fundlylive.com

Arash Habibi Lashkari's Pioneering Contribution to Unlocking Cybersecurity Insights: The Essential Role of Cybersecurity Dataset

10–12 minutes

Arguing that you don’t care about the right to privacy because you have nothing to hide is no different than saying you don’t care about free speech because you have nothing to say.

- Edward Snowden

Humans enjoy numerous advantages in today’s modern era, characterized by astonishing luxuries and technological advancements. However, with the convenience and accessibility of these advancements, there comes a corresponding increase in potential roadblocks that have the power to wreak havoc within communities. Among the most significant threats to this digital landscape are cybersecurity issues, which have the potential to inflict damage on data, steal valuable financial information, and disrupt computing systems. The consequences of such actions can lead to substantial losses within a community, often resulting in irreparable damage. Recognizing the gravity of this challenge, technology experts have devised various approaches to protect communities from the disruptive activities of cybercriminals. One

notable cybersecurity expert contributing to this field is Arash Habibi Lashkari.

Cybersecurity is paramount in addressing contemporary societal challenges, as it safeguards all data types against theft and malicious activities. This protection encompasses a wide range of sensitive data, including personally identifiable information (PII), protected health information (PHI), personal data, intellectual property, governmental and industry information systems, and much more. Businesses increasingly utilize advanced analytics methods such as machine learning and predictive modeling to identify and mitigate potential cyber risks proactively. Data analytics empowers organizations to fortify their cybersecurity defenses through real-time monitoring, anomaly detection, and data-driven decision-making.

Data analytics is pivotal in identifying patterns and irregularities within network traffic, user behavior, and system performance. Applying statistical and machine learning techniques makes it possible to distinguish between normal and abnormal patterns, evaluate outliers and deviations, and flag potential threats and security incidents. The availability of cybersecurity datasets for analysis and testing is instrumental in developing and fine-tuning effective cybersecurity solutions, ensuring the protection of critical data and systems from the ever-evolving landscape of cyber threats.

Our narrative begins with Professor Lashkari's profound realization that proper comprehension of cybersecurity's intricacies could not be achieved through mere theory alone. Instead, it necessitated acquiring and analyzing real-world data—an insight that laid the foundation for a transformative expedition. Professor Lashkari embarked on this quest with unwavering determination and a commitment to gathering, refining, and contributing to the datasets that would ultimately shape the future of cybersecurity. His journey would take him through the annals of time and technology, witnessing the metamorphosis of these datasets from rudimentary collections of network traffic, malware samples, and URLs into the sophisticated and all-encompassing resources we know today.

1. **Understanding Network Traffic (2015):**2015 Prof. Lashkari and his research team laid the foundation for their dataset creation journey. They focused on comprehending network traffic patterns in Virtual Private Networks (VPN) and non-VPN environments. This effort resulted in the "VPN-nonVPN Network Traffic dataset," a comprehensive resource covering various traffic categories. This dataset equipped researchers with labeled network traffic data and full packet captures, setting the stage for advanced research in network security.

2. **Unveiling Darknet Traffic (2016):**The following year, in 2016, their research shifted towards understanding darknet traffic, precisely the behavior of Tor and non-Tor networks. This endeavor led to the creation of the "Tor-nonTor Network Traffic dataset." By amalgamating ISCXTor2016 and ISCXVPN2016 public datasets, this resource facilitated the classification of Tor and VPN applications, shedding light on anonymization techniques and network behavior.
3. **Android Malware and Malicious URLs (2017):**In 2017, Prof. Lashkari and his team pursued a dual focus on Android malware analysis and malicious URL detection. They introduced two critical datasets:

- At first, they produced "**Android Adware and General Malware Dataset (AAGM)**", including 1900 Android malware using real smartphones instead of simulators.
- The "CICAndMal dataset" became a cornerstone for Android malware detection and classification solutions, featuring over 10,854 samples categorized into various types of malware.
- Simultaneously, the "ISCX-URL dataset" emphasized identifying and categorizing malicious URLs, offering insights into URL-based security challenges.

4. **Advancements in Intrusion Detection (2018):**2018 saw significant progress in intrusion detection datasets by introducing the "CIC-IDS" and "CSE-CIC-IDS" datasets. This resource encompassed seven distinct attack scenarios, enabling the evaluation of intrusion detection systems and providing a robust benchmark for research in intrusion detection and network security.

5. **Focused on DDoS Attacks and improved the Android malware analysis (2019):**In 2019, his focus shifted to Distributed Denial of Service (DDoS) attacks with the creation of the “CICDDoS2019” dataset. This invaluable resource included 12 DDoS attacks following CAPEC standards, offering network traffic captures and 80 features extracted from the captured traffic. Also, Prof Lashkari and his research team produced another Android malware analysis dataset, namely “Investigation of the Android Malware (CIC-InvesAndMal2019)”, which includes permissions and intents as static features and API calls and all generated log files as dynamic features in three steps (During installation, before restarting and after restarting the phone).
6. **Android Malware Analysis and DNS Traffic (2020):**In 2020, Prof. Lashkari and his team continued to make significant contributions by introducing three notable datasets:

- The “CCCS-CIC-AndMal-2020 dataset” furthered Android malware analysis and security research.
- Shifting their focus to DNS traffic, they introduced the “CIRA-CIC-DoHBrw-2020 dataset,” facilitating DNS analysis, testing, and evaluation over HTTPS (DoH) traffic.
- Improving their encrypted traffic analysis dataset by generating “DarkNet (CIC-Darknet2020)” to analyze and characterize the darknet traffic known as network telescopes, sinkholes, or black holes.

7. DNS Traffic Analysis (2021):**In 2021, they emphasized DNS traffic analysis with two datasets:**

- The “CIC-Bell-DNS2021 dataset” focused on the analysis of

malicious DNS traffic.

- Simultaneously, the “CIC-Bell-DNS-EXF-2021 dataset” explored DNS exfiltration traffic.

8. **Memory-Based Malware Analysis and Source Code Attribution (2022):**In 2022, Prof. Lashkari and his team introduced three diverse datasets:

- The “CIC-MalMem-2022 dataset” spotlighted memory-based malware analysis.
- The “CIC-Evasive-PDFMal2022 dataset” focused on evasive PDFs.
- They also explored source code authorship attribution with the “Source Code Authorship Attribution (YU-SCAA-2022)” dataset.

9. **Recent Additions in 2023:**As of 2023, two new datasets have enriched the cybersecurity landscape:

- The “Modbus Dataset 2023 (CIC-Modbus-2023)” focuses on substation network security.
- The “SQL Injection Attack (BCCC-SFU-SQLInj-2023)” dataset expands the malicious SQL dataset, advancing research in SQL injection attack detection and prevention.

By creating and releasing these cybersecurity datasets, Prof. Lashkari and his research team revolutionized cybersecurity AI model training, testing, and evaluation to support complex processes of understanding, categorizing, and defending against various cybersecurity threats. Prof. Lashkari is also well-known as a forerunner in designing and producing Cybersecurity open-source analyzers. The Understanding Cybersecurity Series (UCS) knowledge mobilization program is one of his essential

establishments. His main objective remains to spread awareness regarding the threats involved in the digital realm to solve the adversities of cybersecurity on a massive level. He fashioned efficient and relevant equipment in collaboration with different industries, using admirable research to maximize its reach among the community.

This method has helped the firm define and deliver vital knowledge and awareness to the public, and it is done in a way that is consumable, comprehensible, and accessible to everyone. The proposed resolution of this program will be to support the education and training of IT systems and cybersecurity students, academics, researchers, developers, and industry professionals while, at the same time, sharing some of the more comestible and applicable research findings with the larger community. The greater community will simplify the expenditure of educational materials through social networks and other popular information-sharing tools. UCS creates two kinds of data: academic and technical information and public non-technical materials. The educational and technical materials aim for an audience of academic scholars, advanced educators, post-secondary students, researchers, and developers. Along with collecting academic materials, the non-technical materials will target broad spectators, including youth, seniors, laypeople, and the public.

Prof. Lashkari is currently a Canada Research Chair (CRC) in cybersecurity. As the founder and director of the Behaviour-Centric Cybersecurity Center (BCCC) at York University in Canada, his current work involves developing vulnerability detection technology to protect network systems against cyberattacks, as he has more than two decades of concurrent industrial and development

experience in network, software, and computer security. He oversees many research and development teams simultaneously, all engaged in various projects, including analyzing network traffic, malware, Honeynets, and threat hunting.