РАЗРАБОТКА ПРОГРАММНОГО СРЕДСТВА ПРОГНОЗИРОВАНИЯ ИНЦИДЕНТОВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

А. Ю. Ермакова[1], А. Б. Лось[2]

[1]Российский технологический университет МИРЭА, Москва, Россия
[2]Национальный исследовательский университет «Высшая школа экономики», Москва, Россия
E-mail: a.alla1105@yandex.ru, alexloss2011@mail.ru

В докладе рассматриваются вопросы разработки программного средства для построения прогнозных моделей появления инцидентов информационной безопасности. Под инцидентами информационной безопасности далее понимаются события, приводящие к нарушению штатной работы информационной системы (ИС): компьютерные атаки, попытки несанкционированного доступа, хищения денежных средств, сбои в работе системы. Представлено описание программного средства, позволяющего по данным об интенсивности предыдущих инцидентов строить непрерывные аппроксимирующие функции, наиболее близко расположенные к ключевым точкам (данным об инцидентах) и сохраняющие на прогнозируемом участке статистические свойства исходных данных. На основании указанных функций далее возможно построение функции рисков и оценки времени безопасной эксплуатации ИС.

DEVELOPMENT OF A SOFTWARE TOOL FOR PREDICTING INFORMATION SECURITY INCIDENTS

A. Y. Ermakova, A. B. Los

The report discusses the issues of developing a software tool for the construction of predictive models of the occurrence of information security incidents. Information security incidents are further understood as events leading to disruption of the regular operation of the information system (IS): computer attacks, unauthorized access attempts, theft of funds, system failures. A description of a software tool is presented that allows, based on data on the intensity of previous incidents, to construct continuous approximating functions that are most closely located to key points (incident data) and preserve the statistical properties of the source data in the predicted area. Based on these functions, it is further possible to construct a risk function and estimate the time of safe operation of the IC.

Введение. В докладе излагаются результаты построения программного средства с целью прогнозирования появления различных компьютерных инцидентов. Под инцидентами информационной безопасности далее понимаются события, приводящие к нарушению штатной работы информационной системы: компьютерные атаки, попытки несанкционированного доступа, хищения денежных средств, сбои в работе системы.

Необходимость в создании такого средства возникает при разработке методов оценки защищенности информационных систем, требования к которым отражены в стандартах ИБ ([1-2]).

Основной подход, заложенный в указанных стандартах, носит название

риск - ориентированного, и состоит в составлении перечня актуальных угроз $\{y_i\}$, вероятность реализации которых $p(y_i)$ больше *0*, численной оценке вероятностей $p(y_i)$ и оценке возможного ущерба u_i при их успешной реализации. Далее составляется функция рисков R, представляющая собой усредненное значение ущерба при реализации угроз, определение максимально допустимого ущерба R_0 и, в случае выполнения условия $R \le R_0$ принимается решение о защищенности ИС. Очевидным недостатком такого подхода является невозможность оценки срока безопасной работы ИС, а именно: в течение какого времени будет выполнено условие защищенности $R \le R_0$. Одной из возможностей модификации традиционного риск - ориентированного подхода является введение зависимости величин $p(y_i) = p_{y_i}(t)$ и $u_i = u_i(t)$ от времени *t* ([3]). В этом случае, приравнивая зависящую от времени функцию рисков $R(t)$ величине максимального риска R_0, получаем уравнение относительно переменной *t*, решение которого T_0, в свою очередь, можно рассматривать как оценку времени, в течение которого возможна безопасная работа ИС. Далее возникает задача построения таких временных функций рисков и задача построения оценок параметра T_0.

Одним из подходов к построению временных функций рисков для оценки защищенности ИС, является разработка прогнозных моделей интенсивности появления инцидентов, приводящих к нарушению информационной безопасности, по известным данным за прошлый период. Примеры применения данного подхода представлены в работах [4-10].

В данной работе приведено описание алгоритма разработки прогнозных моделей путем построения непрерывных аппроксимирующих функций наиболее близко отстоящих от заданных значений (ключевых точек) – метод МНК. В качестве прогнозных величин берутся значения построенной функции на временном интервале, где ищется прогноз. При этом в традиционный алгоритм МНК добавлено условие сохранения статистических характеристик исходных данных. Это условие весьма важно, поскольку в большинстве методов аппроксимации за пределами заданного для прогноза интервала, поведение аппроксимирующей функции не соответствует исходным данным. Ярким примером может служить метод аппроксимации на основе полиномов Лагранжа, при котором аппроксимирующая функция за пределами исходного интервала очень быстро уходит в бесконечность и не может использоваться для прогнозирования.

В работе дано описание структуры алгоритма, его интерфейса, а также приведен пример построения с его помощью прогнозной функции количества финансовых операциях без согласия клиентов на основе данных за 2019-2022 года, публикуемых ЦБ РФ ([13]). Построен прогноз количества финансовых операций без согласия клиентов на 2024-2025 года. На основе прогнозной функции построены теоретические оценки времени безопасной работы ИС.

Алгоритм построения непрерывной аппроксимирующей функции

Метод наименьших квадратов (далее – МНК) является одним из способов

аппроксимации таблично заданной функций некоторым базисным набором функций, выбор которых основан на определенном критерии, позволяющем выявить особенности заданных входных значений ([12]).

Исходными данными для построения непрерывной прогнозной функции *F(t)* являются значения состояния системы в предыдущие моменты времени. В нашем случае это могут быть данные о числе инцидентов в ИС, за какой либо предыдущий период.

В соответствии с предлагаемым алгоритмом, непрерывная аппроксимирующая функция *F(t)* ищется как линейная комбинация, так называемых, базовых функций $\{\varphi_i\}$

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

где $\varphi_i(t)$– элементарные непрерывные функции: $\alpha_1 \cdot sin(\beta_1 \cdot t^k)$, $\alpha_2 \cdot cos(\beta_2 \cdot t^k)$, $\alpha_3 \cdot e^{\gamma t}$, $\alpha_4 \cdot t^{\beta n}$, ($\alpha_i$, β_i, γ, κ, n – некоторые коэффициенты) и другие аналогичные функции.

Очевидно, что в этом случае $F(t)$ также является непрерывной.

Алгоритм начинается с выбора количества и вида базовых функций. Затем с применением метода МНК определяются значения коэффициентов c_i при которых достигается минимальное расстояние от значений функции $F(t)$ до значений в ключевых точках.

Заключительным шагом алгоритма является проверка соответствия статистических характеристик прогнозных значений исходным данным. Каждое из прогнозных значений должно отклоняться от эмпирического среднего исходных данных на более чем на $n \cdot \sigma$, где σ- квадратный корень из эмпирической дисперсии, n – параметр алгоритма, обычно, полагается $n \geq 3$. В случае невыполнения данного условия производится замена набора базовых функций.

Полученная таким образом функция далее применяется для построения прогнозных значений состояния рассматриваемой системы, в нашем случае – для прогнозирования появления инцидентов ИБ.

Структура разработанной программы.

Формат данных. Исходные данные (ключевые точки) создаются в виде таблицы значений функции в формате Excel, пример которой представлен в таблице 1.

В представленной таблице первый столбец – значения временного интервала t_i, второй столбец – значения искомой функции y_i. Количество ключевых точек равно количеству представленных в таблице строк, поэтому ниже этой таблицы не должно быть занятых ячеек.

Начало работы с программой предполагает нажатие кнопки «Чтение таблицы» и выбора файла исходных данных. В случае успешного чтения, первая таблица заполняется данными из Excel файла. В случае ошибки при чтении, либо в случае появления пустого файла, появляется сообщение «Ошибка чтения из файла». Основные функции программы становятся доступными после успешного ввода данных.

Таблица 1

Структура начальных данных

№	A	B
1	1	0, 4815
2	2	0,8354
3	3	0,7453
4	4	-0,5432
5	5	-0,6234
6	6	0,3245
7	7	-0,3853
8	8	0,7546
9	9	0,6434
10	10	-0,9651
11	11	-0.2567
12	12	0,9765
13	13	0,5574
14	14	-0,8346
15	15	0,7239

Интерфейс программы

Интерфейс разработанной программы представлен на рис. 1.

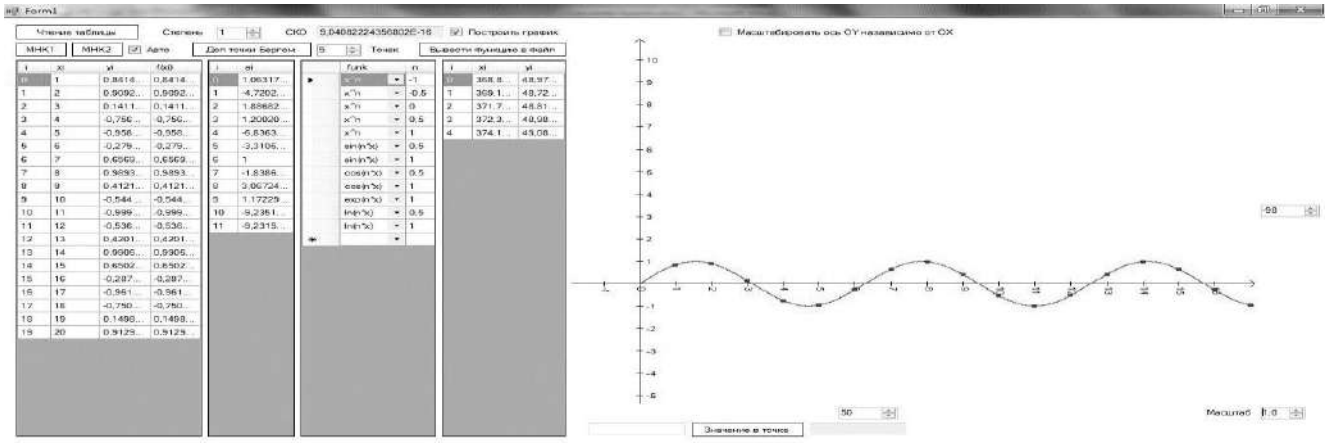

Рис. 1. Интерфейс программы

Клавиша «МНК1» предназначена для запуска алгоритма поиска аппроксимирующей функцией в виде многочлена методом наименьших квадратов с выбранными базисными функциями. Степень многочлена можно менять в поле

«Степень». При этом в первой таблице в четвёртом столбце $F(t_i)$ появятся значения аппроксимирующей функции в заданных точках. Далее строится график искомой функции и посчитывается среднеквадратическое отклонение по узлам данных. Значения коэффициентов многочлена выводятся во вторую таблицу. При изменении степени, многочлен пересчитывается автоматически, и все изменения отображаются на графике и в таблицах.

Клавиша «МНК2» без функции «Авто» включает алгоритм построения аппроксимирующей функции методом наименьших квадратов с базисными функциями, вид и параметры которых находятся в третьей таблице. При этом количество функций, их порядок, комбинация и дополнительные параметры выбираются вручную и могут быть произвольными. После завершения работы алгоритма, результат выводится аналогично описанной выше схеме, однако, коэффициенты во второй таблице уже соответствуют коэффициентам базисных функций построенной аппроксимирующей функции.

Если при нажатии клавиши «МНК2» нажата клавиша «Авто», то выбор базисных функций и соответствующих параметров определяется автоматически путём перебора (с применением циклов) в заданных границах возможных комбинаций параметров. В соответствии с методом МНК, критерием для выбора оптимальной структуры аппроксимирующей функции служит минимальное среднеквадратическое отклонение в узлах таблично заданной функции среди рассмотренных вариантов.

Завершающей процедурой алгоритма служит задание прогнозных временных значений и проверка попадания прогнозных величин для выбранных временных точек в интервал от $m-n\cdot\sigma$ до $m+n\cdot\sigma$, где m и σ – соответственно эмпирическое среднее и квадратный корень из дисперсии исходных данных. В случае невыполнения последнего условия алгоритм возвращается в начало к выбору базисных функций и их параметров. Вывод результатов аналогичен предыдущему случаю, при этом разница состоит в том, что оптимальный набор базисных функций автоматически появляется в третьей таблице.

Вывод построенной функции осуществляется клавишей «Вывести функцию в файл», с помощью которой найденная функция выводится в указанный файл. Пример построенной функции приведен на рис. 2.

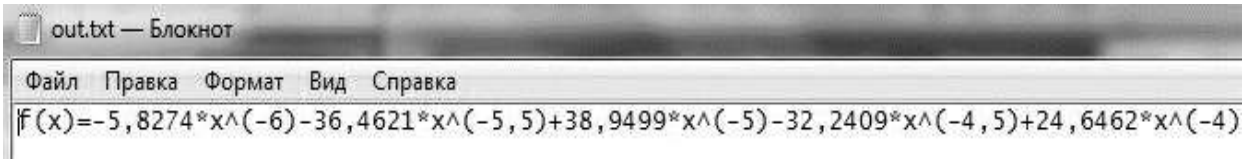

Рис. 2. Пример построенной «аппроксимирующей» функции.

Пример построения графика указанной функции приведен на рис. 3.

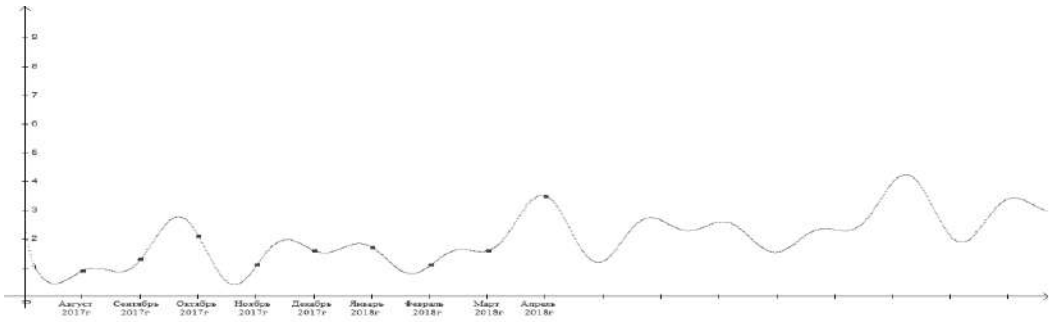

Рис. 3. Пример построения графика аппроксимирующей функции.

Построение прогнозной модели инцидентов - количества финансовых операций без согласия клиента

В качестве примера предлагаемого подхода приведем результаты эксперимента по построению прогнозной модели инцидентов информационной безопасности - количества финансовых операций без согласия клиентов.

В таблице 1 приведена поквартальная статистика количества финансовых операций без согласия клиентов за последние 4 года. Данные взяты из официального сайта Банка России [13], на котором приводятся отчеты об инцидентах информационной безопасности.

Таблица 2

Количество финансовых операций без согласия клиента

№	Период	Количество финансовых операций без согласия клиента (тыс.)
1	1 квартал 2019 г.	133
2	2 квартал 2019 г.	137
3	3 квартал 2019 г.	163
4	1 квартал 2020 г.	170
5	2 квартал 2020 г.	193
6	3 квартал 2020 г.	182
7	1 квартал 2021 г.	238
8	2 квартал 2021 г.	237
9	3 квартал 2021 г.	256
10	1 квартал 2022 г.	258
11	2 квартал 2022 г.	211
12	3 квартал 2022 г.	230

В рассматриваемом эксперименте для построения прогнозной функции *F(t)* использовались данные за период с 1 квартала 2019 года по 4 квартал 2022 года, за нулевое значение по оси OX принята дата - 4 квартал 2018 года.

Для данного эксперимента аппроксимирующая функция имела вид:

$$F(t) = 28.79 \cdot \ln[t] + 379.34 - 775.22 \cdot \frac{1}{t} + 534.95 \frac{1}{t^2} - 6.94 \cdot \sin[2 \cdot t] + 9.31 \cdot \cos[2 \cdot t] -$$

$$-13.15 \cdot \cos[4 \cdot t] - 5.84 \cdot \cos[3 \cdot t] + 15.34 \cdot \sin[t] - 14.37 \cdot \cos[t] + 1.62 \cdot \sin[3 \cdot t] +$$

$$+15.54 \cdot \sin[4 \cdot t]$$

График, прогнозной функции *F(t)* представлен ниже на рис. 4.

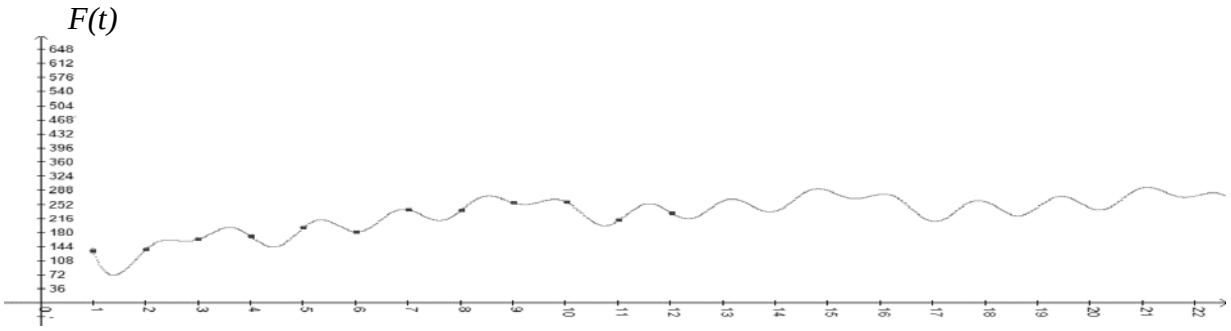

Рис. 4. График функции *y=F(t).*

В табл. 3 приведены результаты эксперимента по построению прогнозных значений количества финансовых операций без согласия клиента на период с 1 квартала 2024 года по 4 квартал 2025 года.

Таблица 3

Прогноз количества финансовых операций без согласия клиента

№	Период	Количество финансовых операций без согласия клиента (тыс.)
1	1 квартал 2024 г.	260,6
2	2 квартал 2024 г.	234,6
3	3 квартал 2024 г	287
4	4 квартал 2024 г.	278
5	1 квартал 2025 г.	238
6	2 квартал 2025 г.	260
7	3 квартал 2025 г.	253
8	4 квартал 2025 г.	274

Пример построения оценки времени безопасной работы ИС.

Для оценки времени безопасной работы ИС аппроксимирующую функцию *F(t)* представим в виде:

$$F(t)=\alpha_1\cdot\ln[t]+\alpha_2-\alpha_3\cdot\frac{1}{t}+\alpha_4\cdot\frac{1}{t^2}-\alpha_5\cdot\sin[2\cdot t]+\alpha_6\cdot\cos[2\cdot t]-\alpha_7\cdot\cos[4\cdot t]-\alpha_8\cdot\cos[3\cdot t]+$$

$$+\alpha_9\sin[t]-\alpha_{10}\cdot\cos[t]+\alpha_{11}\cdot\sin[3\cdot t]+\alpha_{12}\cdot\sin[4\cdot t],$$

где α_i - соответствующие коэффициенты.

Нетрудно видеть, что при $t\geq 1$ справедливо неравенство:

$$F(t)\leq\alpha_1\cdot\ln[t]+\alpha_{13},$$

где $\alpha_{13}=\sum_{i=2}^{12}\alpha_i$ - сумма коэффициентов.

Обозначим далее через *N* - число ИС, участвовавших в эксперименте, *U* – средний ущерб от финансовой операции без согласия клиента.

Тогда функция рисков *R(t)* для одной ИС принимает вид:

$$R(t) = \frac{1}{N} \cdot U \cdot \left(\alpha_1 \cdot \ln[t] + \alpha_{13} \right),$$

а оценка времени T_0 безопасной работы ИС может быть найдена как решение уравнения

$$R_0 = \frac{1}{N} \cdot U \left(\alpha_1 \cdot \ln[t] + \alpha_{13} \right),$$

где R_0 - допустимый уровень ущерба (потерь).

Из последнего уравнения находим оценку T_0

$$T_0 = \exp\left\{ \frac{N \cdot R_0 / U - \alpha_{13}}{\alpha_1} \right\}$$

При задании соответствующих значений величин N, R_0 и U могут быть получены численные значения времени безопасной работы информационной системы T_0.

Заключение. В докладе представлены результаты разработки программного средства для построения прогнозных моделей инцидентов информационной безопасности. Представлена структура разработанного средства и приведен пример построения с его помощью прогнозной модели инцидентов, связанных с проведением финансовых операций без согласия клиентов, а также пример построения оценки времени безопасной работы информационной системы.

СПИСОК ЛИТЕРАТУРЫ

1. ГОСТ Р ИСО/МЭК 13335-1:2006. Методы и средства обеспечения безопасности. Концепция и модели менеджмента безопасности информационных и телекоммуникационных технологий. Электронный фонд правовых и нормативно-технических документов. [Электроный ресурс]. URL: https://docs.cntd.ru/document/1200048398?ysclid=lq9icrm375902759376 (дата обращения 23.10.2023).

2. ГОСТ Р ИСО/МЭК 13335-3:2007. Методы и средства обеспечения безопасности. Методы менеджмента безопасности информационных технологий. Федеральное агентство по техническому регулированию и метрологии. [Электроный ресурс]. URL: https://altell.ru/legislation/standards/13335-3.pdf?ysclid=lq9igeanrm347850898 (дата обращения 16.10.2023).

3. *Лось А. Б. Кабанов А. С., Трунцев В. И.* Временная модель оценки риска нарушения информационной безопасности // Доклады ТУСУР. 2012. № 1. Ч. 2. С. 87-91.

4. *Ермакова А. Ю.* Оценка качества прогнозирования динамики изменения валютных курсов на основе построения аппроксимирующих функций // Качество. Инновации. Образование. 2013. № 2 (93). С. 71-79.

5. *Ермакова А. Ю.* Исследование качества прогнозирования биржевых курсов драгоценных металлов // Качество. Инновации. Образование. 2014. № 1 (104). С. 49-56.

6. *Ермакова А. Ю.* Построение прогнозной модели динамики изменения цен на древесину // Лесной Вестник 2016. № 6. С.88-97.

7. *Ермакова А. Ю.* Разработка методов прогнозирования на примере анализа средств вычислительной техники // Промышленные АСУ и контроллеры. 2017. № 1. С. 28-34.

8. *Ермакова А. Ю.* Об оценке точности прогнозирования состояния динамической системы методом построения аппроксимирующих функций // Промышленные АСУ и контроллеры. 2018. № 5. С. 36-42.

9. *Ермакова А. Ю. Лось А. Б.* Исследование прогнозных моделей динамической системы на примере прогноза инцидентов информационной безопасности // Компьютерные науки и информационные технологии: сборник статей междун. науч.-практич. конф. 2018. С. 144-149.

10 *Ермакова А. Ю.* Об одном подходе к оценке защищенности информационной системы на основе анализа инцидентов // Системы высокой доступности. 2018. № 4. С. 32-35.

11. *Рыбников К. К.* Введение в дискретную математику и теорию решения экстремальных задач на конечных множествах / М. : Гелиос АРВ, 2010. 318 с.

12. Обзор отчетности ЦБ об инцидентах информационной безопасности при переводе денежных средств. [Электронный ресурс]. URL: https://www.fzlabs.ru/news/czb-opublikoval-obzor-otchetnosti-ob-inczidentah-informaczionnoj-bezopasnosti-pri-perevode-denezhnyh-sredstv/ (дата обращения 17.09.2023).