vulnerability leveling is ready for safe operation.

For graphical interpretation of dependencies, graphic materials are presented, for which simulations were performed in the MatLab system. The graphic materials clearly indicate the possibility of obtaining a state of safe operation of information systems depending on the intensity of stopping attempts by the protection system of illegal access to information, and the intensity of such attempts at the entrance to the protection system.

This will allow developers of information systems and service personnel to have quantitative indicators of the coefficient of readiness for safe operation of the system and decision-making regarding possible vulnerabilities.

Keywords: coefficient of readiness for safe operation, modeling, graphical interpretation, information systems, parameters, attempts of illegal access to information.

Ахрамович Володимир Миколайович, д.т.н., професор, професор кафедри кібербезпека, Державного некомерційного підприємства «Державний університет «Київський авіаційний інститут», м. Київ, Україна.

Akhramovych Volodymyr, Doctor of Technical Science, Professor, Professor, Department of Cybersecurity of the State non-commercial company state university «Kyiv aviation institute», Kyiv, Ukraine.

E-mail: 12z@ukr.net.

ORCID ID 0000-0002-0086-9131.

Чупрун Сергій, аспірант на кафедрі систем інформаційного та кібер- нетичного захисту, державного університету інформаційно - телекомунікаційних технологій, Київ, Україна.

Chuprun Sergii, postgraduate of the Department of Information and Cyber Defense Systems of the State University of information and telecommunication technologies, Kyiv, Ukraine.

E-mail: ua0988888424@gmail.com.

ORCID ID:0009-0007-7336-1068.

Придибайло Роман, аспірант на кафедрі систем інформаційного та кібернетичного захисту, Державного університету інформаційно-телекомунікаційних технологій, Київ, Україна

E-mail: drake0103@gmail.com.

ORCID ID:0009-0003-1747-7518.

Prydybailo Roman, postgraduate of the Department of Information and Cyber Defense Systems of the State University of information and telecommunication technologies, Kyiv, Ukraine.

E-mail: drake0103@gmail.com.

ORCID ID:0009-0003-1747-7518.

УДК 336.71:004.056

МЕТОД ПОБУДОВИ ПРОФІЛЮ КЛЮЧОВИХ ФАКТОРІВ РИЗИКУ КІБЕРБЕЗПЕКИ СУЧАСНИХ РОЗПОДІЛЕНИХ ІНФОРМАЦІЙНИХ СИСТЕМ

Дмитро Палко, Лариса Мирутенко

Оцінка та аналіз ризиків кібербезпеки є фундаментальними аспектами формування надійної і ефективної СУІБ, особливо в умовах стрімкого розвитку технологій та зростання складності сучасних РІС. Традиційні методи оцінки ризиків, що засновані переважно на концептуальних підходах і класичних методах, мають ряд обмежень та є малоефективними в умовах сучасних масштабованих розподілених систем, оскільки не враховують динаміку середовища та не забезпечують ефективний аналіз взаємозв'язків між численними факторами ризику. В дослідженні запропоновано метод побудови профілю ключових факторів ризику сучасних розподілених інформаційних систем на основі урахування кореляційного аналізу та моделювання їх взаємозв'язків, що дозволяє підвищити ефективність процесу оцінки ризиків кібербезпеки в умовах динамічного середовища сучасних масштабованих розподілених інформаційних систем. На основі запропонованого методу здійснено розробку профілю ключових факторів ризику сучасних розподілених систем, проведено аналіз їх статистичної важливості та кореляції, а також визначено і структуровано основні заходи та контролі інформаційної безпеки, які демонструють найкращі показники ефективності в умовах розподіленості середовища, враховують як технологічні, так і організаційні аспекти, забезпечуючи системний підхід до управління ризиками ІБ, зменшення впливу загроз і підвищення стійкості розподілених систем до можливих атак. Запропонований підхід до оптимізації вибору вхідного набору ознак та виокремлення найбільш вагомих факторів ризику на основі спроектованого профілю ключових факторів ризику для сучасних РІС продемонстрував тотожний результат по числовому показнику кількості відібраних для аналізу метрик у порівнянні з факторним аналізом за допомогою методу головних компонент (PCA) – 42 метрики у порівнянні із 40 для PCA, але при цьому забезпечив покращення загальних показників точності класифікації для проєктованих моделей оцінки ризику кібербезпеки в РІС на 4% у порівнянні з контрольною моделлю на основі PCA, що підтверджує його ефективність у контексті адаптивного аналізу ризиків у розподілених середовищах.

Ключові слова: *інформаційна безпека, ризик інформаційної безпеки, фактори ризику, оцінка ризиків, управління ризиками, розподілена інформаційна система, нейронна мережа.*

ВСТУП

Сучасні розподілені інформаційні системи (РІС) відіграють ключову роль у забезпеченні безперервності бізнес-процесів та цифровій трансформації різних сфер людської діяльності, зокрема у фінансовому секторі, охороні здоров'я, енергетиці, телекомунікаціях та державному управлінні. На сьогоднішній день РІС стають фундаментом діяльності більшості організацій, забезпечуючи доступність критичних сервісів, обробку великих масивів даних та інтеграцію широкого спектру інформаційних активів та ресурсів. Завдяки своїй архітектурній гнучкості, масштабованості та можливості обробки великих обсягів даних у реальному часі, розподілені інформаційні системи підвищують ефективність комунікацій, надають користувачам доступ до обчислювальних ресурсів незалежно від їх місцезнаходження, а також дозволяють розподіляти навантаження, підвищувати відмовостійкість та гнучко масштабувати систему залежно від вимог до продуктивності та надійності. Проте, одночасно з перевагами таких систем постає низка критичних викликів, пов'язаних із забезпеченням кібербезпеки, захистом даних та управлінням ризиками інформаційної безпеки (ІБ). Розвиток РІС супроводжується зростанням складності архітектури сучасних систем, появою широкого спектру нових технологічних компонентів та суттєвим підвищенням кількості кіберзагроз, орієнтованих на масштабовані системи. Оцінка ризиків кібербезпеки РІС в таких умовах є надзвичайно важливим завданням, оскільки дозволяє ідентифікувати потенційні загрози, визначати критичні вразливості та формувати ефективні стратегії захисту [1].

За даними щорічного глобального дослідження стану кібербезпеки «STATE OF CYBERSECURITY 2022: GLOBAL UPDATE ON WORKFORCE EFFORTS, RESOURCES AND CYBEROPERATIONS» від компанії ISACA переважна більшість компаній-респондентів принаймні один раз на рік проводять регулярну оцінку ризиків ІБ (рис. 1), і цей показник збільшується з кожним роком, що сигналізує про позитивну тенденцію та демонструє посилення уваги до питань ризик-менеджменту [2].

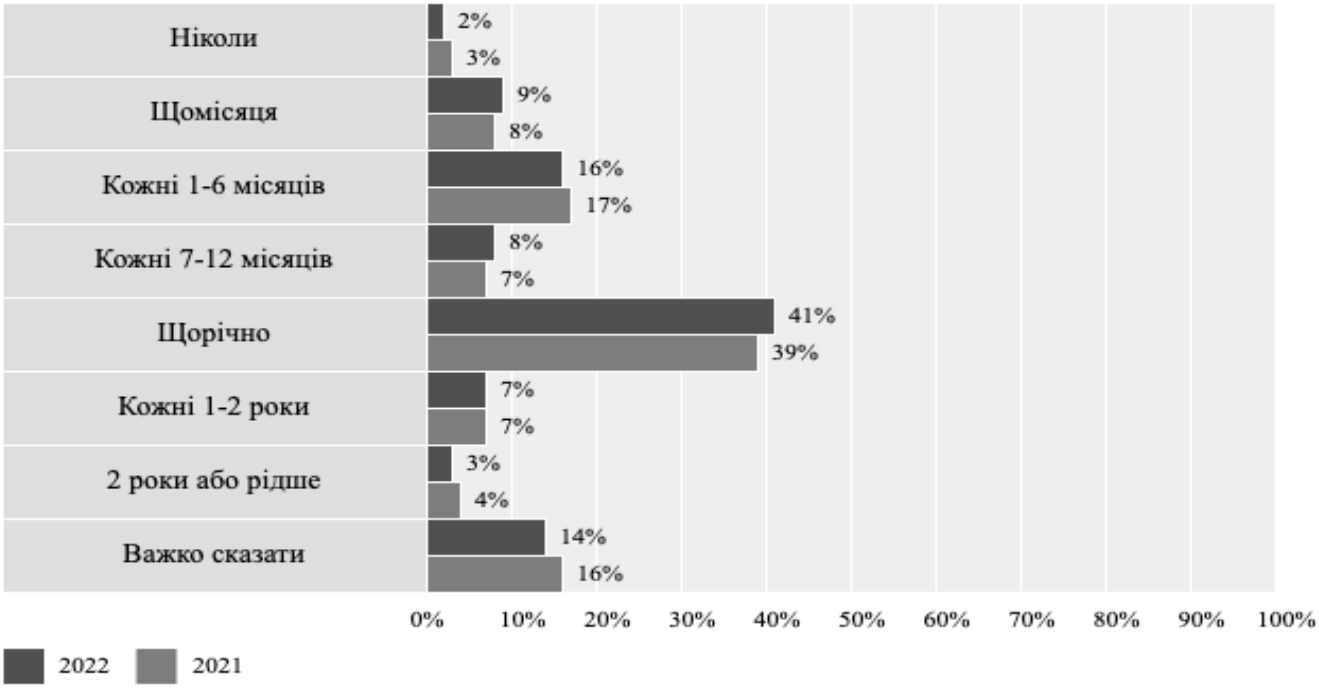

Рисунок 1. Тенденції оцінки корпоративних кіберризиків 2021-2022 рр.

З іншого боку, відповідно до Bitdefender Cybersecurity Assessment Report 2024 більше половини організацій (57%) зазнали порушення або витоку даних за останні 12 місяців (на 6% більше, ніж у попередньому році), при цьому 93% опитаних в наступному році планують збільшити інвестиції в оцінку ризиків ІБ та проактивні стратегії захисту [3].

Загрози кібербезпеці розподілених інформаційних систем постійно еволюціонують, що зумовлює необхідність розробки нових підходів до оцінювання ризиків

кібербезпеки та впровадження ефективних механізмів протидії. Традиційні методи аналізу ризиків ІБ часто виявляються недостатньо ефективними через їх обмеженість у врахуванні специфіки децентралізованих архітектур, змінності загрозового середовища та складності прогнозування кіберінцидентів. Тому особливої актуальності набуває проблема ідентифікації ключових факторів ризику РІС, що є основою для формування адаптивних стратегій ризик-менеджменту та підвищення рівня безпеки інформаційних систем. Недостатня увага до факторів ризику може призвести до значних фінансових, репутаційних та операційних втрат [4].

Відсутність єдиної методології щодо побудови профілю основних загроз та контролів безпеки ускладнює процес прийняття рішень щодо захисту РІС. Зокрема, необхідно розробити підхід, який дозволить не лише виявляти найбільш значущі фактори ризику, а й проводити їх детальний аналіз, оцінювати рівень впливу та визначати взаємозв'язки між різними факторами, встановлювати кореляційні залежності для створення ефективної моделі оцінки загроз та відповідних контрзаходів.

Основні задачі дослідження включають наступні аспекти:

- Аналіз останніх досліджень та існуючих методів оцінки ризиків кібербезпеки РІС.
- Виявлення наявних прогалин в теорії та практиці управління ризиками, визначення недоліків класичних підходів до аналізу ризиків у децентралізованих системах.
- Розробка та обґрунтування методу побудови профілю ключових факторів ризику, що можуть спричинити потенційні інциденти ІБ в умовах фізичної та функціональної розподіленості ресурсів, а також основних контролів безпеки сучасних РІС, на основі кореляційного аналізу та моделювання їх взаємозв'язків. Створення профілю ключових факторів ризику сучасних РІС та обґрунтування його практичного застосування для покращення кібербезпеки розподілених систем та підвищення ефективності процесу ризик-менеджменту.

Таким чином, дане дослідження спрямоване на розробку методу побудови профілю ключових факторів ризику кібербезпеки сучасних розподілених інформаційних систем на основі аналізу їх взаємозв'язків та рівня статистичної важливості. Запропонований підхід дозволить оптимізувати процес ризик-менеджменту, підвищити точність прогнозування кіберзагроз та забезпечити ефективні механізми захисту інформаційних ресурсів у розподіленому середовищі.

Аналіз останніх досліджень і публікацій

Наукові дослідження у сфері оцінки ризиків кібербезпеки охоплюють широкий спектр методів, включаючи класичні математично-ймовірнісні підходи, методи експертного оцінювання, моделювання загроз та аналізу історичних даних. Серед ключових напрямків можна виділити наступні:

Класичні математично-статистичні методи використовують ймовірнісні моделі та статистичні підходи для оцінки на основі історичних даних та поточного стану системи. Вони засновані на аналізі емпіричних даних про попередні інциденти, частоту реалізації загроз, час між атаками та масштаби завданих збитків. Недоліком статистичного підходу є недостатня адаптивність до динамічних умов середовища РІС та значна залежність від якості і репрезентативності історичних даних [5].

Евристичні та експертні методи оцінки передбачають використання чітко формалізованих правил та залучення кваліфікованих фахівців із інформаційної безпеки до суб'єктивної оцінки рівнів ризику. Сутність методу експертних оцінок полягає у проведенні експертного аналізу проблеми з застосуванням кількісної чи якісної оцінки гіпотез та подальшої обробки результатів. Перевагою методу є простота та доступність для практичного застосування. Обмеженнями для цього підходу є висока залежність від компетентності експертів, суб'єктивність оцінок та обмежена відтворюваність.

Методи моделювання передбачають використання формальних, заздалегідь визначених математичних чи логічних моделей для оцінювання ризику. Такі підходи можуть включати ймовірнісні моделі, баєсові мережі, марковські процеси, теоретико-ігрові моделі, нечітку логіку, методи імітаційного моделювання та симуляційні техніки, а також методи засновані на алгоритмах машинного навчання та штучного інтелекту. Методи моделювання є найширше представленими на практиці, оскільки дозволяють найкращим чином відобразити складні взаємозв'язки між загрозами, вразливостями, контролями безпеки та потенційними наслідками. Порівняно з попередніми підходами методам моделювання притаманний ряд переваг, таких як об'єктивність та формалізованість, гнучкість та адаптивність, масштабованість, а також можливість аналізу

складних взаємозв'язків. Методи моделювання, опираючись на формальні математичні принципи та аналітичні інструменти, пропонують більш гнучкий, адаптивний та науково обґрунтований підхід до оцінювання ризиків розподіленого середовища. Вони дозволяють врахувати складну динаміку кіберзагроз і допомагають ухвалювати аргументовані рішення в умовах невизначеності [6-8].

Таким чином, висока складність аналізу ризиків у розподілених середовищах через децентралізовану структуру та динамічний характер РІС, їх масштабованість та широкий спектр наявних загроз, а також обмеженість класичних методів, що не враховують раніше невідомі загрози та передбачають високу суб'єктивність, ресурсоємність і складність впровадження в умовах масштабованих та гетерогенних розподілених систем, вимагають розробки більш гнучких та адаптивних підходів до оцінювання ризиків кібербезпеки розподіленого середовища, що враховували б інтеграцію сучасних методів інтелектуального аналізу, особливості архітектури РІС та взаємозв'язки між факторами ризику.

Дослідження факторів ризику в розподіленому середовищі заслуговує особливої уваги. За даними щорічного дослідження «STATE OF ENTERPRISE RISK MANAGEMENT 2020» компанії ISACA (рис. 2), найбільшими викликами в сфері корпоративних ризиків є фактори, пов'язані з появою нових загроз, змінами / досягненнями в розвитку технологій, а також слабким кадровим потенціалом та відсутністю необхідних навичок і досвіду в спеціалістів та існуючих командах з кібербезпеки [9].

З іншого боку, за даними цього дослідження для попередження/пом'якшення наслідків потенційних проблем безпеки найбільш часто застосовуваним контролем є підвищення рівня обізнаності та проведення тренінгів з кібербезпеки серед персоналу (рис. 3).

80% підприємств-респондентів проводять навчання з підвищення обізнаності, 68 % – використовують стратегії ліквідації наслідків та відновлення після катастроф/інцидентів ІБ, і 67% застосовують загальні контролі з управління та менеджменту ІБ. Менше 50% застосовують страхування як контроль для пом'якшення наслідків; при цьому найбільшими прихильниками такого підходу є підприємства Пн. Америки та Африки [9].

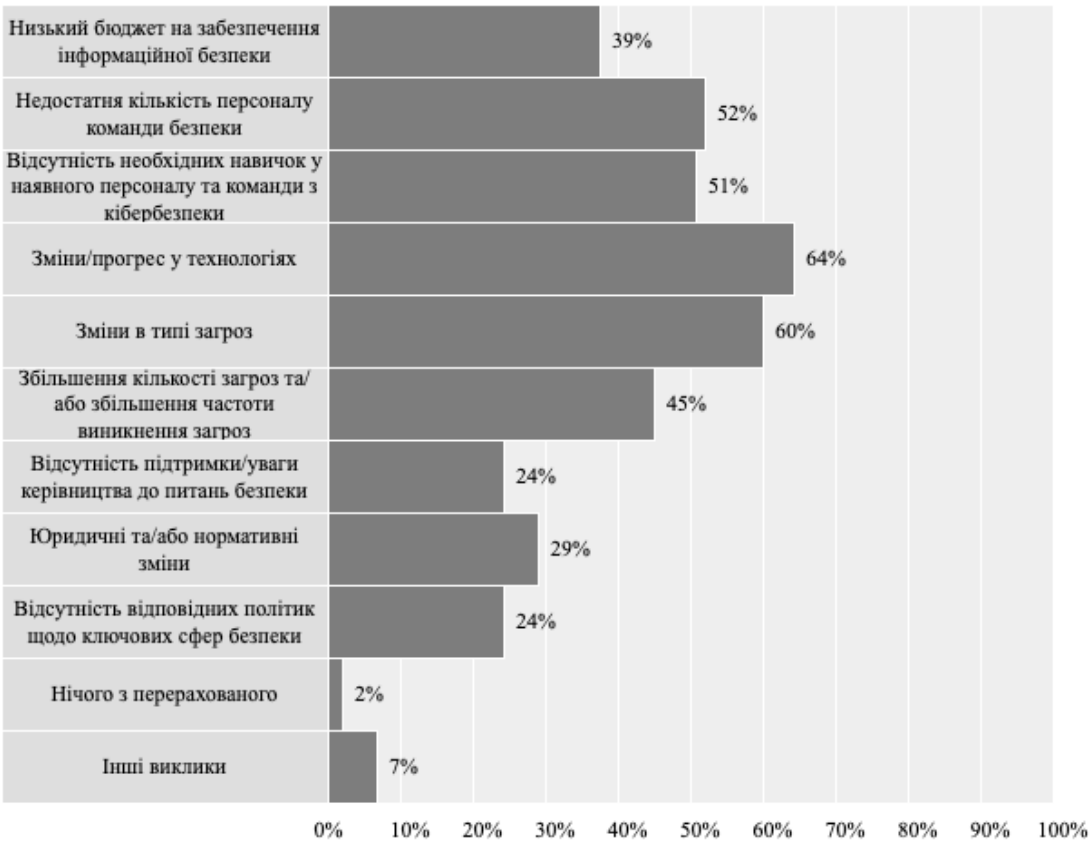

Рисунок 2. Основні виклики кібербезпеки корпоративного середовища

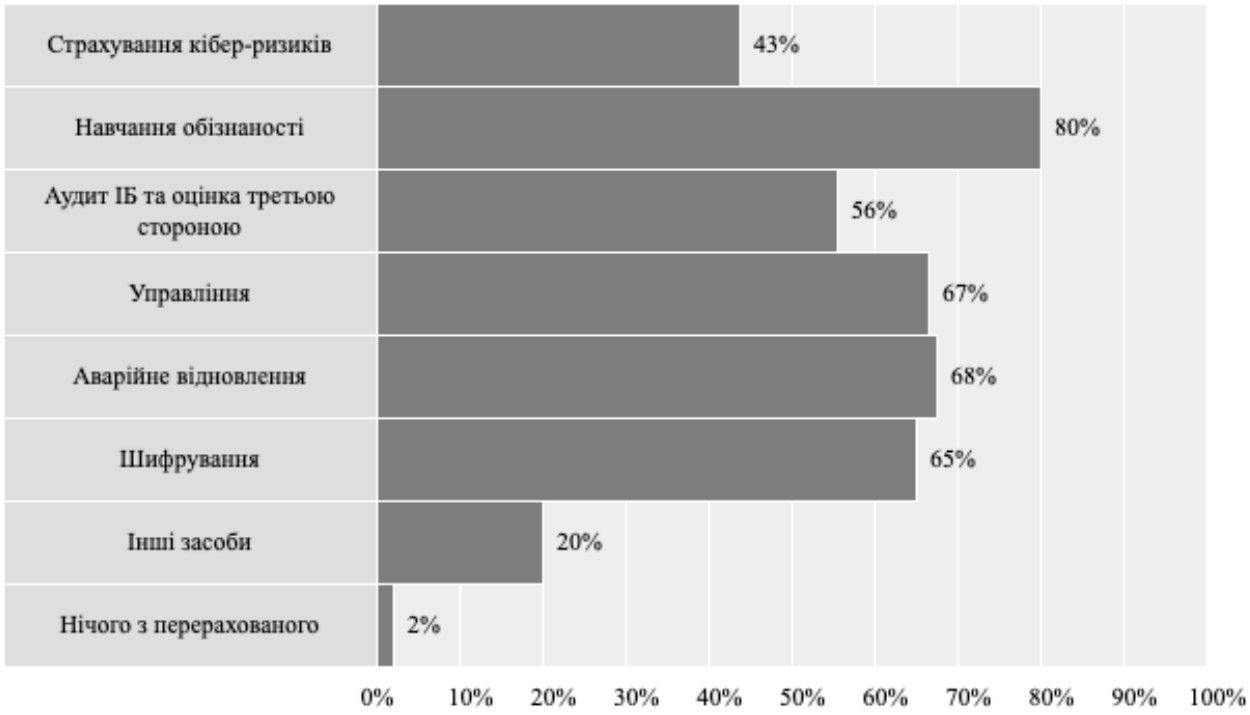

Рисунок 3. Основні контролі кібербезпеки та заходи щодо пом'якшення наслідків

Постановка проблеми

Зважаючи на масштабованість, гетерогенність та динамічність розподілених систем, традиційні методи управління ризиками часто виявляються неефективними або недостатньо точними [10]. Основні методологічні складнощі оцінювання ризиків у РІС пов'язані з наступними аспектами:

1. Висока складність, динамічність та масштабованість середовища – різнорідність обладнання та інфраструктури, постійна зміна мережевих конфігурацій, поява нових сервісів та користувачів, необхідність врахування численних взаємозалежностей між компонентами системи створюють труднощі у визначенні актуальних загроз.

2. Значна фрагментованість та неповнота даних, відсутність єдиного формату представлення у різних джерелах – ефективне оцінювання ризику в розподіленому середовищі потребує збору та оперативної обробки великих масивів актуальних даних, які часто розподілені між різними доменами, можуть бути неповними, складними за структурою та гетерогенними за природою походження.

3. Відсутність єдиних стандартів та методологій – необхідність універсального підходу, що враховував би специфіку розподілених систем та мав уніфіковану процедуру оцінки з врахуванням адаптації до динамічних змін середовища та ландшафту потенційних загроз.

4. Низька ефективність традиційних методів оцінки – класичні підходи виявляють низку обмежень у масштабованому та динамічному середовищі розподілених систем, пов'язаних із складністю і нелінійністю взаємозв'язків між факторами ризику, відсутністю гнучких механізмів аналізу гетерогенних даних РІС та виявлення нелінійних залежностей і прихованих закономірностей, недостатньою адаптивністю та реактивністю, а також необхідністю оперативного аналізу в режимі часу, близькому до реального.

Для розв'язання цих проблем необхідно розробити метод, який дозволяє проводити комплексний аналіз факторів ризику та їх взаємозв'язків, що забезпечить підвищення ефективності процесу оцінки ризиків в динамічних масштабованих середовищах. Необхідність ідентифікації та попереднього аналізу основних факторів ризику, встановлення їх взаємозв'язків та кореляційних залежностей, безпосередньо пов'язана з якістю підготовки вхідних даних для побудови ефективних моделей інтелектуального аналізу ризиків. Попередній етап статистичного дослідження та оцінки факторів ризику дозволяє:

• Оптимізувати вибір ознак (feature selection): Виділення найсуттєвіших факторів ризику та контролів безпеки сприяє

зменшенню розмірності вхідних даних і мінімізує «шум», що у кінцевому підсумку покращує збіжність і стабільність процесу навчання нейронних мереж. Це особливо актуально для глибинних архітектур, де надмірна кількість ознак може призвести до складнощів оптимізації та ризику перенавчання (overfitting).

• Покращити інтерпретованість результатів: Попередній аналіз взаємозв'язків між факторами ризику дає змогу краще зрозуміти структуру вхідних даних. Таким чином, навіть при застосуванні глибоких нейронних мереж, які зазвичай працюють за принципом «чорної скриньки» (black box), стає можливим інтерпретувати вплив окремих груп ознак на рівень ризику. Знання про те, які фактори ризику є ключовими, дозволяє фокусувати увагу на їх динаміці та значущості при подальшій оптимізації моделі.

• Підвищити точність і надійність моделі: Попереднє дослідження причинно-наслідкових зв'язків і взаємозалежностей між факторами ризику допомагає моделі глибокого навчання більш ефективно «навчитися» складним патернам при аналізі даних. Глибинні нейронні мережі здатні виявляти нелінійні закономірності, але наявність початкового, сформованого уявлення про основні детермінанти ризику підсилює впевненість у тому, що модель не буде витрачати обчислювальні ресурси на незначущі або слабо впорядковані ознаки. Узгодити аналітичний підхід із практичними вимогами: У реальних умовах експлуатації РІС, особливо коли йдеться про оцінювання ризиків кібербезпеки, важливо, щоб результати нейромережевих моделей були не лише точними, але й корисними для ухвалення рішень. Попередня статистично-аналітична розвідка дає змогу краще «пояснити» результат моделі та коректно інтегрувати її у процеси управління безпекою.

У дослідженні [11] запропоновано орієнтований на дані прогнозний аналіз загроз кібербезпеки шляхом дослідження ключових факторів ризику та вимірювання значущості ознак на основі отриманих в ході опитування даних. Пропонується вдосконалити використаний в рамках вищеописаного дослідження підхід на основі урахування кореляційного аналізу факторів ризику та моделювання їх взаємозв'язків, що дозволить підвищити ефективність процесу ризик-менеджменту в розподіленому середовищі, а також побудувати актуальний для предметної області динамічних розподілених інформаційних систем профіль ключових факторів ризику та контролів безпеки сучасних РІС. Метою побудови профілю ключових факторів ризику є ідентифікація та виокремлення основних чинників ризику ІБ, що притаманні сучасним розподіленим інформаційним системам, а також аналіз найбільш значущих контролів безпеки для розробки на основі них рекомендацій щодо усунення потенційних загроз.

Результати дослідження

Запропонована методологія побудови профілю ключових факторів ризику кібербезпеки розподілених інформаційних систем заснована на комплексному підході, що поєднує анкетне опитування експертів, статистичний аналіз отриманих результатів та моделювання взаємозв'язків між факторами ризику. Основні етапи цього процесу можна представити наступним чином:

1. **Збір даних через анкетування** – формування репрезентативної вибірки експертів у сфері забезпечення кібербезпеки РІС, проведення опитування у два етапи (пілотне тестування та основне дослідження).
2. **Попередня обробка даних** – виявлення аномальних значень, оцінка якості вибірки, нормалізація та усунення викидів.
3. **Статистичний аналіз** – визначення середніх значень та дисперсії, ранжування факторів ризику та контрольних механізмів.
4. **Моделювання взаємозв'язків** – перевірка нормальності розподілу даних, аналіз кореляційних зв'язків між факторами ризику.
5. **Формування профілю ризиків** – побудова структурованого профілю, що дозволяє ідентифікувати найбільш значущі загрози та оптимізувати заходи кібербезпеки.

Збір даних через анкетування. Для збору вибірки даних для аналізу використовується метод анкетування. В якості респондентів можуть виступати інженери з інформаційної безпеки різного рівня підготовки, спеціалісти з тестування на проникнення і аудиту, та провідні фахівців в області менеджменту проектів ІБ. Всі опитані мають бути обрані вибірково та повинні мати досвід забезпечення безпеки для інфраструктур інформаційних систем різного розміру та масштабу.

Анкетування проводиться у два етапи:

• Пілотне опитування – тестування анкети на обмеженій групі експертів для перевірки інструментарію дослідження, коригування

питань і формулювань.

• Основне опитування – масштабне анкетування цільової групи за допомогою остаточного варіанту анкет за двома категоріями питань: оцінка важливості ключових загроз та факторів ризику, а також оцінка частоти застосування на практиці контрольних заходів та механізмів.

Пілотне дослідження виконується перед проведенням основного анкетування та має на меті перевірку наскільки запропонована модель анкетування підходить для аналізу кінцевих метрик.

При розробці анкет враховуються найбільш поширені фактори ризику, що є загальними для більшості сучасних розподілених інфраструктур. Респондентам пропонується анонімно та суб'єктивно відповісти на питання покладаючись на власний досвід та реальну практику роботи з масштабованими розподіленими інформаційними системами. При цьому пропонується додатково розділити питання щодо факторів ризику на 3 логічні групи:

- Фактори технологічного характеру – логічні (програмні) та фізичні (апаратні) загрози;
- Фактори організаційного характеру;
- Фактори, пов'язані з людським чинником.

Окремо слід виділити ризики, зумовлені людським фактором. Вони включають в себе не тільки помилки співробітників, але і умисні дії, які призводять до поширення конфіденційної інформації.

Відповідно до NIST SP 800-37 Risk Management Framework не слід забувати й про такі категорії ризику як:

- Фінансові ризики;
- Юридичні ризики;
- Бізнес-ризики;
- Політичні ризики;
- Ризики невідповідності законодавству;
- Проектні ризики;
- Репутаційні ризики;
- Ризики безпеки життєдіяльності;
- Ризики стратегічного планування.

Вони не враховуються в рамках запропонованого підходу проте становлять важливу частину будь якого процесу ризик-менеджменту [12].

Респондентам ставляться різноманітні запитання, які використовують шкалу Лайкерта (Likert scale) від 1 до 5. З метою підвищення ефективності та звуження градації можливих результатів для оцінки факторів ризику обрано саме 5-бальну шкалу , в якій показник «неважливо» дорівнює одиниці, а «надзвичайно важливо» дорівнює п'яти. Подібним чином виділено п'ять категорій для оцінки контролів безпеки, так, що показник «ніколи» дорівнює одиниці, а показник «завжди» дорівнює п'яти. Таким чином, усі питання щодо факторів ризику в розподілених системах вимірюються за п'ятибальною від «незначущого» до «надзвичайно важливого», а всі контролі безпеки – від «ніколи» до «завжди». Шкалу Лайкерта досить легко побудувати, вона забезпечує відносну надійність навіть при невеликій кількості суджень, при цьому отримані дані легко обробляти. Відбір суджень для шкали проводиться на основі аналізу літературних джерел в предметній області та в процесі пілотного дослідження методом відбору з первинного списку тверджень з найбільшою дискримінуючою здатністю щодо вимірюваної установки. Для цього створюється початковий перелік тверджень, які пропонуються респондентам з групи, репрезентативної по відношенню до досліджуваної аудиторії (учасникам пілотного дослідження).

Таким чином, при роботі зі шкалою респонденти оцінюють ступінь своєї згоди або незгоди з кожним із запропонованих суджень, від «повністю згоден» до «повністю не згоден» (Табл. 1).

Таблиця 1

Шкала вимірювання критичності факторів ризику та заходів контролю безпеки

Оцінка	**Фактор ризику**	**Контроль безпеки**
1	Неважливо	Ніколи
2	Трохи важливо	Рідко
3	Важливо	Іноді
4	Дуже важливо	Часто
5	Критично	Завжди

Попередня обробка даних. На цьому етапі здійснюється:

Видалення відповідей із високою варіативністю або підозрілими шаблонами, ідентифікація та видалення аномальних або відсутніх значень (методами середнього заповнення, лінійної інтерполяція тощо).

- Нормалізація значень та приведення шкал до єдиного формату для забезпечення коректності подальшого статистичного аналізу.

Статистичний аналіз та ранжування. Після збору та очистки даних проводиться:

- Розрахунок **середніх значень** кожного фактора ризику та контрольного механізму, як середнього значення оцінки кожного показника.
- Визначення **стандартного відхилення** як показника дисперсії, що дозволяє зрозуміти варіативність думок експертів.
- Формування **ранжованого списку** факторів ризику за рівнем загрози та контрольних механізмів за рівнем ефективності при впровадженні.

На даному етапі відбувається визначення рівня важливості факторів ризику та контролів безпеки в життєвому циклі сучасних розподілених інформаційних систем. Нескориговане стандартне відхилення (standard deviation) вибірки S обчислюється (1) для кожної групи факторів ризику.

$$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{x})^2} \quad (1)$$

де $\{x_1, x_2, \dots, x_n\}$ середні значення елементів вибірки, n – розмір вибірки (кількість факторів у кожній групі), та $\tilde{x}$ – середнє значення цієї оцінки (2).

$$\tilde{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (2)$$

Моделювання взаємозв'язків між факторами ризику. Щоб ідентифікувати залежності між факторами ризику, виконується:

1. **Перевірка нормальності розподілу**:
   - Критерії Колмогорова-Смірнова та Шапіро-Вілка.
   - Побудова гістограм частотного розподілу.
2. **Кореляційний аналіз**:
   - Розрахунок коефіцієнтів кореляції (Пірсона, Спірмена, Кендалла тощо).
   - Визначення сили взаємозв'язку (слабкий, середній, сильний).
   - Аналіз значущості залежностей.

На даному етапі виконується перевірка гіпотези про зв'язки між ключовими факторами ризику з використанням коефіцієнтів кореляції. Коефіцієнт кореляції є статистичним показником ймовірності зв'язку між двома змінними, вимірюваними в кількісній шкалі, що дозволяє відповісти на питання про ступінь і напрямок зв'язку між значеннями цих змінних.

Важливим аспектом проектованого профілю ключових факторів ризику кібербезпеки є можливість його застосування для підвищення ефективності інтелектуальних методів оцінки ризиків ІБ, а також покращення та оптимізації процесу вибору вхідного вектору ознак нейромережевих моделей оцінювання ризиків. Перевірка кореляції між вхідними параметрами нейромережевої моделі є критично важливим кроком у науковому дослідженні, оскільки корельовані ознаки можуть впливати на продуктивність моделі та її інтерпретованість. В рамках проектованих моделей оцінки ризику кібербезпеки, що детально описані в дослідженні [13] автора, перевірка кореляції між вхідними параметрами дозволить забезпечити:

- Усунення мультиколінеарності. Якщо вхідні параметри сильно корельовані між собою, це може спричинити мультиколінеарність – ситуацію, коли модель не може чітко визначити вплив кожної змінної на вихідний результат.
- Покращення узагальнюючої здатності моделей. Зменшення кількості надмірно корельованих ознак допомагає уникнути перенавчання. Здатність нейромережевих моделей запам'ятовувати залежності між подібними ознаками, призводить до погіршення узагальнення на нових даних. Видалення або об'єднання корельованих ознак (наприклад, за допомогою PCA) може покращити продуктивність моделі.
- Оптимізацію процесу навчання. Зайві корельовані параметри збільшують вимоги до обчислювальних ресурсів, оскільки модель витрачає час на навчання з дубльованими або надлишковими ознаками.
- Поліпшення інтерпретованості результатів. Видалення або трансформація корельованих ознак робить модель більш зрозумілою, а її результати більш інтерпретованими, особливо в контексті задачі оцінки ризику ІБ.

Для вибору правильного методу

кореляційного аналізу необхідно відповісти на питання чи нормально розподілені досліджувані фактори.

Висувається дві гіпотези (3) для перевірки:

- Нульова гіпотеза (H_0): дані підпадають під зазначений розподіл.
- Альтернативна гіпотеза (H_1): принаймні одне значення не відповідає вказаному розподілу.

$$H_0: P = P_0, H_1: P \neq P_0 \quad (3)$$

де P – розподіл тестової вибірки та P_0– нормальний розподіл.

Перевірка гіпотез відбувається за результатами тестів Колмогорова-Смірнова (Kolmogorov-Smirnov Test, K-S Test) та Шапіро-Вілка (Shapiro-Wilk Test). Якщо значення значимості менше 0,05 – це означає, що дані не мають нормального розподілу. Отже, нульова гіпотеза про те, що дані розподілені нормально має відхилятись.

В залежності від розмірів досліджуваної вибірки даних та результатів перевірки її розподілу приймається рішення про вибір методу кореляційного дослідження. Коефіцієнт рангової кореляції Спірмена (r-Спірмена) застосовується для оцінки сили та напрямку монотонного зв'язку між двома змінними у випадках коли вибірка має невеликий обсяг, не відповідає нормальному розподілу, а дані представлені у вигляді рангів або порядкових шкал. Коефіцієнт кореляції Пірсона використовується для вимірювання лінійного зв'язку між двома змінними, та підходить для даних, що мають нормальний розподіл і лінійний зв'язок. Коефіцієнт кореляції Кендалла (tau-b Кендалла) використовується для оцінки сили та напрямку асоціації між двома змінними у випадках невеликих вибірок, коли необхідна менша чутливість до викидів у даних.

Інтерпретація коефіцієнта кореляції (на прикладі рангової кореляції r-Спірмена) проводиться виходячи з рівня сили зв'язку:

r> 0,01≤0,29 – слабкий позитивний зв'язок;

r> 0,30≤0,69 – помірний позитивний зв'язок;

r> 0,70≤1,00 – сильний позитивний зв'язок;

r> -0,01≤-0,29 – слабкий негативний зв'язок;

r> -0,30≤-0,69 – помірний негативний зв'язок;

r> -0,70≤-1,00 – сильний негативний зв'язок.

Інтерпретація рівня значимості (p-рівня) коефіцієнта кореляції проводиться аналогічно тому, як це робиться для параметричних і непараметричних критеріїв:

якщо p-рівень≤0,05, то зв'язок між змінними є статистично значущим;

якщо p-рівень>0,05, то зв'язок між змінними є статистично незначущим.

Також при інтерпретації p-рівня коефіцієнта кореляції важливим є не тільки сам факт значимості, а й її рівень. Традиційно p-рівень кореляції диференціюється на три рівні:

p≤0,05>0,01 – низька статистична значимість (одна зірочка – *);

p≤0,01>0,001 – середня сила статистичної значимості (дві зірочки – **);

p≤0,001 – висока статистична значимість (три зірочки – ***).

Приймаючи рішення про вибір типу кореляції, при інтерпретації результатів важливо пам'ятати і враховувати, що лінійні кореляції є більш точними, ніж рангові. Ранжування значень при використанні r-Спірмена природним чином знижує міру індивідуальної мінливості виміряного показника.

Результати цього етапу дозволяють визначити взаємопов’язані фактори ризику та розробити рекомендації щодо посилення заходів безпеки.

Відповідно до практичної перевірки запропонованого підходу, проведеної в рамках дослідження [14] автора, на першому етапі створено анкети для опитування, в які включено 40 питань по основним факторам ризику та 14 питань щодо практики застосування контролів безпеки в умовах реальних проектів.

Дослідження демонструє 40 основних факторів ризику в сучасних розподілених інформаційних системах, промарковані від Factor_1 до Factor_40, що досить широко поширені в профільній літературі, часто зустрічаються на практиці та використовуються дослідниками і експертами в області кібербезпеки при вивченні факторів ризику та проведенні заходів ризик-менеджменту. Ці фактори повинні бути ідентифіковані в процесі оцінки та управління ризиками ІБ та контрольовані в подальшому.

Таблиця 2 ілюструє перелік топ-10 ключових факторів ризику розподілених інформаційних систем за результатами даного дослідження.

Таблиця 2

Десять основних факторів ризику для розподілених інформаційних систем

№	N	Mean	Std. Deviation	Відсоток (%)
Factor_21	23	4.391304	0.656376	87.8260
Factor_20	23	4.304348	0.634950	86.0869
Factor_6	23	4.217391	0.735868	84.3478
Factor_10	23	4.173913	0.650327	83.4782
Factor_11	23	4.130435	0.625543	82.6087
Factor_12	23	4.086957	0.668312	81.7391
Factor_33	23	4.000000	0.738549	80
Factor_27	23	3.956522	0.824525	79.1304
Factor_14	23	3.913043	0.792754	78.2608
Factor_8	23	3.869565	0.694416	77.3913

Таким чином, практично всі респонденти виділили фактори, пов'язані з відсутністю політики кібербезпеки та механізмів захисту від мережевих атак, порушеннями автентифікації та управління сеансами, порушеннями контролю доступу та використанням компонентів з відомими вразливостями як найбільш важливі для розподіленого середовища сучасних РІС.

Окрім цього експертній групі були запропоновані можливі категорії заходів по мінімізації ризиків ІБ, що включають в себе організаційно-правовий захист інформації, інженерно-технічні та програмно-апаратні засоби захисту, криптографічні механізми захисту інформації, а також організаційні заходи і заходи фізичного захисту.

Основою для розроблених анкет по можливим контролям та заходам ризик-менеджменту виступив стандарт ISO/IEC 27001 та безпосередньо Додаток А (Appendix A) даного стандарту, що становить собою важливий інструмент для управління інформаційною безпекою [15]. Він містить список заходів безпеки, які повинні застосовуватись для підвищення захищеності інформації, та за своєю структурою сумарно вміщує перелік із 114 контролів безпеки, поділених на 14 розділів. Не всі з цих контролів є обов'язковими для впровадження – компанія може вибрати самостійно, які елементи управління вона вважає застосовними при даних обставинах і в залежності від напрямку бізнесу, стану інфраструктури чи існуючого профілю зовнішніх загроз, та в подальшому реалізувати їх (зазвичай розглядаються по крайній мірі 90% контролів). Більш детальний опис кожного елементу керування додатку А з поясненням того, як його необхідно застосовувати подано в стандарті ISO/IEC 27002. Проте останній не надає ніяких пояснень та підказок як обирати контролі в тій чи іншій ситуації, які елементи керування впроваджувати, як їх вимірювати та як розподіляти обов'язки [16].

В рамках вищезазначеного опитування респондентам було запропоновано оцінити 14 основних груп контролів інформаційної безпеки сучасних РІС, промарковані від Control_1 до Control_14, з точки зору частоти та ефективності їх застосування.

У Таблиці 3 наведено середнє значення та середньоквадратичне відхилення для кожної групи контролів безпеки. Аналіз найбільш поширених категорій механізмів та заходів ризик менеджменту, що застосовуються в сучасних розподілених системах показав, що більшість контролів захисту використовуються часто та є важливими механізмами попередження та мінімізації потенційних ризиків.

Таблиця 3

Середнє значення для кожного контролю безпеки

№	N	Mean	Std. Deviation	Відсоток (%)
Control_1	23	4.260870	0.619192	85.2174
Control_2	23	3.391304	0.940944	67.82608
Control_3	23	2.347826	1.070628	46.95652
Control_4	23	3.782609	0.735868	75.65218
Control_5	23	4.304348	0.634950	86.08696
Control_6	23	4.478261	0.593109	89.56522
Control_7	23	4.478261	0.665348	89.56522
Control_8	23	4.260870	0.619192	85.2174
Control_9	23	4.130435	0.625543	82.6087
Control_10	23	3.000000	1.044466	60
Control_11	23	2.086957	0.900154	41.73914
Control_12	23	2.130435	0.757049	42.6087
Control_13	23	2.478261	0.845822	49.56522
Control_14	23	2.782609	0.951388	55.65218

Узагальнення відповідей опитування, щодо основних груп контролів інформаційної безпеки сучасних розподілених інформаційних системах з точки зору частоти та ефективності їх застосування, продемонструвало що більша частина опитаних виділили контролі, що відповідають за належне та ефективне використання криптографії та інфраструктури відкритих ключів (PKI), логічний та фізичний контроль доступу, операційну безпеку та дотримання політик інформаційної безпеки як найважливіші та найбільш поширені в практиці застосування [17].

Для проведення статистичного аналізу та моделювання даних в рамках вищезазначеного дослідження використано інструментарій програмного забезпечення IBM SPSS Statistics.

Незважаючи на те, що побудовані частотні гістограми на перший погляд достатньо симетричні та добре описуються параболічною кривою для обох тестів Колмогорова-Смірнова та Шапіро-Вілка значення значимості менше 0,05, що означає, що дані не мають нормального розподілу, а отже нульова гіпотеза про те, що дані розподілені нормально відхиляється.

Оскільки об'єм досліджуваної вибірки невеликий (n<30), всі дані мають ранговий характер (представлені у вигляді рангів або порядкових шкал) та розподіл їх значень не відповідає нормальному, то приймається рішення про вибір коефіцієнта рангової кореляції r-Спірмена (4).

$$r = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad (4)$$

де $n = 10$ – кількість факторів, d_i – різниця між двома рангами кожного оцінювання.

В результаті кореляційного аналізу було виявлено помірний негативний зв'язок середнього ступеня статистичної значущості між факторами Factor_20 та Factor_8 – r-Спірмена = -0,528 при p≤0,01, а також помірний негативний зв'язок низького ступеня статистичної значущості між факторами Factor_14 та Factor_8 – r-Спірмена = -0,415 при p≤0,05.

Окрім цього, аналізуючи отримані результати можна дійти висновку, що серед досліджуваних факторів ризику існує помірний позитивний зв'язок низького ступеня статистичної значущості для кореляції змінних Factor_10 та Factor_11 – r-Спірмена = 0,423 при p≤0,05.

Таким чином, отримані результати свідчать про те, що фактори ризику мають значний ступінь взаємозв'язку, часто є взаємопов'язаними і діють в комплексі, що обумовлює їхній комбінований вплив на безпеку PIC. Це підкреслює необхідність застосування комплексного та багатокритеріального підходу до їх аналізу, із

врахуванням усіх можливих взаємозалежностей і впливів.

Отже, результати дослідження демонструють, що всі фактори ризику в процесі життєвого циклу сучасної розподіленої системи є надзвичайно важливими та потребують детального аналізу та врахування при побудові профілю потенційних загроз та здійсненні оцінки ризиків інформаційної безпеки.

Практичне застосування методу

З метою обґрунтування методології дослідження та підтвердження ефективності запропонованого рішення в задачах оцінювання ризиків розподіленого середовища проведено застосування розробленого профілю для оптимізації вибору вхідного набору ознак нейромережевих моделей класифікації рівня ризику на основі інтелектуального аналізу великих обсягів гетерогенних параметрів та метрик про стан інфраструктури розподілених систем. Для цього здійснено порівняння ефективності 2 варіантів оптимізації вхідного простору нейромережевих моделей – на основі побудованого профілю ключових факторів ризику, а також на основі методу головних компонент (PCA).

Узагальнення результатів анкетування та кореляційного аналізу, визначення переліку найважливіших факторів ризику та контролів безпеки, а також формування «карти» взаємозв'язків між чинниками ризику слугує основою для подальшого вибору ознак та адаптивного налаштування нейромережевих моделей з урахуванням виявлених закономірностей [18].

Вхідний масив даних для задач моделювання розмірності [219884 x 341] репрезентує дані типової розподіленої інформаційної системи корпоративного рівня, та містить близько 220 тисяч записів про мережеві активи. Початковий набір містить близько 350 параметрів, що являють собою розподілені метадані і метрики інформаційної системи, та комплексно описують поточний стан інформаційного активу. При цьому цільова ознака може бути представлена 5 основними категоріями, що репрезентують відповідний рівень ризику для ІТ-активу – від R_0 для «Інформаційного» рівня ризику, до R_4, що представляє ризик «Критичного» рівня. Виявлено, що розподілені метрики потребують попередньої обробки, нормалізації та додаткового аналізу через високий рівень гетерогенності та наявність незаповнених значень. Для обох варіантів оптимізації сформовані датасети розділені на навчальну (80%) та тестову (20%) вибірки для тренування проектованих моделей класифікації, та оцінки їх якісних характеристик [19].

З метою дослідження впливу архітектури багатошарового перцептрона на здатність інтерпретації прихованих залежностей та взаємозв'язків у вхідних даних та якість вирішення задачі класифікації прийнято рішення провести оцінку ефективності декількох варіантів архітектури. При цьому використовувались наступні конфігурації нейронної мережі:

- 1 прихований шар з 64 нейронами (Dropout 0.3);
- 2 приховані шари з 128 і 32 нейронами в кожному шарі (Dropout 0.3 для кожного);
- 3 приховані шари з 128, 64 та 32 нейронами у кожному шарі відповідно, (Dropout 0.3 для кожного).

Емпірично проведено підбір оптимальних значень гіперпараметрів. Як функцію активації для прихованих шарів було обрано ReLU (Rectified Linear Unit), для вихідного шару – softmax. Окрім цього, встановлено, що найвищу точність ідентифікації забезпечує класифікатор штучної нейронної мережі із наступними параметрами: алгоритм навчання (оптимізатор) – Adam, швидкість навчання – 0,001, розмір пакету (batch) – 32, функція втрат – категоріальна перехресна ентропія (Categorical Cross-Entropy, CCE).

У першому випадку здійснено навчання ряду нейромережевих моделей, в яких для оптимізації вибору вхідного набору ознак та виділення найважливіших вхідних параметрів моделі (найсуттєвіших факторів ризику) використано **спроєктований профіль ключових факторів ризику** сучасних PIC. На основі результатів проведення багатокритеріального аналізу та кореляції даних ознак із побудованим профілем ключових факторів ризику, що був виконаний у попередньому розділі, здійснено оптимізацію простору вхідного вектору розподілених метрик про роботу інформаційних активів PIC, а їх перелік скорочено до 42 інформативних параметрів, що певним чином відображають стан захищеності активу та можуть бути асоційовані з рівнем ризику.

Результати аналізу навчання моделей з різною кількістю шарів на повному датасеті демонструють, що збільшення кількості прихованих шарів до трьох дозволяє досягти найвищої продуктивності за показником середньої точності (average accuracy) за результатами крос-валідації на рівні 94.39%. Слід відзначити, що усі 3 моделі демонструють високу точність, проте вона більшою мірою зумовлена домінуванням класів R_0 і R_1, що підтверджується вищим показником F1-score для відповідних класів на всіх досліджуваних моделях.

У якості контрольного зразка для порівняння, виконано тренування ряду моделей, для яких у якості методу факторного аналізу для зменшення розмірності та вибору релевантних ознак початкового набору даних застосовано **метод головних компонент (PCA)**. Вхідний набір даних розмірності [219884 x 341] після виділення методом головних компонент скорочується до аналізу 40 інформативних параметрів (що демонструє приблизно такий же рівень оптимізації вхідного простору, як і запропоноване рішення на основі профілю ключових факторів ризику, де цей показник становить 42 метрики)

Таблиця 4

Результати навчання моделі з використанням запропонованого підходу

Показник	**MLP (1 hidden layer)**			**MLP (2 hidden layer)**			**MLP (3 hidden layer)**			
	precision	**recall**	**F1**	**precision**	**recall**	**F1**	**precision**	**recall**	**F1**	**support**
R_0	0.97	0.96	0.97	0.97	0.97	0.97	0.97	0.96	0.97	14912
R_1	0.90	0.96	0.93	0.93	0.95	0.94	0.92	0.96	0.94	12577
R_2	0.95	0.89	0.92	0.95	0.96	0.93	0.97	0.91	0.94	9316
R_3	0.95	0.93	0.94	0.94	0.93	0.94	0.93	0.95	0.94	5202
R_4	0.96	0.95	0.95	0.97	0.93	0.95	0.94	0.95	0.94	1970
macro avg	0.95	0.94	0.94	0.95	0.94	0.95	0.94	0.95	0.94	43977
weight avg	0.94	0.94	0.94	0.95	0.95	0.95	0.95	0.95	0.95	43977
test accuracy	94.26%			94.98%			94.82%			
average accuracy	0.9428			0.9430			0.9439			cross-valid.
standard deviation	0.0019			0.0041			0.0022			

Таблиця 5

Результати навчання контрольної моделі з використанням PCA

Показник	**MLP (1 hidden layer)**			**MLP (2 hidden layer)**			**MLP (3 hidden layer)**			
	precision	**recall**	**F1**	**precision**	**recall**	**F1**	**precision**	**recall**	**F1**	**support**
R_0	0.94	0.95	0.94	0.94	0.95	0.94	0.94	0.94	0.94	14912
R_1	0.82	0.92	0.87	0.83	0.92	0.87	0.83	0.92	0.87	12577
R_2	0.91	0.81	0.86	0.92	0.82	0.86	0.91	0.83	0.87	9316
R_3	0.95	0.87	0.91	0.95	0.87	0.91	0.97	0.88	0.92	5202
R_4	0.93	0.84	0.88	0.89	0.87	0.88	0.92	0.85	0.89	1970
macro avg	0.91	0.88	0.89	0.91	0.89	0.89	0.92	0.89	0.90	43977
weight avg	0.90	0.90	0.90	0.90	0.90	0.90	0.91	0.90	0.90	43977
test accuracy	89.64%			89.93%			90.29%			
average accuracy	0.8978			0.8999			0.8996			cross-valid.
standard deviation	0.0018			0.0023			0.0017			

Побудована модель за результатами крос-валідації демонструє дещо кращі показники ефективності для архітектур з двома та трьома прихованими шарами, причому останній варіант має кращу стабільність та нижчу чутливість до варіативності вхідних параметрів за рахунок найнижчого показника середьоквадратичного відхилення (0.0017). Проте загальні показники точності класифікації приблизно на 4% нижчі ніж запропоноване рішення на основі побудованого профілю ключових факторів ризику. Окрім цього критерій F1-score також демонструє гіршу динаміку у порівнянні з моделлю із застосуванням запропонованого підходу на основі розробленого профілю.

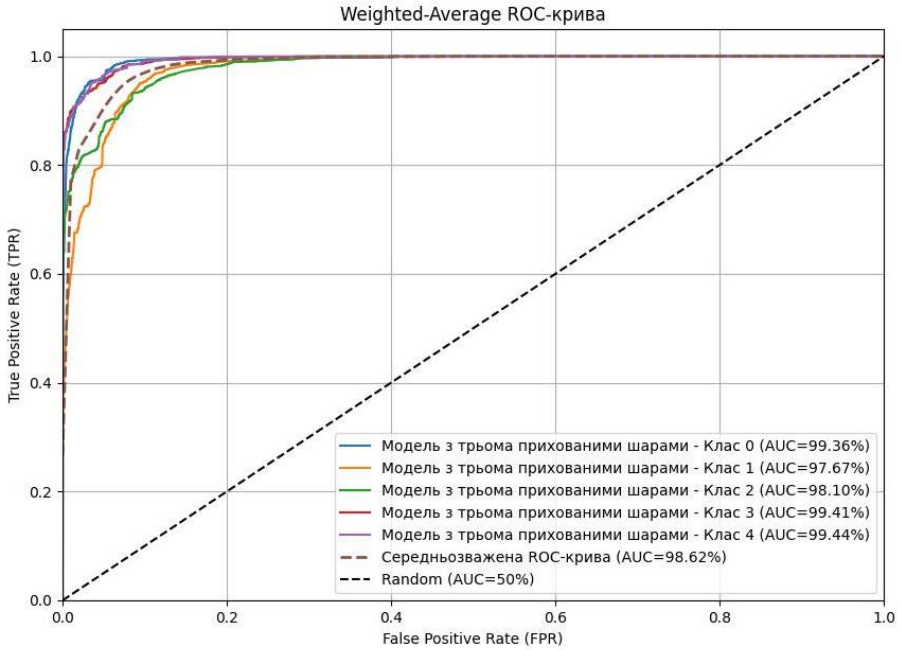

Рисунок 4 Візуалізація кривої ROC-AUC (MLP+PCA)

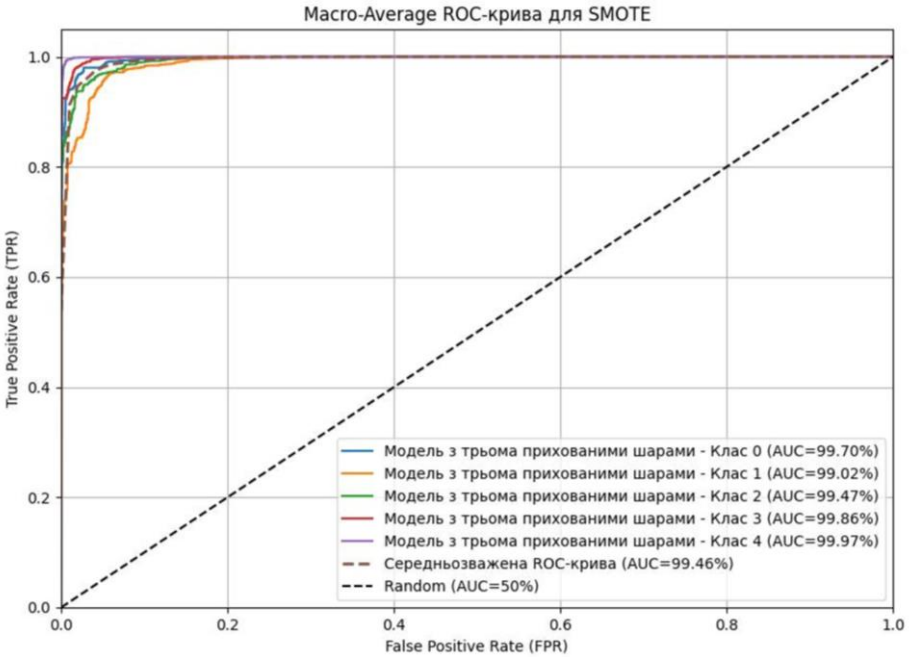

Рисунок 5 Візуалізація кривої ROC-AUC (MLP+запропонований підхід)

Висновки

Отже, в рамках дослідження запропоновано метод побудови профілю ключових факторів ризику кібербезпеки для сучасних розподілених інформаційних систем з використанням інструментарію математичної статистики, що передбачає проведення кореляційного аналізу та моделювання їх взаємозв'язків. В основі запропонованого підходу лежить аналіз результатів експертного опитування із застосуванням багатоступеневого статистичного аналізу, що дозволяє не лише ідентифікувати найважливіші фактори ризику, а й встановити їх взаємозв'язки, що особливо важливо для формування ефективних заходів кіберзахисту. Також визначено та структуровано основні заходи та контролі інформаційної безпеки, які демонструють найкращі показники ефективності в умовах розподіленості середовища, враховують як технологічні, так і організаційні аспекти, забезпечуючи системний підхід до управління ризиками ІБ, зменшення впливу загроз і підвищення стійкості розподілених систем до можливих атак. У ході дослідження здійснено аналіз та ієрархічне впорядкування 40 основних факторів ризику, характерних для типової розподіленої інформаційної системи, та 14 категорій контролів безпеки, оцінених за критеріями їх значущості та частоти виникнення у практичних сценаріях. Запропонований підхід до оптимізації вибору вхідного набору ознак та виокремлення найбільш вагомих факторів ризику, побудований на основі розробленого профілю ключових факторів ризику для сучасних РІС, продемонстрував тотожні результати щодо кількості відібраних метрик у порівнянні з факторним аналізом за допомогою методу головних компонент – 42 метрики у порівнянні із 40 для PCA. Водночас, розроблений підхід забезпечив покращення загальних показників точності класифікації для проектованих моделей оцінки ризику кібербезпеки в РІС на 4% у порівнянні з контрольною моделлю, що використовувала метод головних компонент (PCA) у якості підходу до факторного аналізу, що підтверджує його ефективність у контексті адаптивного аналізу ризиків у розподілених середовищах.

Розроблений метод побудови профілю ключових факторів ризику для РІС дозволяє:

• Сформувати чітку класифікацію факторів ризику розподіленого середовища з урахуванням їх статистичної значущості.

• Ідентифікувати найбільш важливі контрольні заходи для сучасних РІС.

• Визначити взаємозв'язки між факторами ризику та оцінити їх вплив на кібербезпеку.

• Оптимізувати процес управління кіберризиками у РІС шляхом обґрунтованого вибору пріоритетних заходів безпеки.

Таким чином, завдяки запропонованому підходу можна забезпечити більш точну та гнучку систему аналізу ризиків, що дозволить організаціям своєчасно ідентифікувати потенційні загрози, розробляти ефективні заходи з реагування та покращувати загальний рівень кіберстійкості РІС. Практичне застосування запропонованого методу може сприяти вдосконаленню процесу ризик-менеджменту та підвищенню рівня кібербезпеки у складних та динамічних розподілених середовищах.

Література

[1] Andrew S. Tanenbaum, Maarten Van Steen Distributed Systems: Principles and Paradigms, Prentice Hall of India; 2nd edition (January 1, 2007)

[2] The State of Cybersecurity 2022 Report, Global Update on Workforce Efforts, Resources and Cyberoperations. ISACA [Електронний ресурс]. – Режим доступу до ресурсу: https://www.isaca.org/resources/reports/state-of-cybersecurity-2022.

[3] Cybersecurity Assessment Report 2024. Bitdefender [Електронний ресурс]. – Режим доступу до ресурсу: https://www.bitdefender.com/content/dam/bitdefender/business/campaign/2024-Assessment-Report.pdf

[4] Henry K. Risk management and analysis / Kevin Henry // Information Security Management Handbook / Edited by Harold F. Tipton, Micki Krauze. - 6th edition. - Boca Raton: Auerbach Publications, 2017. - Part 1, Section 1.4, Ch. 28. - P. 321-329.

[5] Rot A. IT Risk Assessment: Quantitative and Qualitative Approach // Proceedings of the World Congress on Engineering and Computer Science, 2008. - p. 1073-1078.

[6] Dmytro Palko, Vira Vialkova, Tetiana Babenko «Intellectual models for cyber security risk assessment» // Processing, transmission and security of information : Monografia. Tom 2. / Akademia Techniczno- Humanistyczna w Bielsku-Białej. –Bielsku-Biała : Wydawnictwo Naukowe Akademii Techniczno- Humanistycznej w Bielsku-Białej, 2019. – S. 284–288.

[7] Chang, L.-Y. Applying fuzzy expert system to information security risk Assessment - A case study on an attendance system [Text] / L.-Y. Chang, Z.-J. Lee // 2013 International Conference

on Fuzzy Theory and Its Applications (iFUZZY). - 2013. doi: 10.1109/ifuzzy.2013.6825462

[8] Xin Y. et al. Machine learning and deep learning methods for cybersecurity //IEEE access. – 2018. – Vol. 6. – P. 35365-35381.

[9] State of Enterprise Risk Management 2020 Survey // ISACA, CMMI Institute. - 2019. [Електронний ресурс]. – Режим доступу до ресурсу: https://www.isaca.org/-/media/info/state-of-enterprise-risk-management-survey/index.html

[10] Dmitry Palko, Tetiana Babenko, Larysa Myrutenko, Andrii Bigdan «Model of information security critical incident risk assessment» // Proceedings of the 2020 IEEE International Conference «Problems of infocommunications. Science and technology» PIC S&T'2020, 6-9 October 2020, Kharkiv, Ukraine DOI: 10.1109/PICST51311.2020.9468107.

[11] Johora, F. T., Khan, M. S. I., Kanon, E., Rony, M. A. T., Zubair, M., & Sarker, I. H. (2024). A Data-Driven Predictive Analysis on Cyber Security Threats with Key Risk Factors. arXiv preprint arXiv:2404.00068.

[12] NIST Special Publication 800-30 Rev A. Risk Management Guide for Information Technology Systems, Gary Stoneburner, Alice Goguen, and Alexis Feringa, July 2002.

[13] Palko D, Babenko T, Bigdan A, Kiktev N, Hutsol T, Kuboń M, Hnatiienko H, Tabor S, Gorbovy O, Borusiewicz A. Cyber Security Risk Modeling in Distributed Information Systems. Applied Sciences. 2023; 13(4):2393. https://doi.org/10.3390/app13042393

[14] Dmytro Palko, Hrygorii Hnatienko, Tetiana Babenko, Andrii Bigdan «Determining Key Risks for Modern Distributed Information Systems» // IntSol-2021 Intelligent Solutions, September 28–30, 2021, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine.

[15] ISO/IEC 27001:2022. Information technology - Security techniques - Information security management systems - Requirements. 2022.

[16] ISO/IEC 27002:2022. Information technology - Security techniques - Code of practice for information security controls. 2022

[17] ISO/IEC 27005:2022. Information technology - Security techniques - Information security risk management. 2022.

[18] Haykin S. Neural networks / S. Haykin. − W.: Williams, 2006. − 1104 p. – Режим доступу до ресурсу: https://cours.etsmtl.ca/sys843/REFS/Books/ebook_Haykin09.pdf.

[19] Rassel S. Artificial Intelligence: Modern approach / S. Rassel, P. Norvig. − W.: Williams, 2005. − 1424 p. – Режим доступу до ресурсу: https://www.twirpx.com/file/1626837/.

METHOD FOR BUILDING A KEY CYBERSECURITY RISK FACTORS PROFILE OF MODERN DISTRIBUTED INFORMATION SYSTEMS

The assessment and analysis of cybersecurity risks are fundamental aspects of developing a reliable and effective information security management system, especially in the context of rapid technological advancements and the increasing complexity of modern distributed information systems. Traditional risk assessment methods, which are primarily based on conceptual approaches and classical techniques, have several limitations and prove to be inefficient in large-scale distributed systems. These methods fail to account for the dynamic nature of the environment and do not provide an effective analysis of interdependencies between numerous risk factors. This study proposes a method for constructing a profile of key risk factors in modern distributed information systems based on correlation analysis and modeling of their interrelationships. This approach enhances the efficiency of cybersecurity risk assessment in dynamic environments. Additionally, the proposed method was used to develop a profile of key risk factors for modern distributed systems, analyze their statistical significance and correlation, and identify and structure priority information security measures and controls, which demonstrate high efficiency in distributed environments, considering both technological and organizational aspects, ensure a systematic approach to information security risk management, reduce the impact of threats, and enhance the resilience of distributed systems against potential attacks. The proposed approach to optimizing the selection of input features and identifying the most significant risk factors, based on the developed risk factor profile for modern distributed information systems, demonstrated comparable numerical results with factor analysis using the principal component analysis (PCA) – method 42 selected metrics versus 40 for PCA. However, it provided a 4% improvement in overall classification accuracy for the designed cybersecurity risk assessment models in DIS compared to the PCA-based control model. This confirms its effectiveness in the context of adaptive risk analysis in distributed environments.

Keywords: information security, information security risk, risk factors, risk assessment, risk management, risk evaluation, distributed information system, neural network.

Палко Дмитро Володимирович, аспірант кафедри кібербезпеки та захисту інформації, Факультет інформаційних технологій, Київський національний університет імені Тараса Шевченка, Київ, Україна.
Dmytro Palko, PhD-student, Department of Cyber Security and Information Protection Faculty of information technology, Taras Shevchenko National University of Kyiv, Ukraine.
E-mail: palko.dmytro@gmail.com.
Orcid ID: 0000-0002-2886-1975.

Мирутенко Лариса Вікторівна, к.т.н., доцент, доцент кафедри кібербезпеки та захисту інформації, факультет інформаційних технологій, Київський національний університет імені Тараса Шевченка, Київ, Україна
Larysa Myrutenko, Candidate of Technical Sciences, Associate Professor of the Department of Cyber Security and Information Protection Faculty of information technology, Taras Shevchenko National University of Kyiv, Ukraine.
E-mail: myrutenko.lara@gmail.com.
Orcid ID: 0000-0003-1686-261X.

УДК 004.056.5

АДАПТИВНЕ ПЕРЕШКОДОПОДАВЛЕННЯ У БЕЗПРОВОДОВИХ МЕРЕЖАХ НА ОСНОВІ ШТУЧНОГО ШУМУ

Станіслава Кудренко, Валерій Козловський, Анна Столяр

Адаптивне перешкодоподавлення на основі штучного шуму є перспективним підходом до підвищення безпеки безпроводових мереж. У традиційних системах криптографічного захисту зловмисники можуть використовувати атаки на фізичному рівні для перехоплення сигналу. Одним із ефективних методів захисту є використання штучного шуму (Artificial Noise, AN), який генерує спеціально створені завади для ускладнення несанкціонованого доступу. У статті досліджено принципи адаптивного контролю потужності штучного шуму на основі градієнтного спуску. Запропонований підхід дозволяє динамічно регулювати рівень шуму залежно від параметрів каналу зв'язку, таких як відстань до зловмисника, рівень сигналу та наявність завад. Для моделювання було використано дані безпроводової сенсорної мережі (WSN) з відкритого репозиторію, що дозволило оцінити вплив адаптивного шуму на втрати пакетів і рівень сигналу. Отримані результати показали, що оптимізований метод дозволяє ефективно знижувати ймовірність перехоплення без значного погіршення якості зв'язку для легітимного користувача. Запропонована модель може бути застосована у сучасних системах мобільного зв'язку, мережах IoT, а також у критичних інфраструктурах, що потребують підвищеного рівня захисту даних.
Ключові слова: *перешкодоподавлення, штучний шум, безпроводові сенсорні мережі.*

Вступ

Адаптивне перешкодоподавлення на основі штучного шуму є актуальним напрямом у безпроводових мережах, спрямованим на підвищення захисту даних та покращення якості зв'язку. Використання штучного шуму (Artificial Noise, AN) дозволяє ефективно знижувати ймовірність прослуховування даних сторонніми користувачами (атаки перехоплення) та мінімізувати вплив завад на основний сигнал [1]. Сучасні дослідження у цій сфері включають методи адаптивного управління шумом, динамічного розподілу ресурсів та впровадження технологій безпеки рівня фізичного шару.

Розширення мереж 5G і майбутніх 6G-систем, зростаюча кількість пристроїв Інтернету речей (IoT) та розвиток інфраструктури мобільного зв'язку призводять до зростання потенційних загроз у цифровому середовищі. У той час як традиційні методи криптографічного захисту залишаються основою безпеки переданих даних, вони виявляються вразливими перед фізичними атаками, що спрямовані на безпосереднє перехоплення радіосигналу або його аналіз у реальному часі.[4,5] Зокрема, перехоплення безпроводових сигналів стало однією з ключових загроз, що дозволяє зловмисникам отримати доступ до чутливої інформації без необхідності зламувати криптографічні алгоритми. Одним із поширених методів є атаки «чоловік посередині» (Man-in-the-Middle, MitM), коли зловмисник вставляється між передавачем і приймачем, підміняючи або змінюючи інформацію[5]. Інший метод – аналіз трафіку, що дозволяє визначити структуру переданих даних, навіть якщо вони зашифровані. Окрім цього, новітні методи атаки використовують штучний інтелект для прогнозування і відновлення сигналів навіть у зашифрованих мережах.

У відповідь на ці виклики, дослідники