Cyber Security Challenge

Challenge Overview:

Develop a tool or technique that can successfully and consistently scrape dark web forums that have advanced security measures in place to prevent unauthorized access and scraping using knowledge and skills in cybersecurity, machine learning, and data mining. The tool should be able to bypass or overcome captchas, timeouts, IP banning, and other security measures, without being detected or blocked.

Objectives:

- Develop a tool or technique that can successfully scrape dark web forums continously in real-time.
- Bypass or overcome advanced security measures, such as captchas, timeouts, and IP banning.
- Avoid being detected or blocked by the security measures of the dark web forums.
- Deploy the tool on Cloud like AWS to be able to fetch the latest updates on the forum with least delay as possible.
- Support full text search feature in the tool to search for keywords in the scrapped data.

Evaluation Criteria:

- Ability to successfully scrape dark web forums. The more the number of forums covered, the higher the scoring awarded.
- Effectiveness in bypassing or overcoming advanced security measures. The level of reliability is checked.
- The time taken by the crawler to discover the latest content. The earlier the content is discovered, the higher points awarded.

- The time taken to search a keyword in the crawled data. The quicker the results are returned, the better.

Rules:

- Participants must demonstrate that their tool or system can successfully scrape dark web forums by presenting sample data and results.
- Participants shall leverage existing open source frameworks & tools.
- A team can have upto 4 members & a member shall be a part of only one team. Collaboration between different teams is not allowed.
- Participants shall register new accounts by themselves & take care of any measures needed.
- Participants shall arrange their own payment methods for AWS or other similar cloud platforms.
- The crawling should be as deep and exhaustive as possible, meaning, if there is a thread, all comments & replies should be gathered and stored.

Dark web forums for web scraping:

- Breach Forums – http://breached65xqh64s7xbkvqgg7bmj4nj7656hcb7x4g42x753r7zmejqd.onion
- Dark Net Forum – http://3otgxq7d33rwspxfquwkqlp7r4yoicofniypk7hcxxqefrwhptqp7zad.onion
- Dark Forest Forums - http://dkforestseeaaq2dqz2uflmlsybvnq2irzn4ygyvu53oazyorednviid.onion

Use cases and implementation for this tool

The tool could use web scraping and natural language processing techniques to automatically crawl and extract data from dark web forum. This would involve identifying the relevant forums, and establishing a connection to them, and then using web scraping and NLP algorithms to extract the relevant data.

This could use keyword search and regular expressions to identify specific patterns or keywords that are associated with data and credential leaks. The identified discussions and transactions could be tracked and monitored over time to track the movement and sale of stolen data. This could involve storing the data in a database and using data visualization and analytics tools to track trends and patterns over time.

Web scraping algorithms and tools that can extract data from the forums could be developed. This could involve using open source libraries and frameworks, such as BeautifulSoup and Scrapy, to build custom web scraping tools that can extract the relevant data from the forums. NLP algorithms and tools can be used to process and analyze the extracted data and open source libraries and frameworks, such as NLTK and spaCy, could be used to build custom NLP algorithms that can classify the data, identify potential threats, and extract relevant information from the forums.

The extracted and processed data can be stored in a database, and data visualization and analytics tools could be used to track trends and patterns over time. This could involve using a database management system, such as MySQL or MongoDB, to store the data and dashboards that show the movement and sale of stolen data on the dark web.

FAQs & Doubts

Feel free to join the slack workspace to checkout if we have already answered your question in the pinned messages in the #general channel. We will be glad to help in case you have any questions.

The link to the workspace is:
https://join.slack.com/t/slack-q6m2159/shared_invite/zt-1ln6knasz-hi0mTQhpsAQPJHR4Oy~FJQ

Machine Learning Challenge

Challenge:

Develop a product that can search for news articles mentioning a given name, and determine the sentiment & context associated with the name from those articles.

Objective:

The goal of this challenge is to develop a tool that can help generate a contextual profile analysis of a given person and see how they are being portrayed in the news. The tool should be able to search for news articles mentioning the given name, extract the textual content & identify the passages where the name is mentioned, and determine the sentiment associated with the name of interest. The tool should then make the collected articles and their associated sentiments available to the user through a user-friendly interface. The final objective is to summarize the contextual information for the name.

Requirements:

- The product must be able to search Google for news articles mentioning a given name and collect the resulting articles.
- The product must use NLP techniques to extract the relevant passages from the collected articles and determine the sentiment associated with those passages.
- The product must use techniques such as part-of-speech tagging and named entity recognition to identify the relevant passages, and use machine learning algorithms trained on a large and diverse dataset to classify the sentiment of those passages as positive, negative, or neutral.
- The product must store the collected articles and their associated sentiments in a database and the user should be able to access and query the collected data.

- The product must provide a user interface that allows users to input a name, view the resulting articles and sentiments, and filter and sort the results.
- The user interface should be easy to use and understand, and should include features such as search autocomplete, pagination, and sorting by date or sentiment.

Evaluation Criteria:

The product's ability to accurately determine the sentiment of the collected articles.

- How quickly and effectively it is able to search for and collect the news articles.
- How well it is able to extract the relevant passages from the collected articles and accurately classify their sentiment.
- The ease of use of the product with respect to how easy it is for users to input a name, view the results, and filter and sort the articles.
- The product's overall design and functionality.

Rules:

- Participants should respect the website's terms of service, avoid scraping sensitive or personal information, and should not overload the website with requests.
- Participants should use appropriate NLP techniques and machine learning models for sentiment analysis. Namely, techniques that have been validated and tested on large and diverse datasets, and should avoid biased or unreliable techniques.
- The product should be thoroughly tested on a variety of different test cases, including positive and negative news articles to check for accuracy and performance issues.
- A team can have up to 4 members & a member shall be a part of only one team. Collaboration between different teams is not allowed.

Use cases and implementation for this tool

Develop a web scraper that can search Google for news articles mentioning a given name, and collect the resulting articles. This could involve using a web scraping library or framework, such as Scrapy or Beautiful Soup or SERP API, to navigate the Google search results and extract the relevant information from the collected articles.

Use NLP techniques to extract the relevant passages from the collected articles where the given name is mentioned, and determine the sentiment associated with those using techniques such as part-of-speech tagging and named entity recognition to identify the relevant passages. Tools such as NLTK, spaCy, or Gensim can be used for the NLP tasks, and tools such as scikit-learn or TensorFlow can be used for any machine learning tasks involved in the development of the product.

Develop a web application that provides a user interface and handle user requests using any web framework. The web application should be connected to the database where the collected articles and their sentiments are stored, using an API or other interface to access the data and query it based on the user's input.

The web application could include features that allow the user to input a name, view the resulting articles and associated context. It could also have visualizations, such as charts and graphs, to help the user understand the results.

FAQs & Doubts

Feel free to join the slack workspace to checkout if we have already answered your question in the pinned messages in the #general channel. We will be glad to help in case you have any questions.

The link to the workspace is
https://join.slack.com/t/slack-q6m2159/shared_invite/zt-1ln6knasz-hi0mTQhpsAQPJHR4Oy~FJQ

Attack Surface Discovery Hackathon Challenge

Challenge Overview:

Develop a tool that can scan the internet-facing host systems of an organization's infrastructure for a given domain and gather detailed information about the website's technology stack, and identify potential security vulnerabilities and issues in the domain including any open ports and misconfigured or outdated services with reference to the Common Vulnerabilities and Exposures (CVE) database, that could be exploited by cyber criminals and hackers.

Objectives:

- Develop a tool that can perform detailed scanning of a given domain of an organization and identify if the website belonging to the domain is vulnerable to any known security issues and vulnerabilities in the latest CVE database including DNS attacks, cross-site scripting (XSS), server-side request forgery (SSRF), cross-site request forgery (CSRF), remote code execution (RCE), and local privilege escalation (LPE).
- The tool should be able to perform port scanning to scan for any open ports and services and gather information about the website's technology stack including the operating system, web server, and database server details.
- The tool should gather information about the frameworks, and libraries used by the domain including the programming language and any third–party API or services used by the website by analyzing the source code or other publicly available information using relevant tools and techniques.
- The tool should be able to identify any misconfigurations in the target website, and determine if it is vulnerable to HTTP or TLS compression attacks, and has any vulnerable underlying encryption algorithms such as SHA1 or MD5.

- The tool should scan the website to gather information about any security issues and vulnerabilities and analyze the website's network traffic to get details about the services and third-party connections of the website.
- The tool should map open ports to the corresponding service based on metadata like banners & headers. The version shall be leveraged to check for vulnerabilities & gather information about the severity of the CVE to prioritize the discovered vulnerabilities.
- The tool should include a user-friendly interface that should allow the user to specify a target domain and view the results of the scan with details about the identified security issues and vulnerabilities with respect to the CVE database.

Evaluation Criteria:

- The effectiveness and comprehensiveness of the tool in performing a thorough scan of the given domain, including all subdomains and external connections, to identify potential vulnerabilities and security risks.
- Choice of algorithms and techniques that scan and analyze the domain and provide results and detect potential security issues as quickly as possible.
- The ease of use of the tool with respect to how quickly and easily the users can scan their website's domain and view the results.
- The report generation capacity of the tool with the ability to generate a detailed report on the website's attack surface, including a list of identified vulnerabilities and security issues.

Rules:

- Participants must not attempt to gain unauthorized access to any website or network as part of their tool development or testing.
- Participants must not share or disclose any confidential information that they may come across during the development and testing of their tool.
- Participants should not engage in any activities that could be considered illegal, unethical, or malicious, such as hacking or denial of service attacks.

- Participants should not plagiarize or otherwise use the work of others without proper attribution.
- A team can have up to 4 members & a member shall be a part of only one team. Collaboration between different teams is not allowed.

FAQs & Doubts

Feel free to join the slack workspace to checkout if we have already answered your question in the pinned messages in the #general channel. We will be glad to help in case you have any questions.

The link to the workspace is
https://join.slack.com/t/slack-q6m2159/shared_invite/zt-1ln6knasz-hi0mTQhpsAQPJHR4Oy~FJQ

PROPOSED POLICING CHALLENGES FOR HACKATHON TO BE CONDUCTED BY IIT, GUWAHATI

Following Problem Statements of Assam Police are proposed for consideration as Challenges for the Hackathon.

1. AP-iSMART (Assam Police Integrated System for Monitoring and Alerting in Real-Time using Public and Private CCTV Camera Feeds)

- A Centralised CCTV Camera Feed Monitoring Solution, which shall ingest CCTV Camera Feeds from various participating Public and Private parties.
- The solution shall use AI/Machine Learning powered models to detect various types of incidents like Illegal Trespassing, Theft, etc.
- Integration of FRS (Facial Recognition System) with the Solution to generate real-time alerts and dissemination of the same to the field level officers.

2. Predictive Policing using existing CCTNS Data

Based on an anonymised master data from CCTNS containing various fields like details of the complainant & accused along with their age and gender, sections of law (types of crime), places of occurrence, etc. related to criminal cases, a Detailed Analysis of the data including Statistical and Predictive models which shall assist Assam Police in Crime Detection and Prevention.

3. Smart Naka

There is a database of stolen vehicles in the Vaahan portal maintained by the Ministry of Road Transport and Highways, Government of India. We need to create a mobile app integrating with Vaahan portal through API so that a constable executing naka at the road can check the details of a suspicious vehicle in his mobile at real time.

4. Video Enhancement:

A software or app that can enhance a given Photo or video that will help us identify the target object. Such as vehicle number plates, rendering distorted photos, etc. that will help in the investigation.

Blockchain

The blockchain domain is an open-ended one, so you are free to solve any of the problems from above using Blockchain technologies, or come up with your own idea entirely. There are exclusive prizes from Polygon, Solana for the same, and we encourage you to check them out at https://ethosiitg.devfolio.co .