Predikcia bezpečnostných udalostí použitím viacrozmerných modelov časových radov

Autor: Bc. Miroslav Vojtek
Vedúci práce: JUDr. RNDr. Pavol Sokol, PhD.
Konzultant: RNDr. Andrej Gajdoš, PhD.

Ciele práce:

1. Analyzovať možnosti použitia viacrozmerných modelov časových radov v kybernetickej bezpečnosti.
2. Preskúmať a porovnať metódy viacrozmerných modelov časových radov z pohľadu predikcie bezpečnostných udalostí.
3. Navrhnúť, implementovať a vyhodnotiť systém na predikciu bezpečnostných udalostí.

V súčasnej dobe sme vystavení mnohým kybernetickým hrozbám a každodenne dôjde k veľkému počtu útokov. Preto je potrebné vylepšovať a vyvíjať nielen detekčné systémy, ale aj predikčné systémy. Informácie o útokoch z detekčných systémov môžeme použiť na návrh implementáciu a následné vylepšovanie systémov na predikciu bezpečnostných udalostí.

Táto práca sa bude venovať predikcii bezpečnostných udalostí použitím viacrozmerných modelov časových radov. Pod pojmom bezpečnostná udalosť rozumieme zistený stav systému, služby alebo siete, ktorý indikuje možné porušenie bezpečnostnej politiky alebo zlyhanie bezpečnostného opatrenia, alebo dovtedy neznámu situáciu, ktorá môže mať význam z hľadiska informačnej bezpečnosti [1]. V prvej časti tejto práce sa budeme venovať analyzovaniu možností použitia viacrozmerných modelov časových radov v kybernetickej bezpečnosti. Časovým radom označujeme hodnoty, ktoré sú zhromažďované, zaznamenávané alebo pozorované postupne v čase [2]. Časové intervaly medzi jednotlivými pozorovaniami vo všeobecnosti nemusia byť rovnaké. V tejto práci však budeme uvažovať o časových radoch s rovnakými časovými intervalmi medzi pozorovaniami. Ak v jednom čase pozorujeme hodnoty iba jedného parametra, hovoríme o jednorozmernom časovom rade, naopak, keď v jednom čase sledujeme hodnoty viacerých parametrov, tak hovoríme o viacrozmernom časovom rade. V druhej časti práce preskúmame a porovnáme metódy viacrozmerných modelov časových radov z pohľadu bezpečnostných udalostí. Vyberieme metódy, ktoré je možné použiť na predikciu bezpečnostných údajov a tie v nasledujúcej časti práce použijeme na predikciu.

V tretej časti práce navrhneme systém na predikciu, pričom použijeme vybrané metódy z predchádzajúcej časti. Systém implementujeme a následne vyhodnotíme jeho úspešnosť. Pri predikcii budeme vychádzať z dvoch datasetov bezpečnostných údajov. Prvým datasetom budú časové rady vytvorené zo záznamov o bezpečnostných udalostiach zo systému Warden. ktoré sú ukladané v štruktúrovanom formáte IDEA [3]. Tieto udalosti predstavujú informácie o zdroji hrozby zaznamenané IDS, honeypotmi, monitorovaním útokou na SSH autentizáciu alebo dáta z tretích strán, napríklad Shadowserver alebo Honeynet [4]. Z týchto dát zo systému Warden vytvoríme časový rad pre každý dôležitý parameter, napríklad kategória, cieľ, atď.

Druhým datasetom budú časové rady vytvorené z údajov o známych zraniteľnostiach, teda chybách v systémoch a technológiách, ktoré umožňujú sa prejaviť bezpečnostným hrozbám (napr. malvéru, úniku údajov a pod.) Pre každý pozorovaný parameter zraniteľnosti (napr. závažnosť, vplyv na dôvernosť, integritu, či dostupnosť) vytvoríme osobitný časový rad. Tieto údaje o zraniteľnostiach budeme čerpať z CVE Details [5], čo je informačný systém na jednoduchší prístup k zoznamu zraniteľností CVE od organizácie MITRE [6], ktorý obsahuje aj zraniteľnosti v NVD databáze NIST-u [7]. Po vytvorení všetkých skúmaných časových radov budeme hľadať vzťahy medzi nimi. Je vysoko pravdepodobné, že po zverejnení určitej bezpečnostnej zraniteľnosti sa zvýši počet útokov konkrétneho typu a kategórie. Tiež je ľahko vidieť, že ak sa v poslednom čase objavuje vyššie množstvo podobných útokov smerom, v ktorom zatiaľ zraniteľnosť nepoznáme, tak je možné konštatovať existenciu zraniteľnosti, ktorá ešte nebola zverejnená. Na základe súvislostí medzi skúmanými časovými radmi navrhneme, implementujeme a vyhodnotíme systém, ktorý bude bezpečnostné udalosti predikovať.

Literatúra:

[1] CSIRT.SK [online]. [cit. 16.12.2019]. Dostupné z https://www.csirt.gov.sk/doc/MFSRVzdelavanie/02Vzdelavanie2014/Studijne_materialy/Stud_2014_02_IT_IB_ucitelia.pdf.

[2] Projekt math.sk [online]. [cit. 16.12.2019]. Dostupné z https://www.math.sk/mpm/wp-content/uploads/2017/11/acr.pdf.

[3] Cesnet.cz [online]. [cit. 16.12.2019]. Dostupné z https://idea.cesnet.cz.

[4] Warden.Cesnet.cz [online]. [cit. 16.12.2019]. Dostupné z https://warden.cesnet.cz/cs/architecture.

[5] CVE Details [online]. [cit. 16.12.2019]. Dostupné z https://www.cvedetails.com/index.php.

[6] MITRE CVE [online]. [cit. 16.12.2019]. Dostupné z https://cve.mitre.org/index.html.

[7] NVD [online]. [cit. 16.12.2019]. Dostupné z https://nvd.nist.gov.

[8] BOX, George EP, et al. Time series analysis: forecasting and control. John Wiley & Sons, 2015.

[9] ESLING, Philippe; AGON, Carlos. Time-series data mining. ACM Computing Surveys (CSUR), 2012, 45.1: 12.

[10] DUA, Sumeet; DU, Xian. Data mining and machine learning in cybersecurity. CRC press, 2016.

[11] WERNER, Gordon; YANG, Shanchieh; MCCONKY, Katie. Time series forecasting of cyber attack intensity. In: Proceedings of the 12th Annual Conference on Cyber and Information Security Research. ACM, 2017. p. 18.