**Malicious List Matching**

**Python Code: Project\_DAS\preprocess\MaliciousListMatching\MaliciousURLParse.py**

A screenshot of a computer

Description automatically generated

1. Extract valid URLs from ['assessment']['analysis']['task']['static']['result']['urls\_check']['urls'] path of each app.
2. Convert these URLs to IP addresses.
3. Utilize Dshield Malicious List (<https://www.dshield.org/ipsascii.html>) to find out the malicious IP count for each app. Using Dshield API (<https://isc.sans.edu/api/>) to query “package count (total number of packets blocked from this IP)”, “attacks (number of unique destination IP addresses for these packets)” and “threatfeeds” for each IP address. For a specific IP, if (package count > 200 or attacks > 20 or threatfeeds == true), we identify this IP as a malicious IP and add 1 to malicious URLs count for the app.

A screenshot of a computer

Description automatically generated

1. Insert the malicious URLs count of the apps to the raw dataset for model calculations.

**DShield** is a community-based collaborative [firewall](https://en.wikipedia.org/wiki/Firewall_(networking)) log correlation system.[[2]](https://en.wikipedia.org/wiki/DShield#cite_note-2) It receives logs from volunteers worldwide and uses them to analyze attack trends. It is used as the data collection engine behind the SANS [Internet Storm Center](https://en.wikipedia.org/wiki/Internet_Storm_Center) (ISC).

<https://en.wikipedia.org/wiki/DShield>