине информационной безопасности Российской Федерации [5] отмечается, что состояние информационной безопасности в области государственной и общественной безопасности характеризуется постоянным повышением сложности, увеличением масштабов и ростом кибератак на ОКИИ. Принятие в 2017 г. Федерального закона «О безопасности критической информационной инфраструктуры Российской Федерации» [6], который устанавливает обязательное требование о внедрении государственной системы обнаружения, предупреждения и ликвидации последствий кибератак (СОПКА) на ОКИИ, еще раз подтверждает значимость и актуальность вопросов кибербезопасности ОКИИ для Российской Федерации.

Актуальность проведения исследований в направлении создания и усовершенствования технологий защиты от кибератак на ОКИИ не вызывает сомнений [7]. В частности, в Доктрине информационной безопасности Российской Федерации [5], отмечается недостаточная эффективность научных исследований, направленных на создание перспективных технологий и методов защиты от кибератак. Таким образом, научные исследования по разработке новых методов защиты от кибератак на ОКИИ являются очевидной необходимостью.

Прогнозирование интенсивности кибератак в концепции раннего распознавания и предупреждения

Недостаток большинства современных систем обеспечения кибербезопасности ОКИИ заключается в том, что при идентификации кибератак используются заранее известные сигнатуры или прототипы некоторых процессов или событий [8]. Например, так осуществляется работа антивирусных систем, межсетевых экранов, систем обнаружения и предотвращения вторжений. В работе [8] отмечается, что подобные системы эффективны лишь в отношении начинающих злоумышленников, которые используют типовые приёмы и инструменты для организации кибератак. Против опытных злоумышленников эти

системы, как правило, оказываются неэффективными. Здесь одним из перспективных направлений исследований для решения данной проблемы является направление по прогнозированию интенсивности кибератак на ОКИИ посредством машинного обучения. Например, такой подход удачно интегрируется в изложенную в [9, 10] концепцию раннего распознавания кибератак и предупреждения о них.

Отметим, что под интенсивностью кибератак понимается суммарное число этих атак в единицу времени. Тогда в случае получения прогноза о том, что интенсивность кибератак на ОКИИ превысит некоторую заранее заданную величину, могут приниматься дополнительные (автоматические) меры защиты (например, отключение сети Интернет, выделение дополнительных вычислительных ресурсов и проведение глубокого интеллектуального эвристического анализа трафика и т.п.).

Следует отметить, что в Федеральном законе «О безопасности критической информационной инфраструктуры Российской Федерации» [6], равно как и в «Концепции государственной системы обнаружения, предупреждения и ликвидации последствий компьютерных атак на информационные ресурсы Российской Федерации» [11] подчеркивается необходимость в осуществлении прогнозирования в сфере кибербезопасности.

Таким образом, в направлении, связанном с противодействием кибератакам, кроме оценки соответствующих рисков и использования традиционных систем кибербезопасности, необходимо уделять внимание прогнозированию их интенсивности [12].

В последние несколько лет наблюдается возрастающий интерес исследователей к вероятностному прогнозированию кибератак [13–17]. Это можно объяснить тем, что вероятностные прогнозы позволяют получать не только прогнозы непосредственно будущих событий, но и оценки их вероятностей. Разновидностью вероятностного прогнозирования является интервальное прогнозирование [18]. Суть этого прогнозирования заключается в прогнозировании интервала (из двух заранее заданных интервалов), в котором будет находиться будущее значение показателя на основе оценок вероятностей этих событий. Разделительная граница интервалов задается расчетным способом исходя из статистических характеристик этого показателя. В данной работе выбраны несколько показателей интенсивности кибератак, которые следует рассмотреть подробно.

Показатели интенсивности кибератак и их формализация

В качестве исходных показателей в работе рассматриваются данные о почасовой интенсивности кибератак, полученных с помощью хонипотов (honeypots) [19] с марта по сентябрь 2013 г. с условными названиями хостов «groucho-sydney» (показатель ISD) и «groucho-sa» (показатель ISA) [20]. Учитывая специфику этих данных, нами была проведена их незначительная предобработка:

1. Учитывая, что в исходных данных некоторые значения существенно превышают остальные, мы применили натуральное логарифмирование. Данная процедура полностью обратима, не влияет на интерпретацию полученных результатов, упрощает работу с графиками и в ряде случаев позволяет повысить качество машинного «обучения» и, как следствие, точность интервального прогнозирования.

2. Автором статьи [19] для анализа «фронтов» начала и окончания кибератак было рекомендовано применить к исходным данным преобразование методом простого скользящего среднего (ПСС), т.к. данный метод позволяет уменьшить влияние экстремальных значений и «высокочастотных» шумов. Мы применили к данным, полученным после логарифмирования, ПСС с шириной окна равной двум часам.

Мы намеренно взяли минимальную ширину окна ПСС, чтобы получить некоторый положительный эффект от данной процедуры, но в то же время не «потерять» важные «детали» в закономерностях происходящих процессов.

После этих преобразований показатель ISD был обозначен как SD, а ISA – как SA. Каждый этот показатель был формализован в виде временного ряда:

$$\mathbf{q}=\{q_t : t\in\mathbf{t}\}. \qquad (1)$$

Здесь q_t – значения прогнозируемого показателя в дискретные моменты времени t, где t принимает значения из множества $\mathbf{t}=\{1,...,n\}$, а n – количество (объем) значений показателя. В нашем случае этот объём был сокращён до 4000 значений $(n=4000)$.

Графики показателей SD и SA за последние 168 часов (полная неделя) показаны на рис. 1.

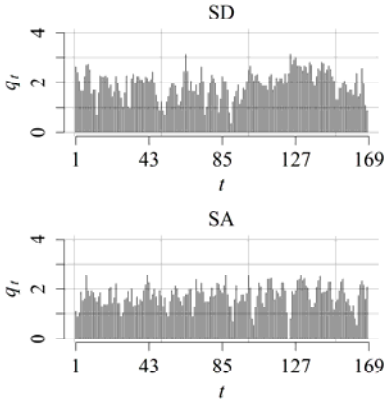

Рис. 1. Графики показателей SD и SA за последние 168 ч

Показатель (1) можно рассматривать как случайную величину с некоторым неизвестным законом распределения вероятностей. Известно, что любое распределение вероятностей можно описать параметром положения, характеризующим центр группирования значений этой случайной величины, и параметром масштаба, характеризующим степень рассеяния значений этой случайной величины относительно центра группирования. В работе [21] был предложен метод классификации какого-либо показателя (1) по четырем классам на основе стационар-

ности (нестационарности) с течением времени по параметру положения и параметру масштаба.

Сделанные расчеты показали, что оба показателя относятся к показателям первого класса (нестационарные по параметру положения и параметру масштаба). В статистическом смысле такие показатели, как правило, являются самыми сложными для прогнозирования в сравнении с показателями других классов.

Формализация интервального прогнозирования показателей интенсивности кибератак

Введем интервал $(q_{\min};q_{\max})$ возможных значений показателя **q** (1) и внутреннюю точку $\hat{q}$ $(q_{\min}<\hat{q}<q_{\max})$. Это позволяет создать два интервала:

$$I^{-}=(q_{\min};\hat{q}],\ I^{+}=(\hat{q};q_{\max}). \qquad (2)$$

Для интервалов (2) значение внутренней точки $\hat{q}$ предлагается определять так:

$$\begin{aligned}\hat{q}&=med(\mathbf{q})+\beta\times med\left(\left|\mathbf{q}-med(\mathbf{q})\right|\right)=\\&=med(\mathbf{q})+\beta\times MAD(\mathbf{q}),\end{aligned} \qquad (3)$$

где $\beta\in[-1;1]$ – коэффициент, который задаётся заранее; $med(\cdot)$ – медиана по множеству значений; $MAD(\cdot)$ – медианное абсолютное отклонение.

В момент времени $t=n$ необходимо определить, в каком интервале (2) будет находиться будущее (неизвестное) значение q_{t+p} на основе оценок вероятностей ρ^{+}_{t+p} и ρ^{-}_{t+p}, где $p=1,...,r$ – время упреждения; ρ^{+}_{t+p} – вероятность того, что будущее значение $q_{t+p}\in I^{+}$, ρ^{-}_{t+p} – вероятность того, что будущее значение $q_{t+p}\in I^{-}$; $\rho^{+}_{t+p}+\rho^{-}_{t+p}=1$.

Пусть $\tilde{\rho}^{+}_{t+p}$ и $\tilde{\rho}^{-}_{t+p}$ – оценки соответствующих неизвестных вероятностей ρ^{+}_{t+p} и ρ^{-}_{t+p}. Интервальное прогнозирование проводится по правилу: будущее значение $q_{t+p}\in I^{+}$, если $\tilde{\rho}^{+}_{t+p}>\tilde{\rho}^{-}_{t+p}$; будущее значение $q_{t+p}\in I^{-}$, если $\tilde{\rho}^{-}_{t+p}\le\tilde{\rho}^{+}_{t+p}$.

Следует рассмотреть выражение (3) подробнее и сделать некоторые пояснения.

Как было отмечено ранее, выбранные показатели рассматриваются как случайные величины с некоторым неизвестным законом распределения вероятностей, характеристиками которого являются параметр положения и параметр масштаба. Так как даже после соответствующих преобразований в данных остаются выбросы, для оценки значения параметра положения целесообразно использовать медиану (это робастный аналог среднего значения), а для оценки значения параметра масштаба использовать медианное абсолютное отклонение (это робастный аналог стандартного отклонения). С учетом этого, в выражении (3) первое слагаемое есть оценка значения параметра положения, а второе слагаемое есть оценка значения параметра масштаба, которая умножается на коэффициент $\beta\in[-1;1]$. Таким образом, мы можем менять значение $\hat{q}$ (2) в диапазоне от $med(\mathbf{q})-MAD(\mathbf{q})$ до $med(\mathbf{q})+MAD(\mathbf{q})$.

На рис. 2 для показателя SD представлены доли кибератак (*sa*), которые попадают в интервал I^{+}, в зависимости от значения β (для показателя SA график выглядит практически идентично).

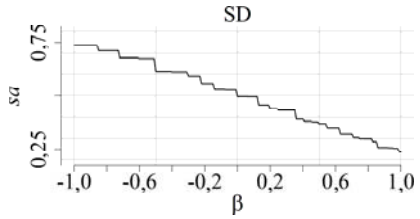

Рис. 2. График доли кибератак, которые попадают в интервал I^{+} (*sa*), в зависимости от значения β

Из графика видно, что с увеличением значения β доля кибератак, которые находятся в интервале I^{+}, уменьшается (для показателя SD значение *sa* находится в интервале от $0{,}24$ до $0{,}74$, а для показателя SA – в интервале от $0{,}22$ до $0{,}73$). Следует подчеркнуть, что округленно при $\beta=-1$ в интервал I^{+} попадает около 70% всех имеющихся значений каждого показателя**,** а при $\beta=1$ в интервал I^{+} попадает около 20% значений. Мы полагаем, что вариация значений $\hat{q}$ в таком диапазоне достаточна для проведения дальнейших исследований в выбранной предметной области.

Значение $\hat{q}$ (3) мы будем называть предустановленным уровнем интенсивности кибератак. Чем больше значение $\hat{q}$ (3), тем интенсивнее должна быть кибератака, чтобы попасть в интервал I^{+} (2). Мы можем задать такое значение β, для которого интервальные прогнозы того, что $q_{t+p}\in I^{+}$, будут свидетельствовать о необходимости принятия дополнительных мер защиты. Тогда прогнозы того, что $q_{t+p}\in I^{-}$, будут рассматриваться как штатная ситуация и оставаться без внимания.

Чем меньше значение β, тем чаще будут происходить прогнозы того, что $q_{t+p}\in I^{+}$, и наоборот. Безусловно, в каждом конкретном случае и в каждой конкретной организации значения β должны выбираться эмпирически или экспертным путем и, следовательно, эти значения могут отличаться.

Формирование обучающего множества

Для осуществления интервального прогнозирования мы выбрали вероятностную нейронную сеть [22] с динамическим обновлением параметра сгла-

живания [23] (ВНС). Модель ВНС была выбрана по нескольким причинам:

1. Преимущества ВНС (применительно к прогнозированию интенсивности кибератак) преобладают над недостатками [24]. Например, ВНС: а) при обучении и прогнозировании устойчива к аномальным выбросам; б) модель относится к моделям «ленивого» обучения и обучается максимально быстро в сравнении с моделями других классов; в) модель устойчива к «дисбалансу» классов обучающего множества; г) результаты работы ВНС легко поддаются интерпретации, так как работа ВНС основана на выявлении «схожих» объектов; д) не требует априорных знаний о статистических характеристиках прогнозируемого показателя. К недостаткам можно отнести: а) «неотделимость» процесса прогнозирования от обучающих данных (в отличие, например, от параметрических моделей, где «обучение» заключается в оценке параметров моделей); б) обучающая выборка должна быть репрезентативной.

2. Модель ВНС демонстрирует лучшую точность интервального прогнозирования различных по своим статистическим характеристикам показателей в сравнении с моделями других классов (например, кластерной или регрессионной) [22, 23, 25, 26].

Учитывая изложенное, а также факт того, что в России и за рубежом исследований интервального прогнозирования кибератак на основе каких-либо известных методов авторам данной работы найти не удалось, в качестве отправной точки была выбрана ВНС.

Алгоритмы построения модели вероятностной нейронной сети и прогнозирования на её основе подробно описаны в [23, 25]. Тем не менее необходимо рассмотреть некоторые особенности формирования обучающего множества этой сети для осуществления интервального прогнозирования.

Пусть при $t=n$ имеется последовательность значений $q_{t-f+1},...,q_t$ количеством f. Сформируем матрицу-строку размером $1\times f$:

$$\mathbf{z}=\left(q_{t-f+1}\Big/\sqrt{\sum_{\zeta=1}^{f} q_{t-f+\zeta}^2},...,q_t\Big/\sqrt{\sum_{\zeta=1}^{f} q_{t-f+\zeta}^2}\right).$$

Делитель $\sqrt{\sum_{\zeta=1}^{f} q_{t-f+\zeta}^2}$ необходим для того, чтобы $\|\mathbf{z}\|_2=1$ (норма l_2) [22].

Пусть имеется зависимая переменная-признак (называемая также откликом) y_{t+p}, истинное значение которой неизвестно, но оно может принимать только два возможных значения: $y_{t+p}=1$, если $q_{t+p}\in I^+$, и $y_{t+p}=-1$, если $q_{t+p}\in I^-$.

При осуществлении интервального прогнозирования, используя $\mathbf{z}$, требуется выполнить прогноз отклика y_{t+p} на основе оценок вероятностей того, что $q_{t+p}\in I^+$ или $q_{t+p}\in I^-$. Напомним, что если $\tilde{\rho}_{t+p}^{+}>\tilde{\rho}_{t+p}^{-}$, то $y_{t+p}=1$, иначе $y_{t+p}=-1$.

Используя значения показателя (1) для $t=1,...,m$, где $m=n-f-p$ (это значение выбрано так, чтобы можно было рассчитать значения откликов по предыстории показателя), построим обучающее множество (training sample) так:

$$\mathbf{x}=\begin{pmatrix} q_1\Big/\sqrt{\sum_{\zeta=1}^{1+f-1} q_\zeta^2} & \cdots & q_{1+f-1}\Big/\sqrt{\sum_{\zeta=1}^{1+f-1} q_\zeta^2} \\ q_2\Big/\sqrt{\sum_{\zeta=1}^{2+f-1} q_\zeta^2} & \cdots & q_{2+f-1}\Big/\sqrt{\sum_{\zeta=1}^{2+f-1} q_\zeta^2} \\ & \cdots & \\ q_m\Big/\sqrt{\sum_{\zeta=m}^{m+f-1} q_\zeta^2} & \cdots & q_{m+f-1}\Big/\sqrt{\sum_{\zeta=m}^{m+f-1} q_\zeta^2} \end{pmatrix}, \quad (4)$$

$$\mathbf{y}=(y_1\ y_2\ ...\ y_m).$$

Здесь $\mathbf{x}$ – матрица предикторов размером $m\times f$, где индекс каждого предиктора указывает на позицию соответствующего элемента в $\mathbf{q}$ (1); $\mathbf{y}$ – матрица-строка откликов размером $1\times m$ (эти отклики рассчитываются по предыстории показателя); m – число «обучающих» примеров или объектов.

Каждой строке матрицы $\mathbf{x}$ соответствует отклик матрицы-строки $\mathbf{y}$ (4). При этом норма l_2 каждой строки матрицы равна 1 [22].

Используя обучающее множество (4), можно построить и обучить вероятностную нейронную сеть, а также осуществлять интервальное прогнозирование, используя $\mathbf{z}$.

Таким образом, алгоритм интервального прогнозирования интенсивности кибератак на основе вероятностной нейронной сети имеет три параметра: f, β и p.

Результаты интервального прогнозирования и их обсуждение

Для анализа результатов интервального прогнозирования интенсивности кибератак на ОКИИ нами использовалось несколько величин. Рассмотрим их подробнее и аргументируем выбор каждой из них.

Прежде всего нас интересует точность, с которой осуществляется прогнозирование событий $q_{t+p}\in I^+$. В самом деле, при получении прогноза о том, что $q_{t+p}\in I^+$, необходимо принять дополнительные меры защиты от кибератак. Чем точнее такие прогнозы, тем реже будут ошибочно приниматься дополнительные меры противодействия кибератакам (ложные срабатывания). Чем меньше ложных срабатываний, тем эффективнее будет работать система защиты от кибератак. Для оценки соответствующей точности прогнозирования предложено использовать величину

$$pr^+=l^+/u^+, \quad (5)$$

где pr^+ – оценка точности прогнозирования событий $q_{t+p}\in I^+$, l^+ – число оправдавшихся прогнозов

того, что $q_{t+p} \in I^+$, u^+ – общее число сделанных прогнозов того, что $q_{t+p} \in I^+$, $0 \le pr^+ \le 1$.

Также нас интересует точность, с которой осуществляется прогнозирование событий $q_{t+p} \in I^-$. Здесь при получении прогноза о том, что $q_{t+p} \in I^-$, система защиты от кибератак продолжает функционировать в штатном режиме. Чем точнее такие прогнозы, тем реже будут возникать ситуации, когда по факту требовалось принятие дополнительных мер защиты от кибератак, но этого не было сделано. И это также влияет на эффективность системы защиты от кибератак. Для оценки соответствующей точности прогнозирования использовалась величина:

$$pr^- = l^- / u^- , \qquad (6)$$

где pr^- – оценка точности прогнозирования событий $q_{t+p} \in I^-$, l^- – число оправдавшихся прогнозов того, что $q_{t+p} \in I^-$, u^- – общее число сделанных прогнозов того, что $q_{t+p} \in I^-$, $0 \le pr^- \le 1$.

Еще одной важной величиной является общая точность интервального прогнозирования событий $q_{t+p} \in I^+$ и $q_{t+p} \in I^-$ без их непосредственного разделения. При этом необходимо, чтобы эта величина учитывала вероятностную природу получаемых оценок $\tilde{\rho}^+_{t+p}$ и $\tilde{\rho}^-_{t+p}$. То есть в некотором смысле «отражала» качество построенной и «обученной» модели. Для этой величины, характеризующей общую точность интервального прогнозирования, выбрана скоринговая оценка Брайера [27]:

$$bs = \frac{1}{u} \sum_{t=1}^{n} \left(\tilde{\rho}^+_{t+p} - \nu_{t+p} \right)^2 , \qquad (7)$$

где bs – оценка Брайера, u – общее число сделанных прогнозов того, что $q_{t+p} \in I^+$ или $q_{t+p} \in I^-$; ν_{t+p} – исход события (равен 1, если $q_{t+p} \in I^+$, и 0, если $q_{t+p} \in I^-$), $0 \le bs \le 1$.

Чем ближе значение величин pr^+ (5), pr^- (6) к единице, а значение bs (7) к нулю, тем точнее интервальное прогнозирование.

Для определения минимально допустимых значений pr^+ (5) и pr^-(6) мы воспользовались результатами работы [28]. В этой работе минимально допустимым значением считается такое (обозначим его как map), нижняя граница доверительного интервала которого больше значения 0,5 для заданного объёма выборки n. Это значение находится расчетным способом [28]. В нашем случае $n = 4000$, а рекомендованное значение $map \approx 0{,}51$. Если величина pr^+ или pr^- будет меньше значения map, такое интервальное прогнозирование следует признать неприемлемым, а модель прогнозирования неадекватной. Мы полагаем, что на практике эти значения должны быть как минимум не меньше 0,6, т.е. $pr^+ \ge 0{,}6$ и $pr^- \ge 0{,}6$.

Нам не удалось найти каких-либо работ в отношении минимально допустимого значения bs, поэтому по результатам предварительных экспериментов мы полагаем, что должно выполняться условие $bs \le 0{,}3$.

Величины (5)–(7) оценивались методом кроссвалидации по отдельным объектам обучающего множества для различных значений β. При этом значение f менялось от 1 до 10 и выбиралось такое значение, при котором оценка bs (7) была минимальной. Параметр $p = 1$.

Для реализации всех алгоритмов использовался свободно распространяемый язык программирования «R» [29]. При этом для реализации и ускорения отдельных функций и процедур использовался язык «C++». Для интеграции языка «R» и «C++» использовался пакет расширения «Rcpp» для «R» [30].

На рис. 3 и 4 для показателя SD и SA представлены значения величин pr^+ (5) и pr^- (6) в зависимости от значений параметра β.

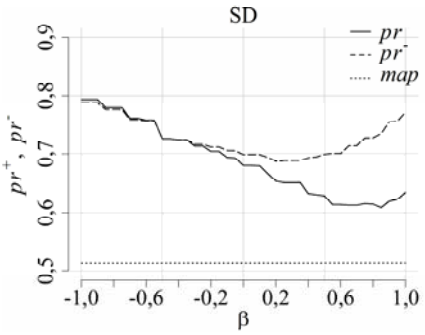

Рис. 3. График значений величин pr^+ (5) и pr^- (6) при изменении параметра β для показателя SD

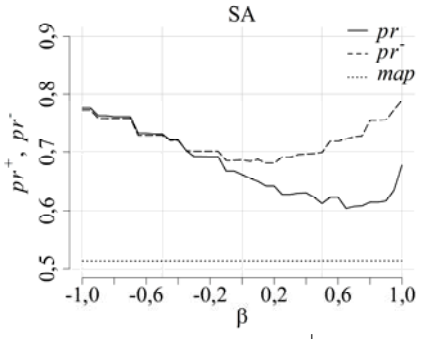

Рис. 4. График значений величин pr^+ (5) и pr^- (6) при изменении параметра β для показателя SA

Из графиков видно, что для каждого показателя все значения pr^+ и pr^- для различных β лежат выше минимально допустимого значения map и рекомендованного нами значения 0,6. При этом с увеличением значения β наблюдается тенденция к уменьшению точности прогнозирования событий $q_{t+p} \in I^+$ (уменьшается значение pr^+). Отсюда следует заключить, что с увеличением предустанов-

ленного уровня интенсивности кибератак становится сложнее прогнозировать события $q_{t+p} \in I^+$.

Также на точность интервального прогнозирования событий $q_{t+p} \in I^+$ влияет тот факт, что с увеличением значения β в обучающем множестве уменьшается число обучающих объектов с откликом $y_{t+p} = 1$ (т.е. когда $q_{t+p} \in I^+$), возникает «дисбаланс» обучающего множества и ухудшается качество «обучения» модели прогнозирования.

Некоторое возрастание значений pr^+ в конце графиков можно объяснить возрастающей погрешностью оценок этих величин, так как с линейным увеличением значения β число сделанных прогнозов того, что $q_{t+p} \in I^+$, уменьшается нелинейно (т.е. число сделанных прогнозов того, что $q_{t+p} \in I^+$, уменьшается быстрее, чем возрастает значение β).

В таблице для показателей SD и SA приведены минимальные и максимальные значения величин pr^+, pr^- и bs.

Минимальные и максимальные значения величин pr^+, pr^-, bs

Показатель	Величина	Минимальное значение	Максимальное значение
SD	pr^+	0,61	0,79
	pr^-	0,71	0,78
	bs	0,15	0,20
SA	pr^+	0,60	0,78
	pr^-	0,70	0,79
	bs	0,15	0,20

Все значения лежат в допустимых пределах, следовательно, интервальное прогнозирование показателей SD и SA следует признать адекватным и приемлемым. Таким образом, интервальное прогнозирование кибератак на основе ВНС демонстрирует приемлемую точность во всем диапазоне значений параметра β.

С учетом полученных результатов мы можем рекомендовать следующую схему противодействия кибератакам с учетом результатов интервального прогнозирования:

1. Определить значение β и соответственно значение предустановленного уровня интенсивности кибератак $\hat{q}$ (3). Например, это можно сделать эмпирически или методом экспертных оценок.

2. Сформировать обучающее множество для различных значений f (4), «обучить» ВНС и выбрать такое значение f, при котором по результатам кросс-валидации по отдельным объектам значение показателя bs (7) минимально.

3. Осуществить на основе «обученной» модели интервальный прогноз на один час вперед ($p = 1$). Если $q_{t+p} \in I^+$, то принять дополнительные меры противодействия кибератакам. Прогноз $q_{t+p} \in I^-$ игнорировать (штатный режим работы).

4. Через один час добавить в конец выборки показателя (1) новое значение об интенсивности кибератак ($n = n + 1$) и вернуться на этап 1.

Дополнительно следует добавить, что предустановленный уровень интенсивности кибератак $\hat{q}$ (3) может регулярно пересматриваться, например в связи с увеличением интенсивности кибератак на выбранный ОКИИ (о чем могут свидетельствовать слишком частые интервальные прогнозы того, что $q_{t+p} \in I^+$). Возможны дополнительные меры противодействия кибератакам, которые должны применяться не при первом попадании будущего значения в интервал I^+, а после некоторого числа попаданий. В этом направлении необходимы дополнительные исследования. Также дополнительные исследования необходимы в направлении формирования обучающего множества на предмет выявления репрезентативных образцов и его «сбалансированности». Возможно, это приведет к улучшению точности интервального прогнозирования.

Заключение

Анализ современных научных работ показал, что на протяжении нескольких лет научным сообществом ведутся исследовательские работы по прогнозированию кибератак различными методами с целью создания адекватных методов заблаговременной защиты от них. Важной задачей государственного уровня является обеспечение безопасности ОКИИ.

В данной работе:

1. Проведено исследование результатов интервального прогнозирования интенсивности кибератак посредством интеллектуального моделирования. В качестве интеллектуальной модели обосновано использование вероятностной нейронной сети с динамическим обновлением параметра сглаживания.

2. Предложенный подход продемонстрировал хорошую точность интервального прогнозирования выбранных нами показателей интенсивности кибератак. Следует отметить, что эти показатели являются нестационарными по параметру положения и параметру масштаба, что подчеркивает их непростую статистическую «природу» и сложные закономерности в происходящих во времени процессах кибератак.

3. Сформулированы необходимые практические рекомендации о применении результатов интервального прогнозирования для противодействия кибератакам на ОКИИ, обозначены возможные направления дальнейших исследований.

Следует отметить, что авторам подобные исследования, связанные с интервальным прогнозированием интенсивности кибератак, не известны, поэтому эту работу можно считать одной из первых. Надеемся, что полученные результаты найдут практическое применение и теоретическое развитие в сфере кибербезопасности ОКИИ.

Литература

1. Edgar T.W. Research Methods for Cyber Security / T.W. Edgar, D.O. Manz. – Rockland: Syngress, 2017. – 428 p.

2. Loukas G. Cyber-Physical Attacks. A Growing Invisible Threat. – Oxford: Butterworth-Heinemann, 2015. – 270 p.

3. Кожевникова А.С. Особенности и тенденции развития информационной безопасности в Российской Федерации / А.С. Кожевникова, В.Н. Лопин // Сб. науч. трудов 6-й Междунар. науч.-практ. конф. «Инновации, качество и сервис в технике и технологиях». – Курск: ЗАО «Университетская книга», 2016. – С. 122–125.

4. Отчет безопасности промышленных предприятий и IOT: прогноз на 2018 год [Электронный ресурс]. – Режим доступа: https://ics-cert.kaspersky.ru/reports/2017/11/30/industrial-enterprise-and-iot-security-threats-forecast-for-2018, свободный (дата обращения: 10.10.2017).

5. Указ Президента РФ «Об утверждении доктрины информационной безопасности Российской Федерации» от 5.12.2016 № 646 [Электронный ресурс]. – Режим доступа: http://pravo.gov.ru/proxy/ips/?docbody=&firstDoc=1&lastDoc=1&nd=102417017, свободный (дата обращения: 01.11.2017).

6. Федеральный закон «О безопасности критической информационной инфраструктуры Российской федерации» от 26.07.2017 № 187-ФЗ [Электронный ресурс]. – Режим доступа: http://pravo.gov.ru/laws/acts/59/49565545 10601047.html, свободный (дата обращения: 03.11.2017).

7. Емельянченко В.А. Кибертерроризм как угроза информационной безопасности / В.А. Емельянченко, Т.О. Игнатенко // Форум молодых ученых. – 2017. – № 5 (9). – С. 730–734.

8. Jones M. Cyber-Attack Forecast Modeling and Complexity Reduction Using a Game-Theoretic Framework / M. Jones, G. Kotsalis, J.S. Shamma // Tarraf D. (eds) Control of Cyber-Physical Systems. Lecture Notes in Control and Information Sciences. – Heidelberg: Springer, 2013. – 380 p.

9. Петренко С.А. Концепция раннего распознавания и предупреждения компьютерного нападения / С.А. Петренко, А.С. Петренко // Матер. Всерос. науч.-практ. конф. «Информационные системы и технологии в моделировании и управлении». – 2016. – С. 82–86.

10. Петренко С.А. Национальная система раннего предупреждения о компьютерном нападении / С.А. Петренко, Д.Д. Ступин. – Иннополис: Изд. дом «Афина», 2017. – 440 с.

11. Выписка из концепции государственной системы обнаружения, предупреждения и ликвидации последствий компьютерных атак на информационные ресурсы Российской Федерации, утв. Президентом Российской Федерации от 12.12.2014 № К1274 [Электронный ресурс]. – Режим доступа: http://www.fsb.ru/files/PDF/Vipiska_iz_koncepcii.pdf, свободный (дата обращения: 12.10.2017).

12. Gandotra E. Computational Techniques for Predicting Cyber Threats / E. Gandotra, D. Bansal, S. Sofat // Procceediongs of Intelligent Computing, Communication and Devices. – 2015. – P. 247–253.

13. Kim Y. A probabilistic approach to estimate the damage propagation of cyber attacks / Y. Kim, T. Lee, H. In, Y. Chung, I. Kim, D. Baik // Proceeding of the 8th International Conference on Information Security and Cryptology. – 2005. – P. 175–185.

14. Wu J. Cyber Attacks Prediction Model Based on Bayesian Network / J. Wu // Proceedings of the 2012 IEEE 18th International Conference on Parallel and Distributed Systems. – 2012. – P. 730–731.

15. Haitao D. Probabilistic Modeling and Inference for Obfuscated Network Attack Sequences [Электронный ресурс]. – Режим доступа: http://scholarworks.rit.edu/theses/8331, свободный (дата обращения: 15.09.2017).

16. Zhan Z. Predicting cyber attack rates with extreme values / Z. Zhan, M. Xu, S. Xu. // IEEE Transactions on Information Forensics and Security. – 2015. – №. 10 (8). – P. 1666–1677.

17. Werner G. Time series forecasting of cyber attack intensity / G. Werner, S. Yang, K. McConky // Proceedings of the 12th Annual Conference on Cyber and Information Security. – 2017. – P. 224–240.

18. Краковский Ю.М. Прикладные аспекты применения интервального прогнозирования в системном анализе / Ю.М. Краковский, А.Н. Лузгин // Современные технологии. Системный анализ. Моделирование. – 2017. – № 2 (54). – С. 115–121.

19. Sokol P. Prediction of Attacks Against Honeynet Based on Time Series Modeling / P. Sokol, A. Gajdos // Proceedings of the Computational Methods in Systems and Software. – 2017. – P. 360–371.

20. Data driven Security [Электронный ресурс]. – Режим доступа: http://datadrivensecurity.info/blog/posts/2014/Jan/blander-part1, свободный (дата обращения: 27.10.2017).

21. Kargapoltsev S.K. Nonparametric classification of technical condition parameters based on shift and scale tests / S.K. Kargapoltsev, Y.M. Krakovsky, A.N. Luzgin // 2017 International Conference on Industrial Engineering, Applications and Manufacturing. – 2017. – P. 1–5.

22. Specht D.H. Probabilistic Neural Networks / D.H. Specht // Neural Networks. – 1990. – № 3. – P. 109–118.

23. Kargapoltsev S.K. A dynamic updating algorithm of smoothing parameter values of probabilistic neural networks / S.K. Kargapoltsev, Y.M. Krakovsky, A.V. Lukyanov, A.N. Luzgin // Far East Journal of Electronics and Communications. – 2017. – Vol. 17, № 4. – P. 909–914.

24. Probabilistic neural network [Электронный ресурс]. – Режим доступа: https://en.wikipedia.org/wiki/Probabilistic_neural_network, свободный (дата обращения: 02.04.2018).

25. Краковский Ю.М. Исследование алгоритмов оптимизации размерности обучающих векторов вероятностных моделей интервального прогнозирования на основе методов кросс-валидации // Ю.М. Краковский, А.Н. Лузгин // Baikal letter DAAD. – 2017. – № 1. – С. 23–44.

26. Краковский Ю.М. Алгоритм интервального прогнозирования на основе линейной модели авторегрессии / Ю.М. Краковский, А.Н. Лузгин // Вопросы естествознания. – 2016. – № 3 (11). – С. 16–24.

27. Hoffman R.R. Minding the Weather: How Expert Forecasters Think / R.R. Hoffman, D.S., LaDue H.M. Mogil, J.G. Trafton. – London: MIT Press, 2017. – 488 p.

28. Краковский Ю.М. Стохастический критерий оценки приемлемой точности вероятностного бинарного прогнозирования динамических показателей / Ю.М. Краковский, А.Н. Лузгин // Вестник ВГУ. Сер.: Системный анализ и информационные технологии. – 2017. – № 2. – С. 98–104.

29. The R project of statistical computing [Электронный ресурс]. – Режим доступа: http://www.r-project.org, свободный (дата обращения: 27.10.2017).

30. Seamless R and C++ Integration [Электронный ресурс]. – Режим доступа: https://cran.r-project.org/web/packages/Rcpp/index.html, свободный (дата обращения: 28.10.2017).

Краковский Юрий Мечеславович
Д-р техн. наук, профессор каф. информационных систем и защиты информации
Иркутского гос. ун-та путей сообщения (ИрГУПС)
Чернышевского ул., д. 15, г. Иркутск, Россия, 664074
Тел.: +7 (395-2) 63-83-10
Эл. почта: kum@stranzit.ru

Курчинский Борис Валентинович
Начальник управления специального обеспечения администрации г. Иркутска
Ленина ул., д. 14б, г. Иркутск, Россия, 664025
Тел.: +7 (395-2) 52-02-02
Эл. почта: kurchinsky_b@admirk.ru

Лузгин Александр Николаевич
Канд. техн. наук, зам. начальника управления специального обеспечения администрации г. Иркутска
Ленина ул., д. 14б, г. Иркутск, Россия, 664025
https://orcid.org/0000-0002-2669-3787
Тел.: +7 (395-2) 52-00-54
Эл. почта: alexln@mail.ru

Krakovsky Y.M., Kurchinsky B.V., Luzgin A.N.

Cyber-attack intensity interval forecasting on objects of critical information infrastructure

In the modern world, cybersecurity issues occupy one of the key and extremely significant strategic niches in a state planning and management system. Cyber-attacks occur daily, and their number grows exponentially. In these conditions, the cybersecurity issues regarding to critical information infrastructure objects of a state are becoming especially topical. The relevance of research in the direction of creating and improving technologies for protection against cyber-attacks on relevant facilities is beyond doubt. In this paper are investigated the results of interval forecasting for cyber-attack intensity through an intelligent algorithm based on a probabilistic neural network with a dynamic updating value of the smoothing parameter. The data on the hourly cyber-attack intensity obtained by means of honeypots from March to September 2013 are used as initial data. Obtained results testify to the high forecasting accuracy of the proposed algorithm. According to the study results, the authors give necessary practical recommendations about interval forecasting results application when protecting against cyber-attacks on critical information infrastructure objects.

Keywords: interval forecasting, cyber-attacks, critical information infrastructure.
doi: 10.21293/1818-0442-2018-21-1-71-79

References

1. Edgar T.W. *Research Methods for Cyber Security.* Rockland, Syngress, 2017. 428 p.

2. Loukas G. *Cyber-Physical Attacks. A Growing Invisible Threat.* Oxford, Butterworth-Heinemann, 2015. 270 p.

3. Kozhevnikova A.S., Lopin V.N. Osobennosti i tendencii razvitiya informacionnoj bezopasnosti v Rossijskoj Federacii [Features and trends in the development of information security in the Russian Federation]. Sbornik nauchnyh trudov 6-oj Mezhdunarodnoj nauchno-prakticheskoj konferencii «Innovacii, kache-stvo i servis v tekhnike i tekhnologiyah» [*Proc. of the 6th International Scientific and Practical Conference «Innovations, quality and service in engineering and technology»*]. Kursk, CJSC University Book Publ., 2016, pp. 122–125.

4. Otchet bezopasnosti promyshlennykh predpriyatii i IOT: prognoz na 2018 god. (In Russ.). Available at: https://ics-cert.kaspersky.ru/reports/2017/11/30/industrial-enterprise-and-iot-security-threats-forecast-for-2018 (accessed: 10 Oct. 2017).

5. Ukaz Prezidenta RF «Ob utverzhdenii doktriny informatsionnoi bezopasnosti Rossiiskoi Federatsii» ot 05.12.2016 № 646. (In Russ.). Available at: http://pravo.gov.ru/proxy/ips/?docbody=&firstDoc=1&lastDoc=1&nd=102417017 (accessed: 01 November 2017).

6. Federal'nyi zakon «O bezopasnosti kriticheskoi informatsionnoi infrastruktury Rossiiskoi federa-tsii» ot 26.07.2017 №187-FZ. (In Russ.) Available at: http://pravo.gov.ru/laws/acts/59/4956554510601047.html (accessed: 03 Nov. 2017).

7. Emel'yanchenko V.A., Ignatenko T.O. Cyberterrorism as a threat to information security. *Forum of young scientists,* 2017, no. 5 (9), pp. 730–734 (In Russ.).

8. Jones M., Kotsalis G., Shamma G.S. *Cyber-Attack Forecast Modeling and Complexity Reduction Using a Game-Theoretic Framework.* In: Tarraf D. (eds) Control of Cyber-Physical Systems. Lecture Notes in Control and Information Sciences. Heidelberg: Springer, 2013. 380 p.

9. Petrenko S.A., Petrenko A.S. Kontseptsiya rannego raspoznavaniya i preduprezhdeniya komp'yuternogo napadeniya [Concept of early detection and prevention of computer attack]. Materialy vserossiiskoi nauchno-prakticheskoi konfe-rentsii «Informatsionnye sistemy i tekhnologii v modelirovanii i upravlenii» [*Materials of the All-Russian Scientific and Practical Conference «Information Systems and Technologies in Modeling and Control»]*. Saint Petersburg, 2016, pp. 82–86.

10. Petrenko S.A., Stupin D.D. Natsional'naya sistema rannego preduprezhde-niya o komp'yuternom napadenii [National Early Warning System on Computer Attack]. *Innopolis*, Publishing House «Athena», 2017. 440 p. (In Russ.).

11. Vypiska iz kontseptsii gosudarstvennoi sistemy obnaruzheniya, preduprezhdeniya i likvidatsii posledstvii komp'yuternykh atak na informatsionnye resursy Rossiiskoi federatsii, utverzhdennoi Prezidentom Rossiiskoi Federatsii ot 12.12.2014 № K1274. (In Russ.) Available at: http://www.fsb.ru/files/PDF/Vipiska_iz_koncepcii.pdf (accessed: 12 October 2017).

12. Gandotra E. Bansal D., Sofat S. Computational Techniques for Predicting Cyber Threats. *Proc. of Intelligent Computing, Communication and Devices*, 2015, pp. 247–253.

13. Kim Y. A., Lee T., In H., Chung Y., Kim I., Baik D. Probabilistic approach to estimate the damage propagation of cyber attacks. *Proc. of the 8th International Conference on Information Security and Cryptology*, 2005, pp. 175–185.

14. Wu J. Cyber Attacks Prediction Model Based on Bayesian Network. *Proceedings of the 2012 IEEE 18th International Conference on Parallel and Distributed System*s, 2012, pp. 730–731.

15. Haitao D. Probabilistic Modeling and Inference for Obfuscated Network Attack Sequences. Available at: http://scholarworks.rit.edu/theses/8331 (accessed: 15 Sep. 2017).

16. Zhan Z., Xu M., Xu S. Predicting cyber attack rates with extreme values. *IEEE Transactions on Information Forensics and Security,* 2015, No. 10(8), pp. 1666–1677.

17. Werner G., Yang S., McConky K. Time series forecasting of cyber attack intensity. *Proceedings of the 12th Annual Conference on Cyber and Information Security*, 2017, pp. 224–240.

18. Krakovsky Y.M., Luzgin A.N. Applied aspects of application of interval forecasting of dynamic indicators in system analysis. *Modern technology. System analysis. Modeling,* 2017, no. 2(54), pp. 115–121. (In Russ.).

19. Sokol P., Gajdos A. Prediction of Attacks Against Honeynet Based on Time Series Modeling. *Proc. of the Computational Methods in Systems and Software,* 2017, pp. 360–371.

20. Data driven Security. Available at: http://datadrivensecurity.info/blog/posts/2014/Jan/blander-part1 (accessed 27 October 2017)

21. Kargapoltsev S.K., Krakovsky Y.M., Luzgin A.N. Nonparametric classification of technical condition parameters based on shift and scale tests. *2017 International Conference on Industrial Engineering, Applications and Manufacturing*, 2017, pp.1–5.

22. Specht D.H. Probabilistic Neural. *Neural Networks,* 1990, no. 3, pp. 109–118.

23. Kargapoltsev S.K., Krakovsky Y.M., Lukyanov A.V., Luzgin A.N. A dynamic updating algorithm of smoothing parameter values of probabilistic neural networks. *Far East Journal of Electronics and Communications*, 2017, vol. 17, no. 4, pp. 909–914.

24. Probabilistic neural network. Available at: https://en.wikipedia.org/wiki/Probabilistic_neural_network (accessed 02 April 2018).

25. Krakovsky Y.M., Luzgin A.N. The study of optimization algorithms of dimension for learning vectors of interval forecasting probability models based on cross-validation methods. *Baikal letter DAAD*, 2017, no. 1, pp. 23–44. (In Russ.).

26. Krakovsky Y.M., Luzgin A.N. Interval forecasting algorithm of dynamic indicators based on autoregressive linear model. *Natural science issues*, 2016, no. 3(11), pp.16–24.

27. Hoffman R.R., LaDue D.S., Mogil H.M., Trafton J.G. *Minding the Weather: How Expert Forecasters Think.* London, MIT Press, 2017. 488 p.

28. Krakovsky Y.M., Luzgin A.N. Stochastic criterion for estimating the acceptable accuracy of probabilistic binary forecasting of dynamic indicators. *Vestnik VSU, series: System analysis and information technology*, 2017, no. 2, pp. 98–104. (In Russ.).

29. The R project of statistical computing. Available at: http://www.r-project.org (accessed: 27 October 2017)

30. Seamless R and C++ Integration. Available at: https://cran.r-project.org/web/packages/Rcpp/index.html (accessed: 28 October 2017).

Yuri M. Krakovsky
Doctor of Engineering Sciences, professor,
Department of Information Systems and Information Security,
Irkutsk State University of Railway Transport,
15, Chernyshevsky st., Irkutsk, Russia, 664074
Phone: +7 (395-2) 63-83-10
Email: kum@stranzit.ru

Boris V. Kurchinsky
Head of Special Providing Department
of Irkutsk City Administration
14b, Lenin st., Irkutsk, Russia, 664025
Phone: +7 (395-2) 52-02-02
Email: kurchinsky_b@admirk.ru

Aleksandr N. Luzgin
PhD, Deputy Head of Special Providing Department
of Irkutsk City Administration
14b, Lenina st., Irkutsk, Russia, 664025
https://orcid.org/0000-0002-2669-3787
Phone: +7 (395-2) 52-00-54
Email: alexln@mail.ru