Математична модель для аналізу інформаційних потоків в соціокіберфізичних системах

УДК 004.056

Наталія Дженюк [1], Максим Толкачов [2]

Національний технічний університет "Харківський політехнічний інститут",
[1]natalidzh16@gmail.com, [2]maksymtolkachov@gmail.com

Природа динамічних фізичних середовищ ставить під сумнів здатність соціокіберфізичних систем до адекватного управління фізичними активами у різноманітних ситуаціях. Отже, проектування соціокіберфізичних систем повинно гарантувати не тільки надійну автономію, але й стійкість у експлуатації та безпеку. Запропонований метод базується на інтеграції специфічних (змішаних) загроз через поєднання технічних кіберзагроз і методів соціальної інженерії.

Аналіз публікацій останніх років дозволяє зробити висновок, що під час створення соціокіберфізичних систем і систем, які забезпечують безпечне їх функціонування, недостатньо досліджені питань, пов'язаних з інформаційними процесами соціальної складової, а також семантичними елементами переданої інформації. Тому вказані питання аналізу сприйняття, оцінки, обміну та обробки інформації є актуальними і пояснюють необхідність розробки методик оцінки цих процесів.

Набори даних із різних сфер зазвичай містять семантичні дані, визначені за широким набором атрибутів, між якими існують різні ступені кореляції.

Ми зосереджуємося на виявленні об'єктів даних, які демонструють аномальну поведінку за підмножиною атрибутів, названих поведінковими, у відношенні до інших, названих контекстними. Основний вклад полягає у розробці моделі для опису законів кореляції, прихованих у розподілах даних між парами поведінкових та контекстних атрибутів. Вводиться ймовірнісний показник, спрямований на оцінку спостережуваних об'єктів на основі того, наскільки їхні дії відхиляються від виявлених законів кореляції [1].

Кореляція між зазначеними атрибутами існує як на рівні топології, так і на рівні семантичного змісту наборів даних, а також у часовому аспекті. Таким чином, результатом цього етапу стане набір кластерів даних, корельованих між собою. Ці кластери вказують на події, які корелюють одна з одною за часовими, семантичними та топологічними показниками.

Математичну модель цього процесу можна розглядати як модель, з використанням якої необхідно визначити маніпулятивні фактори на основі спостережуваних моделей переконань або думок цільової аудиторії [2].

Припустимо, що у нас є потік інформації, що складається з набору N повідомлень. Кожне повідомлення характеризується набором атрибутів, що описують його зміст, контекст, семантичні ознаки. Семантичні ознаки, що вилучені з інформації джерела, дозволяють моделювати особисті характеристики користувача. Кожне повідомлення може бути представлене у вигляді вектора ознак, де кожна ознака – це бінарна змінна, яка вказує на наявність або відсутність певної характеристики.

Також визначимо набір шаблонів M або шаблони, що охоплюють ключові характеристики повідомлень, пов'язаних з певними темами.

Для аналізу ентропії інформаційного потоку та вилучення відповідних даних може бути використаний метод, такий як порівняння зі зразком або кластеризація. Це дозволяє ідентифікувати повідомлення, які відповідають шаблонам з бібліотеки. Бібліотеку шаблонів можна представити у вигляді матриці P, де кожен рядок відповідає певному шаблону, а кожен стовпець – певній функції:

Щоб визначити повідомлення, схожі на шаблони у нашій бібліотеці, ми можемо представити повідомлення у вигляді векторів функцій та обчислити їхню схожість з кожним шаблоном, використовуючи такі міри, як косинусна схожість чи схожість Жаккара. Наприклад, представляючи повідомлення у вигляді вектора M, його косинусну схожість з кожним шаблоном у нашій бібліотеці обчислюється наступним чином:

$$similarity(M, P_i) = \frac{dot(M,P_i)}{(norm(M) \cdot norm(P_i))} \quad (1)$$

де $dot(M, P_i)$ – скалярное произведение векторов M и P_i,
$norm(M)$ и $norm(P_i)$ – евклидовы нормы векторов M и P_i.

Ця міра подібності може бути використана для визначення k найкращих шаблонів, які найкраще відповідають кожному з повідомлень, і використовувати їх для витягування відповідних даних чи ідей. Наприклад, якщо повідомлення найбільше відповідає зразку політики, можна зробити висновок, що воно містить політичний контент, і додати його до бази даних політичних повідомлень.

Таким чином, побудована математична модель дозволяє аналізувати ентропію інформаційних потоків та витягувати відповідні дані з використанням бібліотеки шаблонів, що можуть бути корисні для розуміння динаміки розповсюдження інформації та розробки ефективних заходів втручання або протидій.

Запропонований підхід у майбутньому сприятиме покращенню контролю та управління в галузі кібербезпеки, забезпечуючи врахування різних аспектів, включаючи соціальні та сприйняттєві фактори.

1. Fabrizio Angiulli, Fabio Fassetti, Cristina Serrao, Anomaly detection with correlation laws, Data & Knowledge Engineering, Volume 145, 2023, 102181, ISSN 0169-023X, https://doi.org/10.1016/j.datak.2023.102181.
2. Yevseiev, S., Dzheniuk, N., Tolkachov, M., Milov, O., Voitko, T., Prygara, M., Shpak, O., Voropay, N., Volkov, A., & Lezik, O. (2023). Development of a multi-loop security system of information interactions in socio-cyberphysical systems. Eastern-European Journal of Enterprise Technologies, 5(9 (125), 53–74. https://doi.org/10.15587/1729-4061.2023.289467