**Synthetic Datasets for Cybersecurity Threat Intelligence**

These synthetic datasets provide a foundation for exploring a wide range of research questions, developing predictive models, and simulating various cybersecurity threat scenarios. Researchers can leverage these datasets to enhance their understanding of threat intelligence, test security tools, and contribute to the development of more robust cybersecurity defenses.

## **Indicators of Compromise (IoCs)**

### Description

This dataset simulates information about potential Indicators of Compromise (IoCs), including the type of indicator (e.g., IP Address, Domain, File Hash), the indicator value, context, and confidence level.

### Possible Applications

1. Threat detection and incident response.
2. Analyzing patterns of IoCs to identify potential threats.
3. Building and testing IoC-based detection algorithms.

### Possible Research Questions

1. How does the distribution of IoC types vary over time?
2. Can the confidence level of an IoC be used to predict the severity of an incident?
3. What are the commonalities in context between high-confidence IoCs?

## **Tactics, Techniques, and Procedures (TTPs)**

### Description

This dataset simulates information about cyber threat Tactics, Techniques, and Procedures (TTPs), including details on tactics, techniques, and procedures employed by threat actors.

### Possible Applications

1. Identifying and mitigating specific TTPs used in cyber-attacks.
2. Enhancing threat intelligence feeds with TTP information.
3. Training and testing detection systems against known TTPs.

### Possible Research Questions

1. Which TTPs are most commonly associated with specific threat actor types?
2. How do TTPs evolve over time in response to security measures?
3. Can the detection of specific TTPs contribute to predicting the nature of an attack?

## **Vulnerabilities Dataset**

### Description

This dataset simulates information about known vulnerabilities, including vulnerability ID, Common Vulnerabilities and Exposures (CVE) ID, description, and severity.

### Possible Applications

1. Prioritizing and patching vulnerabilities based on severity.
2. Analyzing the correlation between exploited vulnerabilities and incidents.
3. Assessing the risk associated with specific vulnerabilities.

### Possible Research Questions

1. How does the severity of vulnerabilities correlate with their exploitation in incidents?
2. Can the historical exploitation of vulnerabilities be used to predict future incidents?
3. What are the common characteristics of vulnerabilities associated with specific threat actors?

## Malware Dataset

### Description

This dataset simulates information about malware, including malware name, MD5 hash, and a brief description.

### Possible Applications

1. Developing and testing anti-malware solutions.
2. Analyzing the prevalence and characteristics of different types of malware.
3. Investigating the relationship between malware and specific incidents.

### Possible Research Questions

1. How does the distribution of malware types change over different campaigns?
2. Can the characteristics of malware be used to predict the target systems or industries?
3. What patterns exist in the temporal evolution of malware prevalence?

## **Campaigns Dataset**

### Description

This dataset simulates information about cyber campaigns, including campaign ID, name, type, and motivation.

### Possible Applications

1. Analyzing the motives and tactics of different cyber campaigns.
2. Understanding the impact and scope of specific campaigns.
3. Studying the correlation between campaign characteristics and incident severity.

### Possible Research Questions

1. How do campaign types differ in their choice of TTPs?
2. Can the motivation behind a campaign be predicted based on historical data?
3. What is the relationship between campaign characteristics and the targeted assets?

## **Threat Actors Dataset**

### Description

This dataset simulates information about threat actors, including actor ID, name, type, and motive.

### Possible Applications

1. Profiling and tracking threat actors for attribution.
2. Analyzing the strategies and motives of different types of threat actors.
3. Predicting potential threat actor activity based on historical patterns.

### Possible Research Questions

1. How do the motives of threat actors influence their choice of targets?
2. Can the type of threat actor be predicted based on observed TTPs?
3. What patterns exist in the temporal behavior of threat actors?

## Incidents Dataset

### Description

This dataset simulates information about cybersecurity incidents, including incident ID, type, affected systems, description, and timestamp.

### Possible Applications

1. Studying the impact of different incident types on organizations.
2. Developing and testing incident response strategies.
3. Analyzing the temporal evolution of incident occurrence.

### Possible Research Questions

1. How does the distribution of incident types vary across different industries?
2. Can the affected systems provide insights into the motives of threat actors?
3. What are the commonalities in incident descriptions for specific incident types?

# Predictive Analytics and Simulation

## Predictive Modeling

1. Can machine learning models predict the severity of incidents based on historical IoCs, TTPs, and vulnerabilities?

## Temporal Analysis

1. How do the temporal patterns of TTPs correlate with the occurrence of specific incidents over time?

## Threat Actor Attribution

1. Can machine learning algorithms attribute observed TTPs to specific threat actor types with a high level of accuracy?

## Vulnerability Risk Assessment

1. How well can predictive analytics assess the risk associated with different vulnerabilities and prioritize patching efforts?

## Malware Prediction

1. Can historical data on malware prevalence and characteristics be used to predict future trends in malware attacks?

## Campaign Impact Prediction

1. How accurately can predictive models assess the potential impact and scope of a cyber campaign based on early indicators?

## Integrated Analysis

1. Can an integrated analysis of IoCs, TTPs, and threat actor behavior improve the accuracy of predictive models for incident occurrence?

## Simulation Scenarios

1. How do different simulation scenarios, such as changes in threat actor motives or the discovery of new vulnerabilities, impact the predicted risk landscape?