Кластеризация угроз и идентификация рисков нарушения информационной безопасности опасных производственных объектов

М.А. Тукмачева, А.В. Шестаков, К.З. Билятдинов

***Аннотация* — представлены результаты разработки проактивной комплексной модели выявления и идентификации угроз нарушения информационной безопасности надзорной деятельности территориально распределенного объекта повышенной опасности мегаполиса посредством методов и моделей анализа угроз и их кластеризации с учётом возможных рисков на основе междисциплинарных моделей, которые объединяют методические подходы MITRE ATT&CK, IEC 62443, HAZOP/FMEA и показатели характеристик существенных свойств цифровых двойников инфраструктурных компонент реальных объектов, с динамической оценкой сценариев, включающих влияние киберугроз на возникновение физических аварий и пожароопасных чрезвычайных ситуаций.**

Научная новизна результатов исследования заключается в применении при декомпозиции ситуаций территориально распределенного опасного производственного объекта алгоритма иерархической кластеризации Ланса-Вильямса и в учете рисков информационной, функциональной и пожарной безопасности через параметры изменения метрики ребер графа ситуаций.

Проведенное моделирование критических ситуаций с различным соотношением характеристик при изменении объема перечня угроз на порядок при воздействии на различные ресурсы информационной инфраструктуры объекта показало возможность обеспечения стабильной погрешности за счет динамических процедур выбора и применения метода и алгоритма кластеризации.

Представленные результаты предлагается реализовать в виде математического обеспечения систем поддержки принятия решений специалистов – администраторов информационной безопасности системы обеспечения информационной безопасности опасных производственных объектов мегаполиса.

***Ключевые слова* — алгоритмы, кластеризация, риски, информационная безопасность, функциональная безопасность, пожарная безопасность, опасный производственный объект.**

М.А. Тукмачева – адъюнкт ФПКВК ФГБОУ ВО «Санкт-Петербургский университет ГПС МЧС России» (email: mtukmacheva@mail.ru).

А.В. Шестаков – д.т.н., ведущий научный сотрудник ФГБОУ ВО «Санкт-Петербургский университет ГПС МЧС России» (email: alexandr.shestakov01@yandex.ru).

К.З. Билятдинов – д.т.н., профессор ФГАОУ ВО «Санкт-Петербургский государственный университет аэрокосмического приборостроения» (email: k74b@mail.ru).

I. Введение

Глобальное внедрение цифровых технологий в экономику, управление и социальные процессы, характеризуют четвертую промышленную революцию, именуемую как «Индустрия 4.0» или «Цифровая трансформация 4.0».

Технологии «Индустрии 4.0» существенно влияют на глобальные производственные площадки и организационные структуры. Города и мегаполисы, в концепции «Индустрии 4.0», являются центрами экономики, промышленности, рабочей силы и информации. Традиционное разграничение между производственными зонами и жилыми районами, как показывают обзорные исследования (например, Fang C., Yu D. [1, 2017], Risdianto E., Cs M. [2, 2019], Maulidi C., Adiwan F.A, Dwicaksono A., Winarso H. [3, 2024]), стирается, территория мегаполиса используется многофункционально, а в результате структурного сжатия - формируется многоуровневая городская производственная инфраструктура, что отражается на системе рисков для опасных производственных объектов (ОПО).

В мегаполисе появляется информационная инфраструктура технологических процессов, которая усложняет техногенную безопасность городов. В автоматизированных системах управления технологическими процессами (АСУ ТП) ОПО возрастают причинно-следственные связи рисков пожарной, функциональной и информационной безопасности. При рассмотрении возможного потенциального ущерба мегаполису и упрощения последующей статистической обработки данных, а также для группирования наиболее критических ситуаций в системах обеспечения информационной безопасности объектов (СОИБ), целесообразно применять методы кластерного анализа.

Общие «узкие места» современных решений заключаются в отсутствии единого интегрированного фреймворка «ИБ + функциональная безопасность + пожарная безопасность», а также ограниченном взаимодействии специалистов по киберзащите и инженеров по промышленной безопасности ОПО.

Таким образом, объективное развитие технологий и цифровая трансформация территориально-распределенных производственных процессов ОПО и мегаполисов меняют традиционное подходы к пониманию и организации построения их системы

рисков, что определяет несомненную актуальность затронутой проблематики исследования.

II. Постановка задачи исследования

Постановка задачи кластеризации рисков трёх типов (информационной, функциональной и пожарной безопасности) для информационной инфраструктуры и технологических процессов опасного производственного объекта территориально-распределенного в различных районах мегаполиса с применением различных методов нормализации, различных методов кластерного анализа, с ограничениями и допущениями, проверкой качества кластеризации и выбором последовательности реагирования на киберинциденты представляется исходными данными, целевой функцией, системой ограничений и допущений.

Обозначения и исходные данные

Множество рисков S представлено в виде совокупности n-оценённых рисковых S_n событий (факторов)

$$S = \{S_1, S_2, \dots, S_n\}. \quad (1)$$

Каждое i-рисковое событие S_i задано как вектор $f(S_i)$ исходных признаков

$$f(S_i) = \{f_{1i}, \dots, f_{di}\} \in R^d. \quad (2)$$

где d – число количественных и/или категориальных метрик (действие злоумышленника, оценка уязвимости, оценка ущерба, зона ответственности, технологический участок и т. п.).

Указан y_i тип риска: для каждого S_i известно доменное поле $y_i \in \{1,2,3\}$, где $1, 2, 3$ – информационная, функциональная и пожарная безопасность соответственно.

Каждый i-риск информационной безопасности (R_{InfSec}), функциональной безопасности (R_{FunSaf}), пожарной безопасности ($R_{FireSaf}$) представлен в виде совокупности рисков соответственно

$$R_{InfSec} = \{R_{InfSec1}, \dots, R_{InfSeci}, \dots, R_{InfSecN}\}; \quad (3)$$
$$i=\overline{1, N}.$$

$$R_{FunSaf} = \{R_{FunSaf1}, \dots, R_{FunSafi}, \dots, R_{FunSafN}\}; \quad (4)$$
$$i=\overline{1, N}.$$

$$R_{FireSaf} = \{R_{FireSaf1}, \dots, R_{FireSafi}, \dots, R_{FireSafN}\}; \ i = \overline{1,N}. \quad (5)$$
$$i=\overline{1, N}.$$

Задан географический признак δ_i как район мегаполиса, в котором находится опасный производственный объект (объект защиты, технологический участок)

$$\delta_i \in D = \{d_1, d_2, \dots, d_M\}. \quad (6)$$

Определены методы нормализации

Пусть набор f→f операторов нормализации представлен множеством M_N методов нормализации как

$$M_N = \{N_1, N_2, \dots, N_P\}. \quad (7)$$

где N_1 – минимально-максимальная нормировка представлена в виде

$$\hat{f}_j = (f_j - \min k\ f_{jk})/(\max k\ f_{jk} - \min k\ f_{jk}), \quad (8)$$
$$\text{при } j=\overline{1, N},\ j \in N;$$

N_2 – z-оценка (стандартизация) представлена в виде

$$\hat{f}_j = (f_j - \mu_j) / \sigma_j; \quad (9)$$

N_3 – робастная нормировка (медиана / IQR - Inter-Quartile Range, межквартильный размах) представлена в виде

$$\hat{f}_j = (f_j - med_j) / (Q_{3j} - Q_{1j}); \quad (10)$$

N_P – другие доменные преобразования (логарифм, ранговая шкала и т.п.).

Определены методы (алгоритмы) кластеризации

Пусть набор M_C методов (алгоритмов) кластеризации задан как

$$M_C = \{A_1, A_2, \dots, A_Q\}. \quad (11)$$

где A_1 – k-means (с жёстким ограничением на k);

A_2 – иерархическая агломеративная (complete/linkage);

A_3 – DBSCAN (плотностной);

A_4 – COP-k-means (constrained k-means с ограничениями по размерам и меткам);

A_Q – иные методы (спектральная кластеризация, GMM и т.п.).

Постановка задачи для каждого сочетания нормализации и алгоритма

Для каждого $N_p \in M_N$ и $A_q \in M_C$ выполнить следующие действия:

а) получить нормализованные векторы

$$\hat{X}_p = \{\hat{x}_i = N_p(f(x_i))\} \in \mathbb{R}^d, \quad (12)$$

б) задать параметры алгоритма A_q, в частности:

число кластеров k (если применимо);

пороги ε, min P_{ts} для DBSCAN;

весовые коэффициенты для метрик и т.п.

в) сформировать разбиение (для DBSCAN – с шумом $C_0\{p,q\}$)

$$C\{p,q\} = \{C_1\{p,q\}, \dots, Ck\{p,q\}\}. \quad (13)$$

Целевая функция и критерии качества

Внутрикластерная дисперсия (для алгоритмов на основе центра):

$$J(C) = \sum\{p=1, .., k\} \sum\{\hat{x}_i \in C_p\} D(\hat{x}_i, \mu_p), \quad (14)$$

где μ_p = средний/медианный вектор кластера C_p.

Межкластерное расстояние (разделимость):

$$S(C) = \min\{p \neq q\}\ D(\mu_p, \mu_q). \quad (15)$$

Внешние M_{QJ} метрики качества

$$M_Q = \{Q_1, Q_2, \dots, Q_I\}. \quad (16)$$

Можно представить (без учёта меток y, δ):

коэффициент Силуэта (Silhouette coefficient S_i) для каждого $\hat{x}_i$ и среднее Sil(C);

индекс Дэвиса–Булдина DB(C) (Davies–Bouldin index DB(C));

индекс Данна (Dunn index Dn(C)).

Приоритизация кластеров по «уровню риска»

для каждого кластера p вычисляем

$$R_p = w_1 \cdot \sum\{\hat{x}_i \in C_p\} \text{Prob}(x_i) + \quad (17)$$
$$+ w_2 \cdot \sum_{(i)} \text{Sever}(x_i) + w_3 \cdot |C_p|,$$

где Prob и Sever – исходные признаки вероятности и тяжести, w_k – веса.

Ограничения

При формировании каждого C{p, q} должны быть соблюдены:

размер кластера

$$L \le |C_p\{p,q\}| \le U, \quad (18)$$

доля одного вида риска в кластере $\forall j \in \{1, 2, 3\}$:

$$|\{x_i \in C_p : y_i = j\}| \le \alpha_j |C_p|, \quad (19)$$

географическое покрытие:

либо кластер «локален» (δ_i одинаково для всех $x_i \in C_p$),

либо не более β разных районов в одном кластере:

$$|\{\delta_i : x_i \in C_p\}| \le \beta, \quad (20)$$

максимальный внутрикластерный диаметр

$$\max\{x_i, x_j \in C_p\}\, D(\hat{x}_i, \hat{x}_j) \le \tau_{max}, \quad (21)$$

ограничение времени

$$T(N, d, k\{p, q\}, I\{p, q\}) \le Tmax, \quad (22)$$

где I{p, q} – число итераций/итерационных проходов алгоритма A_q.

Допущения

Метрики D и алгоритмы нормализации N_p корректно отражают «сходство» рисков.

Признаковые векторы $f(x_i)$ полны или упрощённо восстановлены до нормализации.

Доменные метки y_i и географические δ_i заданы экспертно и не изменяются.

Распределение рисков статично в анализируемый период. Параметры α_j, β, τ_{max}, L, U, T_{max}, w_k отражают реальные бизнес-ограничения и стратегию опасного производственного объекта.

III. Методы и модели кластеризации угроз (рисков) нарушения информационной безопасности опасных производственных объектов мегаполиса

Для кластеризации угроз нарушения информационной безопасности опасных производственных объектов мегаполиса используются методы нормализации и методы кластерного анализа. Методы позволяют выявить кластеры угроз, сходных между собой по определенным свойствам, и оценить последствия их реализации. Кластеризация угроз необходима для управления рисками, их минимизации, а также для разработки мер по защите информации и иных профилактических мероприятий.

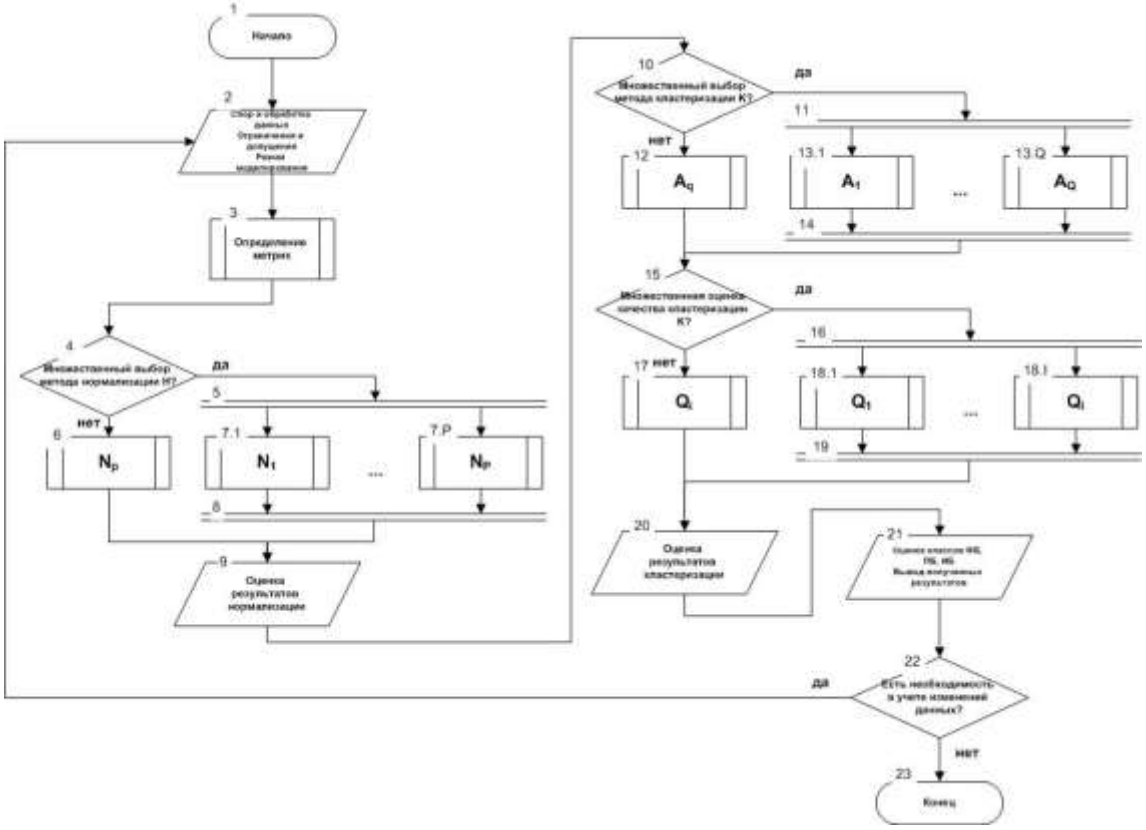

Рис. 1 — Обобщенная схема выявления (идентификации) угроз и кластеризации рисков информационной, функциональной и пожарной безопасности опасных производственных объектов

Схема выявления и идентификации угроз нарушения информационной безопасности, а также кластеризации рисков информационной, функциональной и пожарной безопасности опасных производственных объектов представлена на рисунке 1. Основными этапами являются: ввод данных; нормализация; кластеризация;

качество моделирования; результаты выявления и идентификации угроз и кластеризации рисков. В рамках пяти этапов осуществляется 23 последовательных шага, представленных на рисунке.

Этап 1: Ввод данных. Осуществляется сбор и обработка данных ручным или автоматизированным способом, определяются требования к данным в соответствии с постановкой задачи. Выполняются шаги 1-3. Формируются ограничения и допущения для входных данных, с учетом формул (18) – (22). Актуализация изменений данных (при необходимости).

Учет особенностей данных по угрозам (рискам) информационной безопасности.

В аналитическом отчете «Лаборатории Касперского» показано, что в 6 из 13 регионов мира, показатели изменений доли компьютеров АСУ ТП, на которых были заблокированы вредоносные объекты, в I квартале 2025 года выросли по сравнению с предыдущим кварталом, причём наибольшее изменение произошло в России (рисунок 3).

Рис. 2 – Изменения доли компьютеров АСУ ТП, на которых были заблокированы вредоносные объекты, в I квартале 2025 года по сравнению с предыдущим кварталом

В системах промышленной автоматизации решения «Лаборатории Касперского» для обеспечения безопасности заблокировали в первом квартале 2025 года вредоносное программное обеспечение (ПО) из 11 679 различных семейств вредоносных программ в различных категориях (рисунок 4).

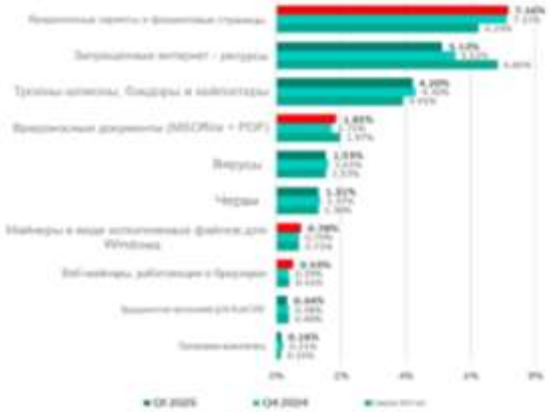

Рис. 3 – Доля компьютеров АСУ, на которых была заблокирована активность вредоносных объектов различных категорий

Специалистами «Лаборатории Касперского» на основе анализа статистических данных указывается, что основной категорией вредоносных программ, используемых для первоначального заражения компьютеров АСУ ТП (ICS), являются вредоносные скрипты и фишинговые страницы. Большинство вредоносных скриптов и фишинговых страниц служат для распространения или загрузки вредоносного ПО следующего этапа (шпионских программ, криптомайнеров и программ-вымогателей).

Корреляция между значениями для вредоносных скриптов и фишинговых страниц, а также для шпионских программ представлена в явном виде на рисунке 5.

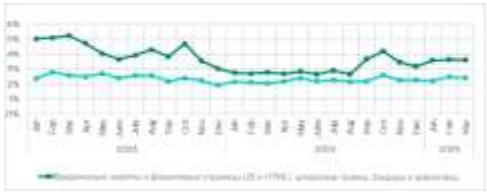

Рис. 5 – Доля компьютеров ICS, на которых были заблокированы вредоносные объекты (январь 2023 г. - март 2025 г.)

Как и в случае с вредоносными скриптами и фишинговыми страницами, в первые три месяца 2025 года процент компьютеров в сфере промышленного управления и контроля (ICS), на которых было заблокировано шпионское ПО, был выше, чем в те же месяцы 2024 года, как показано на рисунке 6.

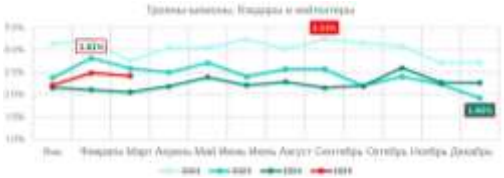

Рис. 6 – Доля компьютеров ICS, на которых было заблокировано шпионское ПО (январь 2022 г. - март 2025 г.)

Статистические данные по угрозам компьютерных атак на информационные инфраструктуры (сетевых атак, программ-вымогателей, эксплойтов, веб-угроз, сканирование по требованию, спама в электронной почте, вредоносной почте, локальных заражений) различных стран мира, обнаруженным на устройствах пользователей продуктов «Лаборатории Касперского», представлены на рисунках 7-14 и в таблицах 2-9 за период 28.07.2025-28.08.2025 в Российской Федерации.

Рис. 7 – Распределение угроз (сетевые атаки) на информационную инфраструктуру РФ за месяц

Таблица 2

Топ-10 обнаруженных угроз (сетевых атак) за месяц

№№ п/п	Перечень обнаруженных угроз	Доля обнаруженных угроз, %
1	Bruteforce.Generic.Rdp.a	50.07
2	DoS.Generic.Flood.TCPSYN	19.8
3	Bruteforce.Generic.Rdp.d	10.8
4	Intrusion.Win.MS17-010.o	7.4
5	Scan.Generic.PortScan.TCP	6.6
6	Scan.Generic.PortScan.UDP	3.2
7	DoS.Generic.Flood.ICMP	0.96
8	Intrusion.Win.MS17-010.p	0.21
9	Intrusion.Generic.CVE-2021-44228.a	0.16
10	Bruteforce.Generic.Rdp.c	0.12

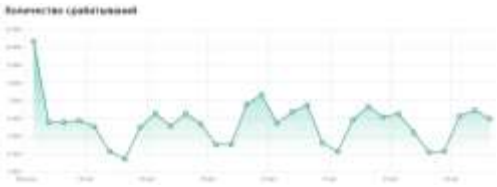

Рис. 8 – Распределение угроз (программы-вымогатели) на информационную инфраструктуру РФ за месяц

Таблица 3

Топ-10 обнаруженных угроз (программ-вымогателей) за месяц

№№ п/п	Перечень обнаруженных угроз	Доля обнаруженных угроз, %
1	Trojan-Ransom.Win32.Blocker.ckeq	14.53
2	Trojan-Ransom.Win32.Mor.bjs	8.31
3	Trojan-Ransom.Win32.Gen.p	5.70
4	Trojan-Ransom.AndroidOS.Rkor.pac	5.10
5	Trojan-Ransom.Win32.Crypmodng.gen	4.58
6	Trojan-Ransom.Win32.Wanna.m	3.73
7	Trojan-Ransom.Win32.Crypren.gen	3.16
8	Trojan-Ransom.Win32.Blocker.jxbh	2.98
9	Trojan-Ransom.Win32.Wanna.ar	2.91
10	Trojan-Ransom.Win32.Dcryptor.b	2.87

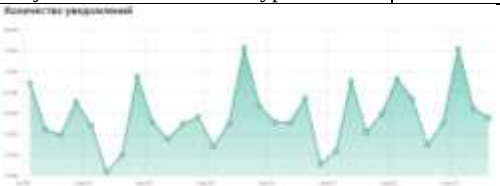

Рис. 9 – Распределение угроз (эксплойтов) на информационную инфраструктуру РФ, за месяц

Таблица 4

Топ-10 обнаруженных угроз (эксплойтов), за месяц

№№ п/п	Перечень обнаруженных угроз	Доля обнаруженных угроз, %
1	Эксплойт. Win32. CVE-2010-2862.a	17,36
2	Эксплойт.MSOffice.Generic.a	11,68
3	Эксплойт. HTTP. CVE-2017-5638.gen	8,80
4	Эксплойт.MSOffice.CVE-2017-11882.gen	8,10
5	Эксплойт. Win32.MS05-036	5,09
6	Эксплойт. Скрипт. Общий	4,34
7	Эксплойт.MSOffice.CVE-2018-0802.gen	4,10
8	VHO: Эксплойт. МИЛФ. Convagent.gen	3,99
9	Эксплойт. Win32. CVE-2017-11882.gen	2,51
10	Эксплойт.MSOffice.CVE-2017-11882.j	2,24

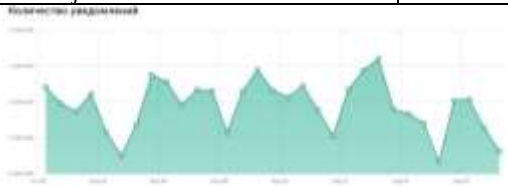

Рис. 10 – Распределение угроз (веб-угроз) на информационную инфраструктуру РФ, за месяц

Таблица 5

Топ-10 обнаруженных угроз (веб-угроз), за месяц

№№ п/п	Перечень обнаруженных угроз	Доля обнаруженных угроз, %
1	Троян.BAT.Майнер.ген	53,08
2	Троян.Скрипт.Универсальный	8,82
3	Троян-загрузчик. PDF. Агент. ген	8,31
4	Троян.Скрипт.Агент.ген	6,50
5	Троян.PDF.Бадур.ген	5,31
6	Троян.Мульти.Preqw.gen	3,17
7	Hoax.PDF.Phish.au	1,43
8	Троян-PSW.Script.Generic	0,98
9	Hoax.HTML.Фишинг.ген	0,95
10	Троян.Java.SAgent.gen	0,80

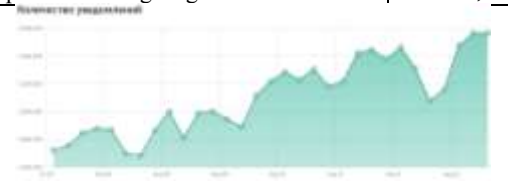

Рис. 11 – Распределение угроз (сканирование по требованию) на информационную инфраструктуру РФ, за месяц

Таблица 6

Топ-10 обнаруженных угроз (сканирование по требованию), за месяц

№№ п/п	Перечень обнаруженных угроз	Доля обнаруженных угроз, %
1	Троян.AndroidOS.Фейковые деньги.v	19,64
2	Троян.AndroidOS.Фейковые деньги.eb	3,80
3	Backdoor.AndroidOS.Triada.z	3,74
4	Троян.AndroidOS.Triada.fe	3.51
5	Троян-банкир.AndroidOS.Creduz.z	2,60
6	Троян.AndroidOS.Triada.gn	2,49
7	Троян.Мульти.Агент.ген	2,05
8	Троян-загрузчик.AndroidOS.Dwphon.a	1,83
9	Троян.AndroidOS.Фальшивые деньги.	1,60
10	Троян.AndroidOS.Фейковые деньги.dz	1,57

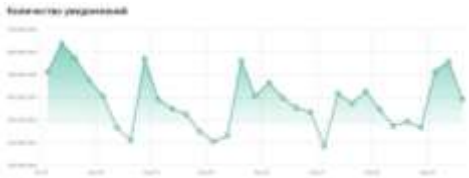

Рис. 12 – Распределение угроз (спам в электронной почте) на информационную инфраструктуру РФ, за месяц

Таблица 7

Топ-6 обнаруженных угроз (спам в электронной почте), за месяц

№№ п/п	Перечень обнаруженных угроз	Доля обнаруженных угроз, %
1	Анализ формальных атрибутов	19,88
2	Автоматизированный анализ Формальных атрибутов	19,10
3	Лингвистический анализ	3,10
4	Анализ атрибутов Отправителя	0,13
5	Анализ сигнатур	0,02
6	Служба принудительного обновления защиты от спама	0,02

Рис. 13 – Распределение угроз (вредоносная почта) на информационную инфраструктуру РФ, за месяц

Таблица 8

Топ-10 обнаруженных угроз (вредоносная почта), за месяц

№№ п/п	Перечень обнаруженных угроз	Доля обнаруженных угроз, %
1	Hoax.Script.Scaremail.gen	15,24
2	Опасный объект. Мульти. Универсальный	13,13
3	Троян.Win32.Badun.gen	10,48
4	Троян.Скрипт.Универсальный	7,93
5	Троян.MSIL.Taskun.gen	5,92
6	Троян-шпион.MSIL.Noon.gen	4,59
7	Троян.WinLNK.Агент.gen	1,67
8	Троян-PSW.MSIL.Agensla.gen	1,55
9	Hoax.HTML.Фишинг.ген	1,37
10	Троян. Win32.Makoob.gen	1,32

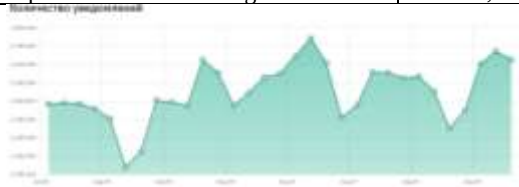

Рис. 14 – Распределение угроз (локальные заражения) на информационную инфраструктуру РФ, за месяц

Таблица 9

Топ-10 обнаруженных угроз (локальные заражения), за месяц

№№ п/п	Перечень обнаруженных угроз	Доля обнаруженных угроз, %
1	Опасный объект. Мульти. Универсальный	11,10
2	Троян.AndroidOS.Triada.ii	4,15
3	Троян.AndroidOS.Fakemoney.dt	4,15
4	Троян.AndroidOS.Triada.fe	3,88
5	Троян.AndroidOS.Фейковые деньги.v	3,48
6	Троян-банкир.AndroidOS.Creduz.z	3,39
7	Троянец-дроппер.AndroidOS.Hqwar.cq	3,12
8	Троян-банкир.AndroidOS.Mamont.hi	2,81
9	Троян. Win32.Hosts2.gen	2,79
10	Троян-шпион. Win32.Agent.gen	2,77

Эффективность мер защиты информации и процессов в количественных показателях представляют метрики в информационной безопасности. Метрики связаны с различными аспектами информационной безопасности, такими как: управление уязвимостями, работа с инцидентами, повышение осведомленности сотрудников, соответствие требованиям регуляторов и др.

Определение метрик может происходить автоматизированным способом из систем защиты информации (антивирусы, межсетевые экраны) или способом сбора данных из технической инфраструктуры, методом экспертных оценок или из аналитических отчетов (компаний по защите информации (Kaspersky) или соответствующих регуляторов (ФСТЭК России)).

Определить какие метрики коррелируют с рисками информационной, функциональной и пожарной безопасностью на опасном производственном объекте в мегаполисе. Данные о метриках характеризуют численные значение свойств, необходимых для дальнейших процедур кластеризации.

Этап 2: Нормализация. Цель этапа – приведение величин разных масштабов и единиц измерения к единообразному виду. Концепция нормализации впервые была предложена Эдгаром Коддом (Edgar Frank Codd, 1970), он разработал теорию реляционных баз данных и предложил теорию нормализации. Выполняют шаги 4-9. Определяются допустимые методы нормализации. Предусматривается возможность применения одного из методов или параллельная обработка данных при нескольких методах. Выбор метода зависит от постановки задачи: минимально-максимальная нормализация сохраняет диапазон; z-оценка центрирует и нормирует дисперсию; робастная нормировка устойчива к выбросам и т.д.

Например, N_1 – минимально-максимальная нормализация; N_2 – z-оценка (стандартизация); N_3 – робастная нормировка; N_4 – нормализация на основе среднего значения; N_5 – десятичное масштабирование; N_6 – квантили и т.д.

При отсутствии множественного выбора нормализации нормирование данных образуется по одному методу, при множественном выборе – расчет параллельный.

Обоснование применимости в обобщенной методике кластеризации различных методов нормализации представляется следующим.

Минимально-максимальная нормализация преобразовывает все значения в диапазон от 0 до 1. Метод не имеет единого авторства, т.к. представляет собой общепринятую статистическую технику [18]

$$S = \{s_1, \ldots, s_i, \ldots, s_n\}, \quad (23)$$

$$min(j) = min\{s_{ij} | i = 1 \dots n\}, \quad (24)$$

$$max(j) = max\{s_{ij} | i = 1 \dots n\}, \quad (25)$$

$$\hat{s}_i = (s_i - min(j))/(max(j) - min(j)), \quad (26)$$

где s_i – исходное значение множества S;

$\hat{s}_i$ – нормализованное значение;

$min(j), max(j)$ – минимальное и максимальное значение множества S соответственно.

Процесс преобразования включает следующие процедуры:

1. Для множества S вычисляется минимальное $min(j)$ и максимальное $max(j)$ значение с помощью ранжирования и упорядочивания чисел.
2. Каждое исходное значение множества S преобразовывается в значение $\hat{s}_i$ по формуле нормализации (26).

Результатом нормализации определяется выборка $\hat{S}$, в которой минимальное значение равно 0, а максимальное значение 1.

При нормализации значений в диапазоне [a, b], формула имеет следующий вид

$$\hat{s}_i = (s_i - min(j))/(max(j) - min(j)) \times (b - a) + a. \quad (27)$$

Ограничения метода: чувствителен к экстремумам (выбросам), нестабилен в динамике данных, требование максимального и минимального значения не равных друг другу.

Z-оценка (стандартизация) нормализует данные на основе среднего значения набора данных (μ) и стандартного отклонения (σ) данных. Метод сформировался в результате развития теории нормального распределения и статистических мер рассеяния, с вкладом К.Ф. Гаусса, К. Пирсона и Р.А. Фишера [18]

$$\hat{s}_i = (s_i - \mu)/\sigma, \quad (28)$$

где μ – среднее значение набора данных,

σ – стандартное отклонение набора данных.

Среднее значение набора данных (μ) имеет следующий вид

$$\mu = \frac{\sum_i s_i}{n} | \ i = 1 \dots n, \quad (29)$$

где s_i – i-е значение признака;

n – число наблюдений в выборке.

Стандартное отклонение (σ) имеет вид

$$\sigma = \sqrt{\frac{\sum_i (s_i - \mu)^2}{n-1}}., \quad (30)$$

Процесс преобразования включает следующие процедуры:

1. Вычисление среднего значения набора данных по формуле (29) и стандартного отклонения по формуле (30).
2. Каждое исходное значение множества S преобразовывается в значение $\hat{s}_i$ по формуле нормализации (28), т.е. из каждого исходного значения вычитают среднее значение набора данных (μ) и делят результат на стандартное отклонение (σ).

Ограничения метода: чувствителен к экстремумам (выбросам), нестабильность при малых выборках, требование стандартного отклонения не равное нулю.

Робастная нормировка (медиана/IQR) применяется для выборок логнормального типа, среднее значение заменяют на медиану, а среднеквадратичное отклонение — на половину расстояния между верхним и нижним децилями. В развитие робастной статистики внесли вклад Дж. Бокс (George E. P. Box, 1953), Дж. П. Хьюбер (Peter J. Huber, 1964), Ф. Хампель (Frank Hampel, 1968), Р.А. Фишер и Б.М. Болес (Ronald Aylmer Fisher, Robert C. Bolles, 1981) и др. Формальное представление нормировки имеет следующий вид [19]

$$\hat{s}_i = (s_i - M(s_i))/IQR, \quad (31)$$

где $M(s_i)$ – медиана множества S = Q_2;

IQR – межквартильный размах.

Для нахождения медианы множества S необходимо отсортировать выборку в порядке возрастания

$$S = \{s_{1i} \leq \dots \leq s_{ii} \leq \dots \leq s_{ni}\}, \quad (32)$$

где n – число наблюдений в выборке.

Если n – нечетное, n=2m+1, то медиана $M(s_i) = s_i$ (m+1).

Если n – четное, n=2m, то медиана $M(s_i) = s_i$ (m) + s_i (m+1)/2.

Расчет IQR (межквартильного размаха) выполняется в следующие процедуры:

1. Находится первый квартиль (Q_1) — медиана нижней половины данных $M(s_{1i})$

$$M(s_{1i}) = Q_1. \quad (33)$$

2. Находится третий квартиль (Q_3) — медиана верхней половины данных $M(s_{3i})$

$$M(s_{3i}) = Q_3. \quad (34)$$

3. Из третьего квартиля вычитается первый и рассчитывается IQR

$$\text{IQR} = Q_3 - Q_1. \quad (35)$$

Преимущество IQR заключается в устойчивости к выбросам и экстремальным значениям, отсутствует зависимость от потенциально аномальных крайних точек данных.

После преобразования данных в единый формат оцениваются результаты, их сопоставимость, и оценка удобства дальнейшего анализа. Оценка результатов нормализации возможна с помощью таких методов как: анализ матрицы решений; сравнение результатов классификации и кластеризации; анализ эффективности метода нормализации и др.

Этап 3: Кластеризация. Определяются методы кластеризации и предусматривается возможность применения одного из методов или параллельная обработка данных при нескольких методах. Выполняют шаги 10-14. Процесс группировки объектов на кластеры происходит таким образом, чтобы внутри одного кластера объекты были схожи по определенным признакам, а между кластерами (группами) различны. Существует ряд методов кластеризации, например: A_1 –

метод k-средних (k-means); A_2 – иерархическая агломеративная; A_3 – DBSCAN (плотностной); A_4 – COP-k-means; и др.

Метод k-средних (K-means) минимизирует суммарное квадратичное отклонение точек кластеров k от центров μ_i этих кластеров. Метод предложен Г. Штейнгаузом и С. Ллойдом (1950) [20]

$$V = \sum_{i=1}^{k} \sum_{x_j \in S_i} (x_{ij} - \mu_i)^2, \quad \text{при } i = 1, \ldots, k; j = 1, \ldots, n_j, \tag{37}$$

где V – функция, которую необходимо минимизировать;

k – количество кластеров;

n_j – количество точек в кластере;

x_{ij} – i-я точка в кластере j;

S_i – образовавшиеся кластеры;

μ_i – центры масс векторов $x_j \in S_i$, центроид кластера j.

В двумерном пространстве расстояние между двумя точками можно вычислить в следующем виде:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}. \tag{38}$$

В многомерном пространстве евклидово расстояние имеет вид:

$$d(x, y) = \sqrt{\sum (x_i - y_i)^2}. \tag{39}$$

Метод k-средних осуществляется с повторяющими циклами – итерациями, включая две основные процедуры: назначения и обновления. Процедура назначения заключается в назначении точек к ближайшему центроиду. Процедура обновления заключается в пересчете центроида как среднего арифметического точек в кластере

$$(x', y') = \left(\frac{x_1 + \ldots + x_n}{n}, \frac{y_1 + \ldots + y_n}{n}\right), \tag{40}$$

где n – количество точек в кластере;

x, y – координаты точки.

На каждой итерации выбирается центр масс для каждого кластера – выбор случайно выбранного значения в области кластера. Затем данные разбиваются на кластеры в соответствии с новыми центрами, ближайших по выбранной метрике.

Завершается метод при отсутствии изменений центра масс кластеров на какой-либо итерации.

Иерархический метод представляет собой дендрограмму – дерево, построенное по определенному алгоритму. При определении одного элемента в одном кластере и в другом, расстояние определяется между объектами:

$$R(\{x\}, \{y\}) = p(x, y). \tag{41}$$

При определении нескольких элементов в одном кластере и несколько в другом, расстояние между кластерами U и V определяется с помощью функций в зависимости от специфики задачи.

Метод одиночной связи заключается в определении расстояния между кластерами как между двумя наиболее близкими объектами из различных кластеров:

$$R_{min}(U, V) = \min_{u \in U, v \in V} p(u, v). \tag{42}$$

Метод полной связи заключается в определении расстояния между двумя дальними элементами из различных кластеров:

$$R_{min}(U, V) = \max_{u \in U, v \in V} p(u, v). \tag{43}$$

Метод средней связи представляет собой анализ расстояния между кластерами как среднее расстояние между элементами из разных кластеров:

$$R_{avg}(U, V) = \frac{1}{|U| * |V|} \sum_{u \in U} \sum_{v \in V} \mathrm{p(u, v)}. \tag{44}$$

Центроидный метод представляет собой анализ расстояния между двумя кластерами как расстояние между центрами этих кластеров (средних)

$$R_c(U, V) = p^2\left(\sum_{u \in U} \frac{u}{|U|}, \sum_{v \in V} \frac{v}{|V|}\right). \tag{45}$$

Метод Уорда направлен на объединение ближайших по расположению кластеров и создает кластеры малого размера.

$$R_{ward}(U, V) = \frac{|U| * |V|}{|U| + |V|} p^2\left(\sum_{u \in U} \frac{u}{|U|}, \sum_{v \in V} \frac{v}{|V|}\right). \tag{46}$$

Метод DBSCAN основан на плотности, группирует данные, которые тесно расположены и помечает как выбросы данные из областей с малой плотностью. Метод применим к нерегулярным и сложным распределениям данных в пространстве. Разработали метод М. Эстер, Г.-П. Кригель, Й. Сандер и С. Сяовэй (Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xu Xiaowei, 1996).

Формально кластер в DBSCAN определяется как максимальная плотность множества связанных точек, к шуму относят точки, не вошедшие ни в один кластер.

Метод не может хорошо кластеризовать наборы данных с большой разницей в плотности, не полностью однозначен.

Ключевые параметры DBSCAN:

ε – радиус окрестности, определяющий ближайших друг к другу данных (точек);

minPts – минимальное количество точек в ε-окрестности для признания точки ключевой.

Принцип метода:

1. Произвольную точку проверяют является ли она ключевой (если не ключевая – относят к шумовой, если ключевая – создают новый кластер с этой точкой).

2. Соответствующим образом добавляют в кластер точки, находящиеся в пределах расстояния ε от любой точки кластера.

3. Обработка следующих данных.

С помощью методов кластеризации данных определяется структура модели угроз (рисков) нарушения информационной безопасности на опасном производственном объекте мегаполиса. Показатели рисков нарушения информационной безопасности можно сгруппировать таким образом, что их угрозы внутри одного кластера будут максимально схожи друг с другом, а из разных групп (кластеров) максимально различны. Таким образом будет выявлен кластер с максимальными рисками нарушения информационной безопасности, ведущие за собой риски в сфере функциональной и пожарной безопасности, что может привести к авариям и чрезвычайным ситуациям в мегаполисе.

Этап 4: Качество моделирования. Качество применения методов кластеризации к данным необходимо оценить посредством оценки успешности применения метода кластеризации с помощью

дополнительных метрик (14) – (16). Внешние метрики качества представляются как: Q_1 – коэффициент Силуэта; Q_2 – индекс Дэвиса–Булдина; Q_3 – индекс Данна.

Коэффициент Силуэта показывает сходство объектов внутри кластера по сравнению с объектами из других кластеров. Метод предложил бельгийский статистик Питер Руссо (Peter Rousseeuw, 1987). Значение коэффициента Силуэта варьируется от -1 до 1 (близкое к 1 – объект хорошо согласуется внутри кластера; близкое к 0 – объект находится на границе двух кластеров; близкое к -1 – объект скорее всего отнесен не к верному кластеру).

$$S(x) = \frac{b(x) - a(x)}{max(a(x), b(x))}, \quad (47)$$

где *a(x)* — среднее расстояние от точки *x* до других точек в этом кластере *Ci*;

b(x) — минимальное среднее расстояние от точки *x* до точек в любом другом кластере.

Индекс Дэвиса–Булдина измеряет суммарное сходство кластеров и их компактность, оценка кластеров через анализ среднего сходства между каждым кластером и наиболее похожим на него кластером – отношение суммы разброса внутри кластера к расстоянию между кластерами (чем ниже значение DBI, тем лучше кластеризация, более высокие значения DBI соответствуют худшим решениям кластеризации)

$$DB = \frac{1}{k}\sum_{i=1}^{k} \max_{i \neq j} R_{ij}. \quad (48)$$

Индекс Данна сравнивает межкластерное расстояние с диаметром кластера, определяется соотношением между минимальным расстоянием пар точек из разных кластеров и максимальным расстоянием пар точек из одного кластера (при большем значении индекса лучше кластеризация)

$$DI = \frac{min(\delta(Ci, Cj))}{max(\Delta(Ck))}, \quad (49)$$

где: δ (Ci, Cj) — межкластерное расстояние между кластерами Ci и Cj;

Δ(Ck) — внутрикластерный диаметр кластера Ck;

min (δ(Ci, Cj)) — наименьшее расстояние между любыми двумя кластерами;

max(Δ(Ck)) — наибольший диаметр в любом кластере.

Качество моделирования в моделях угроз и рисков нарушения информационной безопасности необходимо для системного выявления и анализа факторов, способных нанести вред информационной системе и повлечь за собой возможные неблагоприятные последствия.

Оценка качества моделирования позволяет упорядочить и формализовать информацию о возможных негативных сценариях; приоритизировать меры защиты; оценить риски и принять соответствующие решения, а также разработать эффективные планы реагирования и провести превентивные мероприятия, что поможет заранее подготовиться к возможным инцидентам, сократить время реагирования и минимизировать ущерб.

Этап 5: Результаты выявления и идентификации угроз.

Результатами выявления и идентификации угроз (рисков) является актуализация угроз и рисков, их источников, способов, возможностей и сценариев реализации.

Они могут привести к нарушению безопасности обрабатываемой информации (конфиденциальности, целостности, доступности) и/или к нарушению или прекращению функционирования систем и сетей.

Результаты оценки угроз и рисков обосновывают выбор организационных и технических мер по защите информации, а также средств защиты и их функциональных возможностей.

Контрольный пример

Апробация предложенной методики проведена на наборе условных данных.

1. Исходные данные представлены в перечне, в таблице 10 и 11:

- число объектов n=8;
- двумерные векторы $f(s_i)$= ($Prob_i$, $Sever_i$);
- метка типа $y_i \in$ {1=инф., 2=функц., 3=пож.};
- гео-метка $\delta_i \in$ {A, B, C}.
- число объектов (узлов ИТ-инфраструктуры + технологических площадок) n=8

$$O=\{o_1,\ldots,o_n\};$$

- для каждого o_i вектор рисков y_i = ($R_{InfSec\,i}$, $R_{FunSaf\,i}$, $R_{FunSaf\,i}$) $\in \mathbb{R}^3$;
- целевая задача: разбить O на k кластеров C = {C_1, …, C_k}, чтобы внутри кластеров суммарный «разнобой» рисков был минимален, а между кластерами — максимален.

Формализуем минимизацию внутрикластерной дисперсии:

$$\min \sum_{j=1..k} \sum_{o \in C_j} d(x_o, \mu_j)^2,$$

где μ_j — центр кластера C_j, $d(x_o, \mu_j)$ — выбранная метрика.

Таблица 10

Исходные данные

№№	Prob	Sever	y_i	δ_i
1	0,10	3	1	A
2	0,40	5	2	A
3	0,05	9	1	B
4	0,80	2	3	B
5	0,60	7	3	C
6	0,30	4	3	C
7	0,90	8	1	B
8	0,20	6	2	A

Таблица 11

Исходные данные

№№	R_{InfSec}	R_{FunSaf}	R_{FunSaf}
1	7,5	5,2	8,0
2	4,0	3,5	6,5
3	9,0	7,1	9,2
4	5,1	4,8	5,9
5	2,8	2,5	3,1
6	3,2	2,9	4,0
7	8,4	6,5	8,7
8	5,5	4,2	6,0

2. Применены следующие методы нормализации:

$$M_n = \{N_1 = \min\text{–}\max, N_2 = \text{z-score}\}$$

a) N^1 (min–max) по каждой координате j = Prob, Sever

$$\hat{f}_i^j = (f_i^j - \min_k f_k^j)/(\max_k f_k^j - \min_k f_k^j)$$

– для Prob: min=0.05, max=0.90;
– для Sever: min=2, max=9;

b) N^2 (z-score)

$$\hat{f}_i^j = (f_i^j - \mu_j)/\sigma_j$$

– $\mu_{Prob} \approx 0{,}45$; $\sigma_{Prob} \approx 0{,}28$;
– $\mu_{Sev} \approx 5{,}5$; $\sigma_{Sev} \approx 2{,}39$.
– для R_{InfSec}: min=2.8, max=9.0;
– для R_{FanSaf}: min=2.5, max=7.1;
– для R_{FunSaf}: min=3.1, max=9.2.

Ограничения и допущения:

– $2 \le k \le 4$ (административное требование)
– Must-link: площадки o_1 и o_3 принадлежат одному дата-центру → они должны оказаться в одном кластере.
– Cannot-link: o_5 и o_7 расположены на удалённых площадках без быстрого канала → не в одном кластере.
– Метрика: Евклидово расстояние.

3. Приняты следующие алгоритмы кластеризации:

$$Mc = \{A_1 = \text{k-means}(k = 2),$$
$$A_3 = \text{DBSCAN}(\varepsilon = 0{,}3, \min P_{ts} = 2)\}$$

а) K-means с учётом ограничений (COP-Kmeans)
б) Иерархическая кластеризация (Ward)
в) DBSCAN (для выявления выбросов)

Пример расчёта (k-means, k=3)

Нормализация (Min–Max)

$R_{InfSec}'_1 = (7.5–2.8)/(9.0–2.8)=4.7/6.2\approx0.758$;
$R_{FanSaf}'_1=(5.2–2.5)/(7.1–2.5)=2.7/4.6\approx0.587$;
$R_{FunSaf}'_1=(8.0–3.1)/(9.2–3.1)=4.9/6.1\approx0.803$.
$R_{InfSec}'_2=(4.0–2.8)/6.2\approx0.194$;
$R_{FanSaf}'_2=(3.5–2.5)/4.6\approx0.217$;
$R_{FunSaf}'_2=(6.5–3.1)/6.1\approx0.557$.

Инициализация центроидов (учитывая must-/cannot-link)

Предполагаем стартовые центры:

μ_1 = норм. o_3, μ_2 = норм. o_5, μ_3 = норм. o_4

Итерации

– Присвоение: каждый o_i к ближайшему μ_j с учётом ограничений
– Пересчёт центров как средних векторов текущих кластеров
– Повтор до сходимости

Итоговое разбиение (например):

$$C_1=\{o_1,o_3,o_7\},\ C_2=\{o_2,o_4,o_8\},\ C_3=\{o_5,o_6\}$$

4. Блок «Полного перебора». Для каждого сочетания (Np, Aq) делаем:

– нормализация → $\hat{X}$;
– задание параметров (k=2 или ε, minPts);
– кластеризация → разбиение C{p,q};
– расчёт метрик качества:

внутрикл. дисперсия $J(C)=\sum_p\sum_{x\in C_p} D(x,\mu_p)$;
силуэт Sil(C) (среднее S_i по всем точкам);
Dunn-index и т. д.

(Для краткости опишем только полученные «видимые» результаты.)

Оценка качества сегментации

Силуэт (silhouette)

– Средний silhouette ≈ 0.42 → умеренно сильная кластерная структура

Davies–Bouldin index ≈ 0.85 (чем меньше, тем лучше).

Проверка ограничений: все must-/cannot-link соблюдены

5. Пример результата

5.1. k-means(k=2) + min–max

Разбиение:

$$C_1 = \{1, 2, 4, 5\},$$
$$C_2 = \{3, 6, 7, 8\}$$

J(C) = 0,72;
Sil(C) = 0,42.

5.2. k-means(k=2) + z-score

J(C) = 0,68;
Sil(C) = 0,45 (лучш.).

5.3. DBSCAN ($\varepsilon = 0.3$, $\min P_{ts} = 2$) + min–max

Кластеры:

два «ядра» + 2 шума → Sil ≈ 0,30.

5.4. Выбор лучшей схемы по Силуэту → k-means + z-score.

6. Учёт ограничений и «корректировка» разбиения.

Заданы ограничения:

$$L = 2 \le |C_p| \le U = 5$$

на долю рисков каждого типа в кластере $\le \alpha_j = 0{,}5$;
гео-покрытие: максимум $\beta = 2$ разных δ;
внутрикл. диаметр $\le \tau_{max} = 0.8$.

Оказалось, что при исходном $C_1=\{1, 2, 4, 5\}$ гео δ={A, A, B, C} = 3 > 2 → необходимо «перебросить». Пересмотренный вариант (с сохранением k=2 и близкой конфигурации):

$$C_1' = \{1(A),\ 2(A),\ 6(C),\ 8(A)\}$$
$$C_2' = \{3(B),\ 4(B),\ 5(C),\ 7(B)\}$$

Проверка:

– размеры 4 и 4 ∈ [2, 5];
– типы в C_1': y = {1, 2, 3, 2} доля каждого ≤ 0,5;
– типы в C_2': y = {1, 3, 2, 1} – OK;
– гео C_1': {A,C}=2; C_2': {B,C}=2
– внутрикл. диаметр (по z-score) <0.8 – OK

7. Приоритизация кластеров по «уровню риска» R_p:

$$R_p = w_1\cdot\sum\text{Prob} + w_2\cdot\sum\text{Sever} + w_3\cdot|C_p|$$

Пусть $w_1=w_2=0.4$, $w_3=0.2$

–C_1': $\sum$Prob=0.10+0.40+0.30+0.20=1.00;
$\sum$Sev=3+5+4+6=18; $|C_1'|=4$
$R_1 = 0.4\cdot1.00 + 0.4\cdot18 + 0.2\cdot4 = 0.4+7.2+0.8 = 8.4$

–C_2': $\sum$Prob=0.05+0.80+0.60+0.90=2.35;
$\sum$Sev=9+2+7+8=26; $|C_2'|=4$
$R_2 = 0.4\cdot2.35 + 0.4\cdot26 + 0.2\cdot4 = 0.94+10.4+0.8 = 12.14$

8. Итоговая последовательность реагирования

Так как $R_2>R_1$, первыми отрабатываем события из кластера C_2', затем из C_1':

– 1-я волна: s_3,s_4,s_5,s_7
– 2-я волна: s_1,s_2,s_6,s_8

Продемонстрирован цикл «перебор нормализаций → перебор алгоритмов → выбор по качеству → учёт ограничений → финальное разбиение → приоритизация кластеров».

На каждом этапе опираемся на заданные целевые функции (J, Sil, Dunn) и на бизнес-ограничения (L, U, α, β, τ_{max}).

Конечный результат – упорядоченный список инцидентов для первоочередного реагирования.

Сравнение с другими методами:

– Иерархическая (Ward) дала близкое деление, но «слила» o_4 и o_8 с C_3 (нарушение cannot-link), поэтому отвергнута;

– DBSCAN при eps=0.3, minPts=2 пометила o_5,o_6 как отдельный кластер-«шум» → подтверждает их низкий риск.

Формирование очередности реагирования:

для каждого кластера вычисляем средний суммарный риск $S_j = avg(R_{InfSec} + R_{FanSaf} + R_{FunSaf})$.

– C_1 (o_1,o_3,o_7): высокие значения → $S_1 \approx (7.5+9.0+8.4 + \ldots)/3 \approx 24.0$ → Критический

– C_2 (o_2,o_4,o_8): средние → $S_2 \approx (4.0+5.1+5.5 + \ldots)/3 \approx 15.0$ → Средний

– C_3 (o_5,o_6): низкие → $S_3 \approx (2.8+3.2+2.9 + \ldots)/2 \approx 8.5$ → Низкий

Порядок реагирования:

1) Кластер C_1 – экстренный аудит, изоляция узлов, патч-менеджмент, форсированный мониторинг.

2) Кластер C_2 – плановая проверка, дополнительная защита периметра, стресс-тест.

3) Кластер C_3 – текущий мониторинг, обновление политик безопасности.

Вывод: Приведённый пример демонстрирует полный цикл: от формальной постановки до выработки приоритетной схемы реагирования на киберинциденты. Такой подход позволяет сгруппировать однотипные по уровню риска объекты и оптимизировать распределение ресурсов при инцидентах.

IV. Обсуждение результатов: реализация проактивной комплексной модели выявления и идентификации угроз нарушения информационной безопасности надзорной деятельности территориально-распределенного объекта повышенной опасности мегаполиса

Замечания по реализации:

комбинация нормализации и алгоритма подбирается «по факту» на контрольной выборке;

при большом N предпочтительны распределённые и эвристические методы (mini-batch k-means, DBSCAN с ускорением);

мульти-criteria выбор (п.7) может быть формализован через взвешенную сумму нормированных показателей качества;

реактивная последовательность (п.8) допускает гибкую настройку весов w_k под актуальные угрозы.

Формализация позволяет:

гибко конструировать и тестировать разные схемы нормализации и кластеризации,

контролировать качество и управлять ограничениями,

автоматизировать выбор оптимального метода,

выстраивать чёткую последовательность реагирования на киберинциденты.

V. Заключение

Результаты исследования являются обоснованными, достоверными и актуальными.

Актуальность представленных материалов обусловлена цифровой трансформацией промышленности (Индустрия 4.0), усложнением информационной инфраструктуры опасных производственных объектов, ростом киберугроз в промышленных системах управления и необходимостью комплексного подхода к обеспечению безопасности производственных процессов опасных производственных объектов в мегаполисах, что подтверждение статистикой киберугроз от "Лаборатории Касперского", ссылками на современные исследования в области информационной безопасности.

Представленные в работе результаты основаны на комплексном междисциплинарном подходе к исследованию рисков информационной, функциональной и пожарной безопасности с применением формальной математической постановки задачи с четкими ограничениями и допущениями, современных средств кластеризации и многоэтапной методикой с контролем качества промежуточных результатов на каждом этапе расчетов.

Существующие работы, освещающие различные аспекты методологии кластеризации, основаны на статических методах кластеризации, ограниченном количестве методов нормализации и характеризуются отсутствием гибких настроек ограничений.

Предложенный подход характеризуется низкопороговой адаптацией, возможностью поэтапного внедрения, экономической эффективностью и масштабируемостью методики.

Научная новизна комплексной методики заключается в интеграции методов оценки рисков информационной, функциональной и пожарной безопасности, динамическом выборе метода кластеризации, формализации процесса группировки и приоритизации рисков. Предложенная модель может быть реализована как математическое обеспечение систем поддержки принятия решений с минимальными затратами.

Несмотря на использование условных примеров, а не реальных данных с конкретного опасного производственного объекта, а также отсутствие результатов натурных экспериментальных исследований, что обусловлено значимостью объектов применительно к рискам чрезвычайных ситуаций мегаполисов, достоверность результатов подтверждается применением математически обоснованных методов нормализации и кластеризации, использованием нескольких метрик оценки качества кластеризации (Силуэт, индекс Дэвиса-Булдина, индекс Данна), проведением сравнительного анализа различных методов, учетом ограничений и экспертных требований.

Методика имеет высокий потенциал практического применения в системах обеспечения безопасности опасных производственных объектов.

Практическая значимость результатов исследования заключается в том, что предоставленная методика может быть использована в системах поддержки принятия решений систем обеспечения информационной безопасности органов государственного пожарного надзора, позволяет оптимизировать распределение ресурсов при реагировании на инциденты подразделений обеспечения информационной безопасности органов государственного пожарного надзора, применимы для различных территориально-распределенных объектов.

Статья подготовлена в рамках выполнения НИР «Кибермониторинг» по государственному заданию МЧС России (ЕГИСУ НИОКТР №125031703734-4).

Библиография

[1] Fang, C., Yu, D. Urban agglomeration: an evolving concept of an emerging phenomenon // Landsc Urban Plan. №162. 2017. Pp.126-136. DOI:10.1016/j.landurbplan.2017.02.014.

[2] Risdianto, E., Cs, M. Analisis Pendidikan Indonesia di Era Revolusi Industri 4.0 // Universitas Bengkulu, Geoscience. April. 2019. Pp.1-16.

[3] Maulidi, C., Adiwan, F.A, Dwicaksono, A., Winarso, H. Urban Transformation Under Technological Disruption: A Literature Review // Evergreen. №11 (2), June, 2024. Pp.1028-1039. DOI:10.5109/7183392.

[4] Финогеев А.А. Проактивный подход к мониторингу событий в сложных распределённых системах интеллектуального города с использованием технологий больших данных и предиктивной аналитики // НИР: грант № 20-71-10087. Российский научный фонд. 2020.

[5] Салихова А.Х., Швырев Е.А., Михалин В.Н. Применение методов статистического анализа при изучении состояния пожарной опасности производственных объектов // Современные проблемы гражданской защиты (Предыдущее название «Вестник Воронежского института ГПС МЧС России») 3(40) / 2021, ISSN 2658-6223. С.47-52.

[6] Кадиев Ш.К., Хабибулин Р.Ш. Модель и алгоритм проведения кластерного анализа чрезвычайных ситуаций техногенного характера // Современные проблемы гражданской защиты (Предыдущее название «Вестник Воронежского института ГПС МЧС России») 1(46) / 2023. С.20-28. ISSN 2658-6223.

[7] Вилисов В.Я., Хабибулин Р.Ш. Кластеризация пожаров на объектах топливно-энергетического комплекса по ретроспективным статистическим данным для выявления рангов пожаров // Пожаровзрывобезопасность/Fire and Explosion Safety. 2024. Т. 33. № 1. С. 83–93. DOI: 10.22227/0869-7493.2024.33.01.83-93.

[8] Орлова Д.Е., Куприенко П.С., Фурсов И.В. Алгоритмы кластеридентификации ситуаций при управлении процессами обеспечения техногенной и пожарной безопасности // Моделирование, оптимизация и информационные технологии. 2021. №9(2). С.1-15. DOI: 10.26102/2310-6018/2021.33.2.020.

[9] Бухарев, Д. А. Применение иерархического кластерного анализа для кластеризации данных информационных процессов АСУ ТП, подвергающихся воздействию кибератак / Д. А. Бухарев, А. Н. Соколов, А. Н. Рагозин // Вестник УрФО. Безопасность в информационной сфере. – 2023. – № 1(47). – С. 59-68. – DOI 10.14529/secur230106. – EDN FYCUHE.

[10] Васильев В. И., Вульфин А. М., Гвоздев В. Е., Картак В. М., Атарская Е. А. Обеспечение информационной безопасности киберфизических объектов на основе прогнозирования и обнаружения аномалий их состояния // Системы управления, связи и безопасности. 2021. № 6. С. 90-119. DOI: 10.24412/2410-9916-2021-6-90-119.

[11] Котенко И.В. Аналитическая обработка больших массивов гетерогенных данных о событиях в сфере кибербезопасности в целях оценки состояния, поддержки принятия решений и расследования компьютерных инцидентов в критически важных инфраструктурах // Отчет о научно-исследовательской работе № 21-71-20078. Российский научный фонд. 2023.

[12] Котенко И.В. Аналитическая обработка больших массивов данных о событиях кибербезопасности с применением суперкомпьютерных вычислений / И.В. Котенко, И.Б. Саенко, И.Б. Паращук [и др.] // Программные продукты и системы. 2024. № 4. С. 487-494. DOI 10.15827/0236-235X.148.487-494. – EDN OOENUN.

[13] Шкарупета Е.В. Влияние цифровой устойчивости и информационной безопасности на устойчивое развитие промышленных предприятий / Е.В. Шкарупета, Е.А. Ильина, А.В. Холманских // Организатор производства. 2023. Т. 31. № 3. С. 64-77. DOI 10.36622/VSTU.2023.80.72.006. – EDN XLXHDL.

[14] Абрамова Т.В. Обнаружение аномалий и нейтрализация угроз в распределенных автоматизированных системах управления на основе мониторинга сетевых информационных потоков / Дисс. на соиск. ученой степ. канд. технич. наук / Абрамова Таисия Вячеславовна, 2024 – 235 с. – EDN KFSLDN.

[15] Aly, S., Tyrychtr J., Kvasnicka, R., Vrana, I. Novel methodology for developing a safety standard based on clustering of experts' assessments of safety requirements // Safety Science 140 (2021) 105292 https://doi.org/10.1016/j.ssci.2021.105292.

[16] Huang, J., Xu, Z., Yang, F., Zhang, W., Cai, S., Luo, J., Xie, G., Li, T. Fire Risk Assessment and Warning Based on Hierarchical Density-Based Spatial Clustering Algorithm and Grey Relational Analysis // Mathematical Problems in Engineering, 2022, 7339312, 8 pages, 2022. https://doi.org/10.1155/2022/7339312.

[17] Deng, F., Gu, W., Zeng, W., Zhang, Z., Wang, F., Hazardous Chemical Accident Prevention Based on K-Means Clustering Analysis of Incident Information // Institute of Electrical and Electronics Engineers vol. 8, pp. 180171-180183, 2020, doi: 10.1109/ACCESS.2020.3028235.

[18] Старовойтов В.В., Голуб Ю.И. Нормализация данных в машинном обучении. *Информатика*. 2021;18(3):83-96. https://doi.org/10.37661/1816-0301-2021-18-3-83-96

[19] Винюков, И. А. Финансовая состоятельность регионов РФ в 2010-2014 годах: продолжение классификационного анализа / И. А. Винюков, Е. В. Маевский, П. В. Ягодовский // Мир новой экономики. – 2017. – № 2. – С. 58-69. – EDN YSPLCT.

[20] Тюрин, А. Г. Кластерный анализ, методы и алгоритмы кластеризации / А. Г. Тюрин, И. О. Зуев // Вестник МГТУ МИРЭА. – 2014. – № 2(3). – С. 86-97. – EDN QJOXHN.

Clustering of threats and identification of risks of information security breaches at hazardous industrial facilities

M.A. Tukmacheva, A.V. Shestakov, K.Z. Bilyatdinov

Annotation — The paper presents the results of developing a proactive integrated model for identifying and identifying threats to information security at a geographically distributed high-risk facility in a metropolitan area. This model utilizes methods and models for threat analysis and clustering, taking into account potential risks. These models are based on interdisciplinary models that integrate MITRE ATT&CK, IEC 62443, and HAZOP/FMEA methodologies and indicators of the essential properties of digital twins of infrastructure components of real-world facilities. These models also utilize dynamic scenario assessments that include the impact of cyber threats on the occurrence of physical accidents and fire emergencies. The scientific novelty of the research results lies in the application of the Lance-Williams hierarchical clustering algorithm to decompose situations at a geographically distributed hazardous industrial facility and in the consideration of information, functional, and fire safety risks through parameters for changing the edge metrics of the situation graph. The conducted modeling of critical situations with varying ratios of characteristics, with the threat list volume varying by an order of magnitude while impacting various resources of the facility's information infrastructure, demonstrated the possibility of ensuring stable error through dynamic procedures for selecting and applying the clustering method and algorithm. The presented results are proposed to be implemented as mathematical support for decision support systems for information security administrators of information security systems for hazardous industrial facilities in a metropolitan area.

Key words — algorithms, clustering, risks, information security, functional safety, fire safety, hazardous production facility.

REFERENCES

[1] Fang, C., Yu, D. Urban agglomeration: an evolving concept of an emerging phenomenon // Landsc Urban Plan. No. 162. 2017. Pp. 126-136. DOI:10.1016/j.landurbplan.2017.02.014.

[2] Risdianto, E., Cs, M. Analisis Pendidikan Indonesia di Era Revolusi Industri 4.0 // Universitas Bengkulu, Geoscience. April. 2019. Pp.1-16.

[3] Maulidi, C., Adiwan, F.A, Dwicaksono, A., Winarso, N. Urban Transformation Under Technological Disruption: A Literature Review // Evergreen. №11 (2), June, 2024. Pp.1028-1039. DOI:10.5109/7183392. [2]

[4] Finogeev A.A. Proactive approach to monitoring events in complex distributed systems of a smart city using big data and predictive analytics technologies // R&D: grant No. 20-71-10087. Russian Science Foundation. 2020.

[5] Salikhova A.Kh., Shvyrev E.A., Mikhalin V.N. Application of statistical analysis methods in studying the fire hazard status of industrial facilities // Modern problems of civil defense (Previous name "Bulletin of the Voronezh Institute of GPS of the Ministry of Emergencies of Russia") 3(40) / 2021, ISSN 2658-6223. Pp.47-52.

[6] Kadiev Sh.K., Khabibulin R.Sh. Model and algorithm for conducting cluster analysis of man-made emergencies // Modern problems of civil defense (Previous title "Bulletin of the Voronezh Institute of the GPS of the Ministry of Emergency Situations of Russia") 1 (46) / 2023. Pp. 20-28. ISSN 2658-6223.

[7] Vilisov V.Ya., Khabibulin R.Sh. Clustering of fires at fuel and energy facilities based on retrospective statistical data to identify fire ranks // Fire and Explosion Safety. 2024. Vol. 33. No. 1. Pp. 83–93. DOI: 10.22227/0869-7493.2024.33.01.83-93.

[8] Orlova D.E., Kuprienko P.S., Fursov I.V. Algorithms for cluster identification of situations in managing processes of ensuring technogenic and fire safety // Modeling, optimization and information technology. 2021. No. 9 (2). P. 1-15. DOI: 10.26102 / 2310-6018 / 2021.33.2.020.

[9] Bukharev, D. A. Application of hierarchical cluster analysis for clustering data of information processes of APCS exposed to cyber attacks / D. A. Bukharev, A. N. Sokolov, A. N. Ragozin // Bulletin of the Ural Federal District. Security in the information sphere. - 2023. - No. 1 (47). - P. 59-68. - DOI 10.14529 / secur230106. – EDN FYCUHE.

[10] Vasiliev V. I., Vulfin A. M., Gvozdev V. E., Kartak V. M., Atarskaya E. A. Ensuring information security of cyber-physical objects based on forecasting and detecting anomalies in their state // Control, Communications and Security Systems. 2021. No. 6. pp. 90- 119. DOI: 10.24412/2410-9916-2021-6-90-119.

[11] Kotenko I. V. Analytical processing of large arrays of heterogeneous data on events in the field of cybersecurity in order to assess the state, support decision-making and investigate computer incidents in critical infrastructures // Report on research work No. 21-71-20078. Russian Science Foundation. 2023.

[12] Kotenko I. V. Analytical processing of big data arrays on cybersecurity events using supercomputer computing / I.V. Kotenko, I.B. Saenko, I.B. Parashchuk [et al.] // Software products and systems. 2024. No. 4. pp. 487-494. DOI 10.15827/0236-235X.148.487-494. – EDN OOENUN.

[13] Shkarupeta E.V. The impact of digital resilience and information security on the sustainable development of industrial enterprises / E.V. Shkarupeta, E.A. Ilyina, A.V. Kholmanskikh // Production organizer. 2023. Vol. 31. No. 3. pp. 64-77. DOI 10.36622/VSTU.2023.80.72.006. – EDN XLXHDL.

[14] Abramova T.V. Anomaly detection and threat mitigation in distributed automated control systems based on monitoring of network information flows / Diss. for the candidate of technical sciences / Abramova Taisiya Vyacheslavovna, 2024 – 235 p. – EDN KFSLDN.

[15] Aly, S., Tyrychtr J., Kvasnicka, R., Vrana, I. Novel methodology for developing a safety standard based on clustering of experts' assessments of safety requirements // Safety Science 140 (2021) 105292 https://doi.org/10.1016/j.ssci.2021.105292.

[16] Huang, J., Xu, Z., Yang, F., Zhang, W., Cai, S., Luo, J., Xie, G., Li, T. Fire Risk Assessment and Warning Based on Hierarchical Density-

M.A. Tukmacheva - adjunct student at the Faculty of Advanced Training and Qualification at the St. Petersburg University of the State Fire Service of the Ministry of Emergency Situations of Russia (email: mtukmacheva@mail.ru).

A.V. Shestakov - Doctor of Engineering, Leading Researcher at the St. Petersburg University of the State Fire Service of the Ministry of Emergency Situations of Russia (email: alexandr.shestakov01@yandex.ru).

K.Z. Bilyatdinov - Doctor of Engineering, Professor at the St. Petersburg State University of Aerospace Instrumentation (email: k74b@mail.ru).

Based Spatial Clustering Algorithm and Gray Relational Analysis // Mathematical Problems in Engineering, 2022, 7339312, 8 pages, 2022. https://doi.org/10.1155/2022/7339312.

[17] Deng, F., Gu, W., Zeng, W., Zhang, Z., Wang, F., Hazardous Chemical Accident Prevention Based on K-Means Clustering Analysis of Incident Information // Institute of Electrical and Electronics Engineers vol. 8, pp. 180171-180183, 2020, doi: 10.1109/ACCESS.2020.3028235.

[18] Starovoytov V.V., Golub Yu.I. Data Normalization in Machine Learning. Informatika. 2021;18(3):83-96. https://doi.org/10.37661/1816-0301-2021-18-3-83-96

[19] Vinyukov, I. A. Financial Solvency of the Russian Federation Regions in 2010-2014: Continuation of the Classification Analysis / I. A. Vinyukov, E. V. Maevsky, P. V. Yagodovsky // World of the New Economy. – 2017. – No. 2. – Pp. 58-69. – EDN YSPLCT.

[20] Tyurin, A. G. Cluster Analysis, Methods, and Algorithms of Clustering / A. G. Tyurin, I. O. Zuev // Vestnik MGTU MIREA. – 2014. – No. 2(3). – Pp. 86-97. – EDN QJOXHN.