## CyberSecurity Data Sources

## Classification of Cybersecurity Data Sources

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Data Type** | **Description** | **Examples** | **Advantages** | **Limitations** |
| Static (Offline) Data | Pre-collected and labeled datasets stored in files. Used for training ML models. | KDD Cup 1999, NSL-KDD, CICIDS2017, UNSW-NB15 | Easy to access, standardized, labeled | May not represent the latest attack patterns |
| Dynamic (Live) Data | Real-time data captured from active networks or simulated environments. | Captured via Wireshark, Zeek, or network sensors | Reflects real-world, current threats | Hard to label, privacy risks, high setup cost |

## 

## Overview of Commonly Used Datasets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Year | Source | Key Features | Suitable For |
| KDD Cup 1999 | 1999 | DARPA Intrusion Detection Evaluation | 41 features, labeled attacks | Basic intrusion detection research |
| NSL-KDD | 2009 | University of New Brunswick | Improved KDD dataset with reduced redundancy | Educational use, ML benchmarks |
| UNSW-NB15 | 2015 | Australian Centre for Cyber Security | 49 features, modern attacks | Deep learning-based IDS |
| CICIDS2017 | 2017 | Canadian Institute for Cybersecurity | 80+ features, realistic network traffic | Realistic enterprise-level intrusion detection |
| TON\_IoT | 2020 | UNSW Canberra | IoT device traffic, telemetry & attacks | IoT security AI models |

## Static vs Live Data: In-depth Comparison

|  |  |  |
| --- | --- | --- |
| Criteria | Static Dataset | Live Data Capture |
| Data Collection | Already available | Requires sensors, packet capture tools |
| Cost | Free / Open source | Expensive setup |
| Data Labeling | Pre-labeled | Manual or semi-automated |
| Privacy Risk | None | Possible data exposure |
| AI Compatibility | Directly usable for supervised ML | Requires preprocessing |
| Best Use | Training & validation | Real-time testing & monitoring |

## 