УДК 004.047

А.Б. КАЧИНСЬКИЙ
Навчально-науковий фізико-технічний інститут Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна,
e-mail: *akachynsky@gmail.com*.

М.С. СТРЕМЕЦЬКА
Навчально-науковий фізико-технічний інститут Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна,
e-mail: *mira.stremetska@gmail.com.*

МЕТОД ГОЛОВНИХ КОМПОНЕНТ ЯК ІНСТРУМЕНТ АНАЛІЗУ МЕРЕЖЕВОГО ТРАФІКУ ДЛЯ ВИЯВЛЕННЯ DDoS-АТАК

Анотація. Для аналізу мережевого трафіку на основі методу головних компонент досліджено ознаковий простір високої розмірності. За допомогою сформованої системи показників виконано кластерний аналіз та візуалізацію мережевих з'єднань, використовуючи спостереження потоків мережевого трафіку, які зафіксовані під час шести різних типів DDoS-атак: DDoS Syn Flood, UDP Lag DDoS, UDP Flood DDoS, NetBIOS DDoS, LDAP DDoS, MSSQL DDoS. Результати дослідження провалідовано в умовах перехресної перевірки. Метод головних компонент є потужним інструментом розвідувального аналізу даних для підтримки моніторингу та виявлення підозрілих подій в кіберпросторі.

Ключові слова: метод головних компонент, перехресна перевірка, кластерний аналіз, оцінка, аномалія, кіберзагроза.

МЕТОД ГОЛОВНИХ КОМПОНЕНТ ДЛЯ ДАНИХ СПОСТЕРЕЖЕНЬ ПОТОКІВ МЕРЕЖЕВОГО ТРАФІКУ

Методологія операційної аналітики передбачає автоматизований збір значної кількості ознак потоків мережевого трафіку. Однак на практиці кількість цих ознак є занадто великою, а їхні значення — корельовані. Метод головних компонент (МГК) дає змогу розв'язати проблему, використовуючи меншу кількість інформативних показників, які пояснюють більшу частину дисперсії початкових даних [1].

Метод головних компонент для даних спостережень потоків мережевого трафіку полягає у розрахунку головних компонент із подальшим його використанням для інтерпретації цих даних [2]. Напрямки головних компонент — це напрямки в просторі ознак, уздовж яких початкові дані мають велику дисперсію. Ці напрямки також визначають лінії та підпростори, які максимально наближені до хмари початкових даних. Окрім обчислення похідних змінних МГК слугує також інструментом візуалізації та кластеризації даних. Отже, МГК — це спосіб конструювання ознакового простору за допомогою моделей, побудованих на спостереженнях потоків мережевого трафіку, що мають лінійні залежності.

Метою роботи є конструювання та валідація системи параметрів для аналізу даних потоків мережевого трафіку за допомогою МГК. Дослідження орієнтоване на конструювання спеціалізованої моделі параметрів, що за допомогою методів кластерного аналізу дасть змогу ефективно узагальнювати інформацію про потоки мережевого трафіку та кластеризувати мережеві з'єднання відповідно до типів атак DDoS (Distributed Denial-of-Service). Результати роботи сприятимуть підвищенню якості операційної аналітики та своєчасному виявленню підозрілих подій у кіберпросторі.

РОЗРАХУНОК ГОЛОВНИХ КОМПОНЕНТ З УРАХУВАННЯМ МУЛЬТИКОЛІНЕАРНОСТІ ПАРАМЕТРІВ

Концепція МГК базується на визначенні достатньої кількості факторів, які найбільше впливають на дисперсію даних та є зваженими лінійними комбінаціями