УДК 004

ОБҐРУНТУВАННЯ ДОЦІЛЬНОСТІ ВИКОРИСТАННЯ R-ПРОГРАМУВАННЯ ДЛЯ ОЦІНКИ РИЗИКІВ

М. Д. Синицін[1], Ю. Г. Даник[1]

[1] *Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Фізико-технічний інститут*

Анотація

В ході роботи було розглянуто класифікацію моделей ризику. Проаналізовано можливість використання мови, призначеної для статистичної обробки даних, в аналізі ризиків. Обґрунтовано на прикладі роботи аналітика доцільність використання R-програмування для реалізації аналізу ризиків методом Монте-Карло.

Ключові слова: ризик, модель ризику, R-програмування

Вступ

З розвитком технологій та активною автоматизацію робочих процесів на об'єктах також зростає увага до питань безпеки. На відміну від традиційних інформаційних систем, в автоматизованих системах управління технологічними процесами, які застосовують, наприклад на об'єктах критичної інфраструктури, існує досить тісний взаємозв'язок між цими автоматизованими системами та фізичними процесами і виконавчими пристроями. Тому, порушення кібербезпеки таких систем може призвести до негативних, важко прогнозованих наслідків. Прикладами наслідків, від реалізації кіберзагроз, можуть слугувати фінансові втрати або можливість втрати репутації, а також, у гіршому випадку завдання шкоди життю та здоров'ю людини, або підвищення ризиків виникнення екологічних катастроф. У цих умовах надзвичайно важливу роль відіграє всебічне забезпечення безпеки, у тому числі і кібербезпеки. Одним із ключових процесів при управлінні та забезпеченні кібербезпеки є процес оцінки ризиків.

1. Оцінка ризиків

1.1. Аналіз літературних джерел

На сьогодні існує чимало публікацій вітчизняних та закордонних вчених, в яких розглядаються дослідження пов'язані з оцінкою ризику реалізації різних загроз та використання для цього різноманітних методів, інструментів та програмних продуктів. Наприклад у роботі [1] модель оцінки ризиків запропонована з використанням апарату нечіткої логіки. Дана модель враховує чотири фактори ризику: уразливість, загроза, ймовірність та вплив. Однак існує і інший підхід, описаний роботою [2], в якій досліджено широкий спектр загроз, які мають високий рівень ризиків та шляхи реалізації цих загроз, створено базу даних фактичних втрат у випадку реалізації загроз, здійснено аналіз втрат з використанням методів статистики та актуарної математики. Нові ж методи оцінки ризику, шляхом адаптації існуючих існуючих методів для розрахунку ризиків і невизначеностей представлені в роботі [3]

1.2. Класифікація моделей ризику

Оскільки ризики різноманітні, їх можна поділити на безліч груп, тобто класифікувати за різними ознаками, то для того, щоб обрати ефективний метод оцінки ризиків треба визначитися з моделлю породження ризику. Розрізняють такі моделі :

- модель « небезпека–ризи »
- модель « невизначеність–ризик »
- модель « можливості–ризик/шанс »

Модель « небезпека–ризик » полягає в тому, що існування ризикової ситуації ґрунтується наявністю множини небезпек різного походження, які і можуть слугувати причиною виникнення ризикової ситуації. Для моделі характерне існування повного обсягу відомостей, необхідних для обчислення профілю ризиків та середнього ризику загроз, який є інтегральною характеристикою рівня небезпек. Саме існування повного обсягу відомостей достатньо для створення необхідних методик нормалізації стану системи, а також протидії загрозам.

Модель « невизначеність–ризик », на відміну від попередньої моделі, не має чіткої структури множин небезпек, а також майже не відомі види та форми загроз. Головна особливість моделі полягає в тому, що відомими є лише дані про наслідки небезпек, які представлені у формі опису загрози, як можливість або неминучість певної шкоди. Ризикова ситуація, за таких умов, утворюється базуючись на відомості про можливості впливу негативних факторів небезпечних явищ і процесів. Саме аналіз таких даних фактично виключає можливість запобігти розвитку чинних загроз та істотно зменшує способи нейтралізації і зменшення наслідків реалізації загроз. Тоб-

то характерною особливістю концептуальної моделі «невизначеність–ризик» є реальна можливість визначення оптимального виходу з ризикової ситуації.

Розглянемо модель « можливість–ризик/шанс ». Ця модель використовується досить широко через те, що за особливостями практичної діяльності виникає потреба в аналізі множини можливих варіантів дій, подій при прийнятті рішень. Це обумовлено тим, що багато альтернативність ризикової ситуації обумовлюється не тільки впливом небезпек або дією різних факторів, але й створюється людиною штучно, зазвичай з метою пошуку прийнятних рішень в певній ситуації. Досить широко дана модель застосовується для аналізу ризиків, що виникають у фінансовій та економічній сфері тому, що в даній моделі характерні так звані спекулятивні ризики, для яких існує ймовірність одержати як негативні (збитки), так і позитивні (прибутки) результати. Оскільки модель передбачає обробку штучно створених факторів ризику, це зумовлює появу множини варіантів апроксимативних моделей. Тобто ризик або ймовірний результат, від реалізації j-го проєкту можна представити формулою :

$$R_j = \sum_{1}^{n} p_i q_{ij} \quad j = 1, m$$

В даній формулі p_i це елемент вектора $P = \{p_1, p_2, p_3, ..., p_n\}$, що є ймовірністю настання події s_i з кортежу $S = \{s_1, s_2, s_3, ..., s_n\}$, а q_{ij} – прибутки (або втрати) пов'язані із функціонуванням проєкт. Кожному з варіантів властиві свої позитивні та негативні якості, тобто виникає проблема вибору кращого варіанту апроксимативної моделі. Процес вибору кращого варіанту детально описаний в роботах [4] та [5].

Постановка завдання дослідження

З точки зору обробки даних модель « можливість–ризик/шанс » охоплює досить великий обсяг інформації, що значно збільшує час роботи аналітику. Крім того, важливим є зручне й інформативне зображення результатів обробки даних та розрахунків, що не більшість існуючих програмних продуктів для ситуації, що розглядається вирішують недостатньо ефективно.

Однак використання мови програмування R, розробленої у 1993 році Россом Айхэка та Робертом Джентлменом, дає змогу ефективно працювати з таким об'ємом даних.

Метою дослідження обрано обґрунтування можливості використання R-програмування для оцінки ризиків є важливим і актуальним.

2. Використання мови програмування R для оцінки ризиків

2.1. Мова статистичних обчислень R

Мова R – являє собою середовище в котрому є набір програмних пакетів за допомогою яких є можливим проведення обчислень та побудови графіків, а також для маніпуляцій з даними.Тобто R програмування використовують для статистичних обчислень. Ця мова має набір алгоритмів котрі досить широко застосовуються у сфері машинного навчання, а конкретніше — в аналізі часових рядів, класифікацій, кластеризації, лінійному моделюванні, що значно зменшує час та спрощує обробку великих даних. Також одна з найголовніших переваг мови R, в порівнянні з іншими мовами полягає в можливості формування якісної графіки типографського рівня, яка може бути експортована у відомі графічні формати, та використана для публікації або презентації. Саме створення чітких графіків та діаграм, в межах самого коду, дає змогу заощадити час спеціаліста при формуванні звітності [6].

В порівнянні з іншими мовами, наприклад Python, R створювався для аналізу даних, тому його конструкції та синтаксис широкі та зрозумілі, на відміну від Python, який вважається універсальним та багатоцільові, що ускладнює його розуміння. Також мова R має велику кількість додаткових пакетів, що значно збільшує функціонал, а також дозволяє завантажувати у вигляді таблиць дані, запропоновані в більшості відкритих пріоритетних форматах. Так мова дає змогу оброблювати таблиці у звичайній текстовій формі, або таблиці Excel різних версій, так само як і дані в форматах CSV, XML.

2.2. Використання мови R аналітиками

Прикладом використання R-програмування для аналізу ризиків було обрано проєкт аналітика Corey Nesky, який мав за мету перетворити експертні оцінки ризику витоку інформації у звітність для зацікавлених сторін

Аналітик запропонував набір скриптів зібраних у пакет[7], написаних мовою R, в яких вхідними даними було обрано експертні оцінки можливості впливу ризику витоку інформації на збитки. Саме використання оцінок, в якості параметрів, при моделюванні даних, дало змогу аналітику, використовуючи написаний код, отримати ймовірнісні твердження для узагальнення отриманих розподілів запропонованих експертами.

Для аналізу ризиків було обрано метод Монте-Карло, який описує усі фактори, котрі впливають на ймовірність та величину ризику, а потім використовує їх для створення тисячі нових випадкових сценаріїв, серед яких обирається найбільш вірогідний. Надалі модель аналізу випадково генерує понад 10 тисяч сценаріїв з діапазону розподілу експертних оцінок, на основі цих даних та повертає аналітику результат. Метод Монте-Карло має низку переваг, в порівнянні з детерміністським аналізом, або аналізом «по точковим оцінкам», а саме:

- **Вірогідні результати.** Результати демонструють не тільки можливі події, а й вірогідність їх появи
- **Аналіз чутливості.** За рідкісним виключенням детерміністський аналіз не дає змогу вирахувати яка зі змінних найбільше впливає на резуль-

тат сценарію. При використанні метода Монте-Карло легко оцінити, які вхідні дані більшою мірою впливають на результати.

- **Аналіз сценаріїв.** Метод дає змогу аналітику чітко визначити які вхідні данні які тим,або іншим чином приводять до певних значень, та простежити появу відповідних наслідків
- **Графічне представлення результатів.** Характер даних, отриманих при використанні метода Монте-Карло дає можливість створювати графіки різних наслідків, а також ймовірностей її появи.

Саме доступність та легкість побудови графіків в мові R значною мірою прискорює час формування звітності аналітиками. Результатом роботи пакету, запропонованого Corey Nesky було отримано графіки, ссилаючись на які за досить короткий термін було сформовано звітність, що до

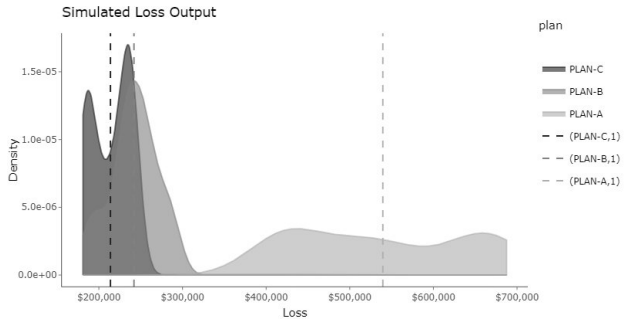

Рис. 1. Графік очікуємого збитку внаслідок можливого витоку інформації за трьома сценаріями

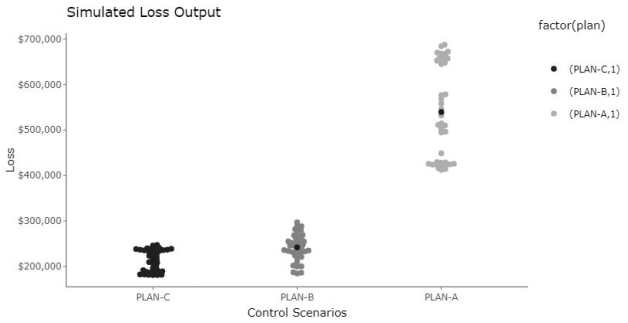

Рис. 2. Графік ймовірносного рзподілу втрат відреалізації трьх сценаріхв ризиків

Висновки

Використання R програмування для аналізу ризиків є доцільним за умови аналізу моделі ризику

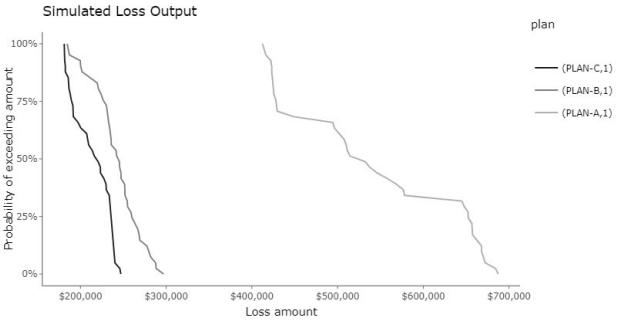

Рис. 3. Графік ймовірності перевищити очікуєму сумму збитку за кожним сценарієм

"можливість-ризик/шанс" через те, що на відміну від конкуруючої мови Python, дана мова програмування має доступний функціонал для роботи та обробки великих даних, та чітко графічно ілюструє результати, що полегшує роботу аналітику, та пришвидшує час формування звітності, а також легша для засвоєння.

Перелік використаних джерел

1. Mansour A. Improving risk assessment model of cyber security using fuzzy logic inference system // Computers and Security. — 2018. — V. 74 — P. 323–339.
2. Eling M., Wirfs J. What are the actual costs of cyber risk events? // European Journal of Operational Research. — 2019. — V. 272 — P. 1109—1119.
3. Radanlieva P., Charles De Rourea D., Nicolescub R., Huthb M., Mantilla Montalvoc R., Cannadyc S., Burnap P. Future developments in cyber risk assessment for the internet of things // Computers in Industry. — 2018. — V. 102 — P. 14—22.
4. Черноруцкий И. Г. Методы принятия решений // БХВ-Петербург. — 2005. — 416с.
5. Катренко А. D., Пасічник В. В., Пасько В. П. Теорія прийняття рішень // Видавнича група BHV. — 2009. — 448с.
6. Мэтлофф Н. The Art of R Programming: A Tour of Statistical Software Design. // Санкт-Петербург. — 2019. — 416с.
7. Nesky C. Security data analyses and quantitative risk assessments to inform executive decisions. // Режим доступа: `https://github.com/theonaunheim/unsuR`.