Analiza i Ocena Bezpieczeństwa Systemów Usługowych i Internetu Rzeczy -- projekt 2019

Tematy:

1. Wykrywanie anomalii ruchu teleinformatycznego z wykorzystaniem metod uczenia nadzorowanego
2. Wykrywanie anomalii ruchu teleinformatycznego z wykorzystaniem metod uczenia nienadzorowanego.
3. Ocena wpływu metody selekcji cech na efektywność i skuteczność wykrywania anomalii ruchu teleinformatycznego.
4. Ocena czułości wybranych algorytmów na typ obserwowanej anomalii ruchu teleinformatycznego.
5. Ocena wpływu stopnia agregacji obserwowanego ruchu teleinformatycznego na możliwość detekcji anomalii.
6. Wykrywanie anomalii w wartościach rekordów przechowywanych w bazie danych
7. Wykrywanie anomalii w strumieniach video
8. Wykrywanie anomalii w plikach/strumieniach audio
9. Wykrywanie zdarzeń nietypowych w logach systemów teleinformatycznych
10. Wykrywanie zdarzeń nietypowych w ruchu obiektów
11. Ocena jakości detekcji anomalii wybranych algorytmów w zależności od metody próbkowania ruchu teleinformatycznego.
12. Opracowanie wirtualnego stanowiska badawczego umożliwiającego testowanie różnorodnych algorytmów wykrywania anomalii w ruchu teleinformatycznym.
13. Wykrywania anomalii w ruchu teleinformatycznym z wykorzystaniem metody analizy szeregów czasowych – implementacja i porównanie wybranych metod.
14. Wykrywania anomalii w ruchu teleinformatycznym na podstawie klasteryzacji – implementacja i porównanie wybranych metod.
15. Wykrywania anomalii w ruchu teleinformatycznym wykorzystujące analizę statystyczną – implementacja i porównanie wybranych metod.
16. Zastosowanie przetwarzania równoległego na potrzeby wykrywania zdarzeń nietypowych w dużych zbiorach danych (np. MapReduce, Hdoop)
17. Ocena możliwości identyfikacji użytkowników serwisu webowego na podstawie charakterystycznych cech przeglądarki (browser fingerprinting)
18. Zdalna identyfikacja urządzeń mobilnych na podstawie wybranego zbioru atrybutów (device fingerprinting)
19. Wykrywanie ataków DDoS (symulator, badanie algorytmów, złożone środowisko zwirtualizowane np. DeterLab)
20. Zapobieganie atakom DDoS (symulator, badanie różnych strategii, złożone środowisko zwirtualizowane np. DeterLab)
21. Analiza poufności i anonimowości danych w systemach rozproszonego składowania i przetwarzania danych (na przykładzie Hadoop (Sector/Sphere))
22. Analiza metod anonimizacji danych. (Porównanie wydajności i ocena jakości algorytmów).
23. Projekt systemu bezpieczeństwa dla elektronicznego kantoru wymiany walut.
24. Usługa bezpiecznej poczty z gwarancją poufności i integralności korespondencji.
25. Projekt systemu bezpieczeństwa dla elektronicznej giełdy kryptowalut.
26. Projekt bezpiecznej usługi monitoringu wizualnego.
27. Projekt usługi uwierzytelniania z wykorzystaniem danych kontekstowych.
28. Analiza możliwości rozwiązań dla środowiska typu Security Operations Center na przykładzie OpenSOC i Apache Metron
29. Ocena możliwości wykorzystania funkcji typu Locality-sensitive hashing (LSH) na potrzeby wykrywania nietypowych plików, zdarzeń, itp. https://github.com/trendmicro/tlsh
30. Ocena skuteczności różnych metod łamania haseł (np. algorytm BFM, algorytm Weir, itp.)dla wybranych rzeczywistych zbiorów haseł)
31. Inny - uzgodniony

Zasady:

1. Przed drugimi zajęciami projektowymi należy dokonać wyboru tematu realizowanego projektu.
2. Każdy temat może być realizowany w zespole złożonym z maksymalnie 3 osób.
3. W trakcie semestru, należy przedstawić raporty częściowe zawierający aktualny stan prac (optymalnie powinny być to części raportu końcowego). Raport częściowe powinny być dokonane co najmniej raz w miesiącu (marzec, kwiecień, maj).
4. Ocena końcowa jest wystawiana na podstawie semestralnej pracy, raportu końcowego, praktycznego wyniku w postaci usługi/symulatora/aplikacji/rezultatów weryfikacji eksperymentalnej oraz wiedzy z zakresu poruszanych zagadnień.
5. Końcowe rozliczenie semestralnej pracy , odbiór wyników projektu oraz wystawienie oceny odbędzie się na ostatnich zajęciach w semestrze, t.j. w tygodniu 12.06.2019-18.06.2019.
6. Raport końcowy powinien zawierać następujące elementy:
    a. Strona tytułowa
    b. Lista autorów
    c. Lista zrealizowanych zadań (dla każdego z autorów osobna)
    d. Spis treści
    e. Wstęp (cel projektu)
    f. Definicja problemu
    g. Definicja hipotez i zadań
    h. Opis wykorzystanego/wytworzonego środowiska badawczego
        - udostępniane funkcjonalności
        - architektura, założenia, diagram klas, itp.
    i. Badania
        - Cel eksperymentu (co badamy, w jaki sposób)
        - Plan eksperymentu (jakie źródło danych, jakie środowisko wykonania, jaka metoda oceny, jak wygląda stan typowy/normalny, a jak definiujemy anomalię, …)
        - Przebieg eksperymentu (dla źródła danych A wykonujemy analizę przy pomocy algorytmu B, dla algorytmu C i D porównujemy wyniki analizy zbioru danych E, itp. …)
        - Prezentacja wyników (dla algorytmu/metody A i zbioru danych B wykryto poprawnie X anomalii, niepoprawnie zidentyfikowano
        - Analiza i ocena otrzymanych wyników
    j. Podsumowanie (ocena osiągniętej funkcjonalności, wyników badań, możliwości rozwoju, itp.)
    k. Załączniki (raportu bieżące, płyta CD z programem/obrazem MV, dokumenty pomocnicze, …)
7. Raporty częściowe i raport końcowy wraz ze źródłami i kodem należy przekazać prowadzącemu za pośrednictwem platformy eporta.pwr.edu.
8. Środowisko pracy (języki programowani, system operacyjny, symulatory) – należy uzgodnić po wyborze tematu projektu z prowadzącym.
    a. Jedna z ciekawszych możliwości - http://www.isi.deterlab.net/
9. Tematy mogą być kontynuowane/rozwijane w ramach pracy dyplomowej.

Literatura podstawowa:

1. Andrew Moore, Introductory overview of time-series-based anomaly detection algorithms
2. Arindam Banerjee, Varun Chandola, Vipin Kumar, Jaideep Srivastava, Aleksandar Lazarevic, Anomaly Detection: A Tutorial
3. Hans-Peter Kriegel, Peer Kröger, Arthur Zimek, Outlier Detection Techniques
4. Ignasi Paredes Oliva, Anomaly Detection
5. Chandola, V., Banerjee, A., and Kumar, V. 2009. Anomaly detection: A survey. ACM Comput. Surv. 41, 3, Article 15
6. Varun Chandola, Varun Mithal, and Vipin Kumar, A Comparative Evaluation of Anomaly Detection Techniques for Sequence Data
7. Shashank Shanbhag, Yu Gu, Tilman Wolf: A Taxonomy and Comparative Evaluation of Algorithms for Parallel Anomaly Detection. ICCCN 2010 (bezpłatny dostęp przez proxy pwr)
8. R. A. Maxion and R. R. Roberts, Proper Use of ROC Curves in Intrusion/Anomaly Detection
9. LeBlanc, Jonathan, and Tim Messerschmidt. Identity and Data Security for Web Development: Best Practices. " O'Reilly Media, Inc.", 2016.
10. Bhattacharyya, Dhruba Kumar, and Jugal Kumar Kalita. DDoS attacks. CRC Press, 2016.

Literatura dodatkowa:

1. Marina Thottan, Guanglei Liu, Chuanyi Ji, Anomaly Detection Approaches for Communication Networks
2. Pedro Casas Hernández, Statistical Analysis of Network Traffic for Anomaly Detection and Quality of Service Provisioning
3. Gerhard Münz, Traffic Anomaly Detection and Cause Identification Using Flow-Level Measurements
4. Daniela Brauckhoff, Network Traffic Anomaly Detection and Evaluation
5. Fernando Jorge Silveira Filho, Unsupervised Diagnosis of Network Traffic Anomalies
6. George Nychis, An Empirical Evaluation of Entropy-based Anomaly Detection
7. DanielaBrauckhoff*, Arno Wagner, Martin May, FLAME: A Flow-level Anomaly Modeling Engine
8. Marcin Żurakowski, Przemysław Kazienko, Zastosowanie sieci Bayesa w wykrywaniu ataków DoS
9. Wybrane pozycje bibliograficzne z : Chandola, V., et.al. Anomaly detection: A survey.

Źródła danych do testów

1. Canadian Institute for Cybersecurity https://www.unb.ca/cic/datasets/index.html
2. Analyzing Web Traffic ECML/PKDD 2007 Discovery Challenge http://www.lirmm.fr/pkdd2007-challenge/ https://gitlab.fing.edu.uy/gsi/web-application-attacks-datasets
3. THE CTU-13 DATASET. A LABELED DATASET WITH BOTNET, NORMAL AND BACKGROUND TRAFFIC, https://mcfp.weebly.com/the-ctu-13-dataset-a-labeled-dataset-with-botnet-normal-and-background-traffic.html
4. Stratosphere IPS https://www.stratosphereips.org
5. Cyber Range Lab of the Australian Centre for Cyber Security (ACCS) https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cybersecurity/ADFA-NB15-Datasets/
6. LBNL/ICSI Enterprise Tracing Project http://www.icir.org/enterprise-tracing/
7. Security Repo http://www.secrepo.com
8. KDD Cup 1999 Data, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
9. Packet traces from WIDE backbone, http://mawi.wide.ad.jp/mawi/
10. CAIDA Data, http://www.caida.org/data/overview/
11. Generator ruchu (Poisson, on-off, …)
12. D-ITG (Distributed Internet Traffic Generator) - http://www.grid.unina.it/software/ITG/
13. Inne (logi serwerów http, pliki pcap, …)
14. Network Data Sources, http://pajek.imfm.si/
15. Data Mining Community's Top Resource http://www.kdnuggets.com/datasets/index.html
16. Frequent Itemset Mining Dataset Repository http://fimi.ua.ac.be/data/

17. University of Oregon Route Views Archive Project http://archive.routeviews.org/
18. Waikato Internet Traffic Storage http://www.wand.net.nz/wits/catalogue.php
19. http://www.netresec.com/?page=PcapFiles

Przykłady zastosowań:

1. https://www.elastic.co/blog/implementing-a-statistical-anomaly-detector-part-1
2. http://techblog.netflix.com/2015/02/rad-outlier-detection-on-big-data.html
3. http://nerds.airbnb.com/anomaly-detection/
4. https://logentries.com/product/anomaly-detection/
5. https://anomaly.io
6. http://www.machine-analytics.com