DOI 10.62731/mcnd-17.05.2024.007

МОДЕЛЬ ДИНАМІКИ ІНТЕНСИВНОСТІ ІНФОРМАЦІЙНОГО ВПЛИВУ ДЛЯ ВИЯВЛЕННЯ ЦІЛЕСПРЯМОВАНИХ ІНФОРМАЦІЙНИХ АТАК

Косогов Олександр Миколайович
ORCID ID: 0000-0001-6691-273X
канд.військ.наук, ст.наук.спів., доцент кафедри аеродинаміки та безпеки польотів
Національний авіаційний університет, Україна

Управління різними технологічними процесами в авіації базується на використанні інформаційно-телекомунікаційних систем (ІТС), до яких відносяться джерела інформації, засоби її передавання, оброблення, відображення, зберігання, загальносистемне та спеціальне програмне забезпечення. У всіх інформаційних технологічних процесах, а також процесах управління, важливу роль відіграє людський фактор [1].

Функціонування системи цивільної авіації в Україні на сучасному етапі безпосередньо пов'язане, насамперед, із забезпеченням належності та своєчасності інформаційних потоків, впровадженням нових інформаційних технологій, глобалізацією та інтеграцією авіаційних інформаційних систем згідно міжнародних стандартів. При цьому інформаційна безпека виступає домінантною складовою процесів інформаційного забезпечення функціонування системи цивільної авіації.

Існує багато видів інформаційних загроз та атак, а з ними і безліч методів їхньої реалізації. Всі вони несуть значну небезпеку для інформаційних об'єктів авіакомпаній, корпорацій, державних авіаційних установ. Необхідно зосередити увагу на найнебезпечнішій і поширеній інформаційній атаці під назвою: "атака (таргетована атака, APT)". Велика небезпека такої атаки полягає в тому, що вона в собі несе і поєднує безліч різноманітних видів і методів реалізації інформаційних атак.

Цільові атаки - це атаки, спеціально націлені на одну людину, компанію або корпорацію, які проводяться тихо і непомітно [1]. Це не масові атаки, так як їх мета не вразити якомога більше комп'ютерів. Небезпека полягає саме в "замовному" характері такого роду атак, які спеціально розробляються для обману своїх потенційних жертв.

Особливість цілеспрямованих атак (APT) полягає в тому, що зловмисників цікавить конкретна компанія або державна організація. Це відрізняє цю загрозу від масових хакерських атак - коли одночасно атакується велика кількість цілей і найменш захищені користувачі стають жертвою. Цілеспрямовані атаки зазвичай добре сплановані і включають кілька етапів – від розвідки і впровадження до знищення слідів присутності. Зазвичай унаслідок цілеспрямованої атаки зловмисники закріплюються в інфраструктурі жертви та залишаються непоміченими протягом місяців або навіть років - протягом усього цього часу вони мають доступ до всієї корпоративної інформації.

Виявлення цілеспрямованих атак з метою своєчасної протидії їм потребує оперативного аналізу інформаційного простору з використанням спеціалізованих систем моніторингу. Такі системи мають забезпечувати не тільки апаратний аналіз інформаційних атак, а й кількісний аналіз динаміки проявів цих атак з урахуванням їх специфіки. У разі здійснення атаки інтенсивність інцидентів потоку атак, яка являє

собою часовий ряд за кількістю інформаційних інцидентів за певний проміжок часу (як правило, за добу), може містити інформацію як про сам факт APT, так і про фазу сценарію, за яким вона здійснюється.

Завдання з виявлення цієї інформації є доволі складним і містить три складові:

а) вибір (побудова) математичної моделі, що відображає динаміку інцидентів;

б) вибір (розроблення) і застосування відповідного методу оброблення часового ряду;

в) інтерпретація отриманих результатів.

Складність зазначеного завдання обумовлена складністю процесів, які відбуваються в інформаційному просторі, великою кількістю факторів, що визначають ці процеси. Намагання врахувати всі визначальні фактори призводить до необхідності побудови складних математичних моделей інформаційного простору. Втім, ускладнення моделі не дає ніякої впевненості у відповідному зростанні рівня її адекватності. Перевірка моделі, зіставлення її з реальним процесом потребує проведення відповідних натурних експериментів. Проведення ж натурних експериментів, що спрямовані на дослідження соціальних процесів, до яких належать і процеси у інформаційному просторі, стикається, як правило, зі значними обмеженнями щодо управління експериментом та його ресурсного забезпечення.

Крім того, ускладнення математичних моделей звужує можливість знаходження на основі їх використання точних математичних рішень.

Суперечливість між простотою моделі і її адекватністю певною мірою може бути ослаблена шляхом використання стохастичних (імовірнісних) математичних моделей. Адекватність моделі при цьому може бути оцінена опосередковано, виходячи із загальної ефективності вирішення завдання щодо виявлення інформаційних акцій (операцій, кампаній) на основі коректного використання цієї моделі.

У цій роботі пропонується така модель формування тематичного інформаційного потоку, який присутній на вході системи контент-моніторингу.

Інформаційний простір асоціюється з деяким співтовариством агентів, чисельність яких $n(t)$ є функцією часу t. Агенти асоціюються з окремими інформаційними повідомленнями. При цьому припускається, що за певний проміжок часу (наприклад, за добу) кожний агент (випадково), незалежно від інших, з ймовірністю $\vartheta(t)$, $0 \le \vartheta(t) \le 1$, вкидає в інформаційний потік інформаційне повідомленням за даною темою. За такої моделі її параметр $n(t)$ відображає потужність співтовариства агентів, що беруть участь у формуванні тематичного інформаційного потоку, а параметр $\vartheta(t)$ – активність агентів. З огляду на таку інтерпретацію параметрів $n(t)$ і $\vartheta(t)$ природно припустити, що активність $\vartheta(t)$ більш динамічно змінюється в часі, ніж потужність $n(t)$.

Необхідно зауважити, що подібні мультиагентні моделі вже описані у науковій літературі [1].

Неважко побачити, що прийнята модель формування тематичного інформаційного потоку збігається зі схемою проведення експерименту Бернуллі щодо визначення ймовірності виникнення події рівно x разів у n незалежних випробуваннях за умови, що ймовірність події у кожному випробуванні становить ϑ [2]. Тобто інтенсивність вхідного тематичного інформаційного потоку $x(t)$ за прийнятою моделлю являє собою нестаціонарний випадковий процес з біноміальним розподілом імовірностей.

Таким чином, оброблення часового ряду, що відображає динаміку надходження інформаційних інцидентів, полягає в оцінюванні параметрів $n(t)$ і $\vartheta(t)$ нестаціонарного випадкового процесу $x(t)$.

На сьогодні методи оптимального (найкращого за певним критерієм) оцінювання параметрів випадкових процесів достатньо глибоко опрацьовані наукою і висвітлені в науковій літературі. Описана велика кількість модельних задач та відповідних їм оптимальних рішень [2, 3]. Тому в рамках вирішення завдання щодо виявлення інформаційних акцій (операцій, кампаній) за прийнятою моделлю формування тематичного інформаційного потоку достатньо визначити найбільш близьку модельну задачу та використати отримане при її розв'язанні оптимальне рішення.

Виходячи зі специфіки цього завдання, яка полягає у:

1) потребі оперативного виявлення факту і фази інформаційної атаки;

2) потребі забезпечення максимальної достовірності отриманих результатів;

3) неможливості безпосередньо спостерігати випадковий процес, який оцінюється (оцінюються параметри $n(t)$ і $\vartheta(t)$), спостерігається випадковий процес $x(t)$);

доцільно розглядати саме методи оптимального оцінювання параметрів випадкових процесів, що недоступні безпосередньому спостереженню, за вимірювальною інформацією зростаючого обсягу. Свій початок ці методи беруть з праць Р.Е. Калмана, присвячених дискретній та безперервній фільтрації [3]. З огляду на те, що в цьому випадку вхідні дані являють собою часовий ряд, доцільно обмежитись дискретним фільтром Калмана і замість безперервного випадкового процесу $x(t)$ розглядати випадкову послідовність x_k, де $t = k\Delta t$, Δt – деякий проміжок часу.

Необхідно зазначити, що рекурентна форма побудови дискретного фільтра Калмана забезпечує зручність практичної реалізації обчислювальної процедури за допомогою електронно-обчислювальної техніки в реальному масштабі часу. Ця обставина є важливою в плані автоматизації процесів оброблення інформації при здійсненні моніторингу інформаційного простору.

З іншого боку, дискретний фільтр Калмана забезпечує отримання оптимальних за критерієм мінімуму середньоквадратичної похибки оцінок параметрів вхідного випадкового процесу лише за умови, що в наявності повна і точна апріорна інформація про початковий стан досліджуваної системи, її поведінку та про систему спостереження за нею. Як правило, реальні умови функціонування дискретного фільтра Калмана відрізняються від зроблених припущень. Це стосується і вирішуваного завдання. Тому проблема забезпечення повними і точними апріорними даними є однією з основних труднощів при практичному використанні дискретного фільтра Калмана, а всі наступні наукові праці, пов'язані з динамічною рекурентною фільтрацією, були спрямовані на забезпечення адекватності алгоритму фільтрації реальним умовам функціонування за тими чи іншими параметрами [3].

З огляду на умови вирішуваного завдання дискретна модель еволюції системи, що формує тематичний інформаційний потік, може бути записана у вигляді рівняння

$$y_{k+1} = Фy_k + bw_k, \quad (1)$$

де $y_k^T = [n_k \quad \dot{n}_k \quad \vartheta_k \quad \dot{\vartheta}_k \quad \ddot{\vartheta}_k]$ *– розширений вектор стану системи, який, крім поточних на момент часу* $t = k\Delta t$ *значень параметрів випадкової послідовності* x_k *–* n_k *і* ϑ_k*, містить поточні значення різниць цих параметрів за часом* $\dot{n}_k = \frac{n_{k+1}-n_k}{\Delta t}$, $\dot{\vartheta} = \frac{\vartheta_{k+1}-\vartheta_k}{\Delta t}$, $\ddot{\vartheta}(t) = \frac{\dot{\vartheta}_{k+1}-\dot{\vartheta}_k}{\Delta t}$. *Введення до вектора стану систему цих різниць зумовлено необхідністю забезпечення певної динаміки еволюції системи;*

$$\Phi = \begin{bmatrix} 1 & \Delta t & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & \Delta t & (\Delta t)^2/2 \\ 0 & 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$ – матриця переходу;

w_k – шум системи для компенсації похибки моделювання. Передбачається, що w_k є білим гауссовим шумом, має нульове математичне очікування і дисперсію Q_k;

$$b^T = [(\Delta t)^2/2 \quad \Delta t \quad 1]$$ – вектор.

Система (3) може спостерігатися через математичне очікування і дисперсію випадкової послідовності x_k та їх різниці. Для випадку біноміального розподілу ймовірностей математичне очікування і дисперсія визначаються так [2]:

$$Mx_k = m_k = n_k\vartheta_k; \tag{2}$$

$$Dx = D_k = n_k\vartheta_k(1-\vartheta_k) = m_k(1-\vartheta_k). \tag{3}$$

Різниці можуть бути обчислені за такими формулами:

$$\dot{m}_k = \frac{m_{k+1}-m_k}{\Delta t} = \dot{n}_k\vartheta_k + n_k\dot{\vartheta}_k; \tag{4}$$

$$\dot{D}_k = \frac{D_{k+1}-D_k}{\Delta t} = \dot{m}_k(1-\vartheta_k) - m_k\dot{\vartheta}_k; \tag{5}$$

$$\ddot{m}_k = \frac{\dot{m}_k-\dot{m}_{k-1}}{\Delta t} = 2\dot{n}_k\dot{\vartheta}_k + n_k\ddot{\vartheta}_k; \tag{6}$$

$$\ddot{D}_k = \frac{\dot{D}_{k+1}-\dot{D}_k}{\Delta t} = \ddot{m}_k(1-2\vartheta_k) - 2n_k\dot{\vartheta}_k^2. \tag{7}$$

Значення параметрів m_k, $\dot{m}_k$, $\ddot{m}_k$, D_k, $\dot{D}_k$, $\ddot{D}_k$ не можуть бути визначені точно на основі деякої вибірки з випадкової послідовності x_k. Можуть бути визначені з деякою похибкою лише їх оцінки $\hat{m}_k$, $\hat{\dot{m}}_k$, $\hat{\ddot{m}}_k$, $\hat{D}_k$, $\hat{\dot{D}}_k$, $\hat{\ddot{D}}_k$. Тому модель дискретних вимірювань може бути записана у вигляді:

$$z_k = f(y_k) + v_k, \tag{8}$$

де $z_k^T = [\hat{m}_k \quad \hat{\dot{m}}_k \quad \hat{\ddot{m}}_k \quad \hat{D}_k \quad \hat{\dot{D}}_k \quad \hat{\ddot{D}}_k]$ – вектор вимірювань;
$f(y_{k+1})$ – вектор-функція, визначена співвідношеннями (4)–(9);
v_k – шум вимірювань, який відображає похибки оцінювання параметрів m_k, $\dot{m}_k$, $\ddot{m}_k$, D_k, $\dot{D}_k$, $\ddot{D}_k$.

У рамках вирішуваного завдання для отримання складових вектора вимірювань z_k пропонується використовувати вибірку з N останніх елементів випадкової послідовності x_k. Іншими словами, складові вектора z_k є деякими функціями, аргументами яких беруться елементи x_k, x_{k-1}, …, x_{k-N+2}, x_{k-N+1} випадкової послідовності x_k, що потрапляють у рухоме вікно розміром N. Відповідно, складові вектора z_{k+1} є тими ж функціями, аргументами яких виступають елементи x_{k+1}, x_k, …, x_{k-N+1}, x_{k-N}, тобто для обчислення складових вектора z_k і z_{k+1} використовуються спільні аргументи. Це означає, що шум вимірювань $v_k = z_k - f(y_k)$ є автокорельованим, і модель такого шуму може бути записана у вигляді [4]:

$$v_{k+1} = A_k v_k + \eta_{k+1} \quad (9)$$

де A_k – деяка матриця;
η_k – білий гауссів шум з нульовим математичним очікуванням і коваріаційною матрицею R_k.

Використання розкладення в ряд Тейлора із залишенням перших двох членів з метою подолання нелінійності моделі вимірювань (10) і використання методу Брайсона–Хенріксона [4] з метою виключення автокорельованого шуму вимірювань v_k дає змогу отримати дискретний лінеарізований динамічний фільтр [5] для оцінювання вектора стану системи *(1)*, модель спостереження якої задається співвідношеннями *(8), (9):*
однокроковий предиктор

$$\bar{y}_{k+1} = \Phi \hat{y}_k; \quad (10)$$

коваріаційна матриця похибки однокрокового предиктора

$$P_{k+1,k} = \Phi P_{k,k} \Phi^T + bb^T Q; \quad (11)$$

нові вимірювання

$$s_{k+1} = z_{k+1} - A_k z_k; \quad (12)$$

нев'язка

$$\tilde{s}_{k+1} = s_{k+1} - f(\bar{x}_{k+1}) + A_k f(\hat{x}_k); \quad (13)$$

коваріаційна матриця нев'язки

$$C_{k+1} = H(\bar{y}_{k+1}) P_{k+1,k} H^T(\bar{y}_{k+1}) - H(\bar{y}_{k+1}) \Phi P_{k,k} H^T(\hat{y}_k) A_k^T - \\ - A_k H(\hat{y}_k) P_{k,k} \Phi^T H^T(\bar{y}_{k+1}) + A_k H(\hat{y}_k) P_{k,k} H^T(\hat{y}_k) A_k^T + R_{k+1}, \quad (14)$$

де $H(y^0)$ – матриця часткових похідних вектор-функції $f(y_k)$ за елементами вектора y_k у точці y^0;
коефіцієнт підсилення

$$K_{k+1} = W_{k+1} C_{k+1}^{-1}, \quad (15)$$

де

$$W_{k+1} = P_{k+1,k} H^T(\bar{y}_{k+1}) - \Phi P_{k,k} H^T(\hat{y}_k) A_k^T; \quad (16)$$

оцінка стану системи

$$\hat{y}_{k+1} = \bar{y}_{k+1} + K_{k+1} \tilde{s}_{k+1}; \quad (17)$$

коваріаційна матриця оцінки стану системи

$$P_{k+1,k+1} = P_{k+1,k} - K_{k+1} W_{k+1}^T. \quad (18)$$

Для реалізації алгоритму *(10) – (18)* необхідно знати дисперсію шуму системи Q_k і параметри моделі шуму вимірювань *(9)* – матриці A_k та коваріаційної матриці шуму R_k. У цьому випадку таких даних немає. При розв'язанні цієї проблеми доцільно використати підхід, який викладений в [6] і забезпечує стійкість процедури оптимальної динамічної фільтрації.

Таким чином, потужність співтовариства агентів, що беруть участь у формуванні тематичного інформаційного потоку, $n(t)$ і активність агентів $\vartheta(t)$ в рамках вирішення завдання щодо виявлення факту проведення інформаційних акцій (операцій, кампаній) може оцінюватись на основі використання дискретного адаптивного динамічного фільтра.

Список використаних джерел:

1. Advanced Persistent Threat (APT) [Електронний ресурс]. – Режим доступу: http://www.tadviser.ru/index.php – Заголовок з екрану.
2. Корн Г., Корн Т. Справочник по математике для научных работников и инженеров : Пер с англ. / Под. общ. ред. И. Г. Арамановича. – 5-е изд. – М. : Наука. Главная редакция физ.-мат. лит-ры, 1984. – 832 с.
3. Kalman R. E. A new approach to linear filtering and prediction problems // Journal of Basic Enginiering. – 1960. – Vol. 82, № 1. – P. 35–44.
4. Огарков М. А. Методы статистического оценивания параметров случайных процессов. – М. : Энергоатомиздат, 1990. – 208 с. : ил.
5. Сейдж Э. Мелс Дж. Теория оценивания и ее применение в связи управлении : Пер. с англ. / Под. ред. Б. Р. Левина. – М. : Связь, 1976.
6. Moghaddamjoo A., Kirlin R.L. Robust Adaptive Kalman Filtering with Unknown Inputs // Transaction on Acoustics, Speech and Signal Processing. – 1989, August. – Vol. 37, № 8. – P. 1166–1175.