patreon.com

Streamlining Cybersecurity Knowledge Retrieval - The Power of Information Extraction Tools and Prof. Arash Habibi Lashkari’s Groundbreaking | Patrick John

8–11 minutes

In the fastidious developing era of cybersecurity, one should have a strong command of this domain and needs to stay ahead of rising threats and vulnerabilities. Administrations must employ the most recent knowledge about their environments and the ever-intensifying body of cybersecurity information to manage security risks efficiently. Nevertheless, this task is a complicated challenge and is gradually more demanding due to the perpetual updates and distribution of information across different resources.

In cybersecurity, extracting information, such as entities, events, features, and relations, from diverse data sources and then populating a knowledge base with this extracted information constitutes a foundational challenge. Numerous data resources like Network Traffic, Memory, Resource Usage, and logs have not been purposefully designed to facilitate Information Extraction (IE) for cybersecurity applications. This absence of dedicated analytical tools can substantially impede our ability to detect, analyze, identify, and characterize activities to profile benign and malicious behavior and promptly report emerging security threats.

The Challenge of Cybersecurity Knowledge Retrieval

The vibrant cybersecurity environment demands skilled personnel to sustain a real-time understanding of liabilities, attacks, and emerging threats. Often, serious security information can be witnessed upsurging in various sources, such as network traffic, memory, logs, and resource usage. Organizing and shaping this diffused information is essential for spreading awareness and enabling intrusion detection systems to secure against attacks, especially for “zero-day” attacks. Let’s highlight the crucial role of information extraction tools in streamlining knowledge retrieval, commemorating the remarkable contributions of Prof. Arash Habibi Lashkari, a preceding figure in this field.

Prof. Arash Habibi Lashkari - A Cybersecurity Star and Mentor

Born November 3, 1974, Arash Habibi Lashkari stands out as a celebrated face in the kingdom of education, serving as a distinguished educator, associate professor, and author. He is the author of ten published books and more than 110 academic articles on various cybersecurity-related topics. He has received 15 awards at international computer security competitions - including three gold awards - and was recognized as one of Canada's Top 150 Researchers for 2017. He holds the prestigious position of Canada Research Chair (CRC) in cybersecurity and assists as an associate professor at York University. His supremacy in his domain has facilitated him to accomplish significant contributions to the field of cybersecurity. Lashkari is also a senior member of the Institute of Electrical and Electronics Engineers (IEEE), an adjunct professor in the faculty of Computer Science at the University of New Brunswick (UNB) in Canada, and an adjunct graduate faculty member at the North Carolina Agricultural and Technical State University in the States.

Contribution to the Cybersecurity Open-Source Information

Extraction (IE) and Data Analysis

Data extraction tools are software applications that streamline gathering and retrieving data from diverse sources. They are purpose-built to simplify data extraction, whether structured or unstructured, from web pages or other data repositories. A data extraction tool can efficiently acquire valuable data points such as customer information, product details, financial data, or any other pertinent information crucial for analysis or decision-making.

Equipped with user-friendly interfaces and intuitive functionalities, a well-suited data extraction tool facilitates easy access to and retrieval of specific data without requiring intricate coding or manual data entry. These tools automate the extraction process, saving time and effort while ensuring the extracted data’s accuracy and reliability.

Data extraction tools offer several advantages:

· They save time by automating data collection and extraction, enabling efficient retrieval of large volumes of data from diverse sources.

· They enhance efficiency by eliminating manual tasks like searching and copying, allowing users to focus on more critical activities.

· They improve data accuracy by reducing human errors associated with manual entry and extracting information directly from sources.

· These tools provide comprehensive insights by consolidating data from multiple sources, enabling the identification of patterns and trends for informed decision-making.

· They are scalable and can efficiently handle large datasets and diverse data sources.

In 2015, Prof. Lashkari started a journey of designing, developing, and launching cybersecurity Information Extraction (IE) or analysis tools when he arrived in Canada. He initiated this endeavor by collaborating with his two colleagues during his Postdoctoral fellows at the University of NewBrunswick (UNB) to create ISCXFlowMeter.

ISCXFlowMeter served as a network traffic flow generator and feature extractor, capable of generating bidirectional flows with the first packet determining both the forward (source to destination) and backward (destination to source) directions. This unique feature allowed for the independent calculation of statistical time-related features in each direction. ISCXFlowMeter could extract 32 statistical features and convert raw network traffic PCSP files into CSV format. In 2017, they designed and developed a new version of Network Traffic analyzer, CICFlowMeter, which could extract more than 80 features.

In 2019, Professor Lashkari and his research team introduced the Static and Dynamic Android App Analyzer (AndroidApplyzer). This research project focuses on classifying Android samples through static and dynamic analysis. The tool evolves through multiple versions, including data collection and static feature extraction. Subsequent versions integrate AI-based classification models and dynamic analysis modules, enhancing its capabilities for effectively categorizing Android applications.

In 2020, Professor Lashkari's team developed the DNS over HTTPS (DoH) Analyzer (DoHLyzer), a toolkit designed to capture

HTTPS traffic, extract statistical and time-series features, and analyze them with a specific focus on detecting and characterizing DoH (DNS-over-HTTPS) traffic. This tool is invaluable in understanding and monitoring network behavior related to encrypted DNS requests.

In 2021, his research team introduced the Volatility Memory Analyzer (VolMemLyzer). This Python tool is dedicated to memory forensics, a crucial aspect of analyzing malicious activities during live malware infections. VolMemLyzer extracts over 36 features from memory snapshots using the Volatility tool, streamlining the process of identifying and investigating malware-infected systems.

In 2022, as part of Professor Lashkari's knowledge mobilization strategy, Professor Lashkari's research team developed the IMAP Bot AnaLyzer (IMAPBotLyzer) to address the growing concern of credential stuffing attacks targeting the Internet Mail Access Protocol (IMAP). This tool distinguishes between human and bot interactions by employing behavioral biometrics like mouse and keystroke dynamics, thus providing an effective defense mechanism against these attacks. Additionally, the team designed the PDF Malware Analyzer (PDFMalLyzer), which extracts 31 distinct features from PDF files, aiding in detecting malicious PDF documents. Lastly, the Authorship Attribution Analyzer (AuthAttLyzer) was introduced, enabling the identification of the source code's author by extracting various features, including N-grams, word-based embeddings, and abstract syntax tree (AST) characteristics.

These analyzers collectively exemplify the power of information extraction tools in streamlining cybersecurity knowledge retrieval. By developing specialized tools like the CICFlowMeter (formerly

ISCXFlowMeter), Static and Dynamic Android App Analyzer (AndroidApplyzer), DNS over HTTPS (DoH) Analyzer (DoHLyzer), Volatility Memory Analyzer (VolMemLyzer), IMAP Bot AnaLyzer (IMAPBotLyzer), PDF Malware Analyzer (PDFMalLyzer), and Authorship Attribution Analyzer (AuthAttLyzer), Prof. Lashkari and his research team at Behaviour-Centric Cybersecurity Center (BCCC) have harnessed the capabilities of information extraction to simplify the complex process of understanding, categorizing, and defending against various cybersecurity threats. These tools automate data collection, feature extraction, and analysis, enabling cybersecurity experts to efficiently access critical insights and actionable information from diverse data sources, ultimately enhancing our ability to detect, respond to, and mitigate evolving security risks.