

Uninvited Guests: Analyzing the Identity and Behavior of Certificate Transparency Bots - Dataset Artifact Submission

Schema

- request_timestamp: timestamp in which the request to our honeypot was received
- host: domain or IP address request was sent to
- path: URL path of request
- user_agent: HTTP User-Agent of requesting client
- ja3_fingerprint: JA3 TLS fingerprint of requesting client (if request occurred over HTTPS)
- client_ip: IP address of requesting client
- method: HTTP method of request
- scheme: HTTP(S)
- org: organization of client_ip ASN
- country_name: name of country associated with client_ip
- asn_type: category of client_ip
- cert_reg_ts: timestamp of certificate creation
- type: category of domain (i.e., sensitive, baseline, or impersonating)

Reproducibility Example

In order to assist in the evaluation of our dataset artifact, we have included a small Python script which generates a number of general statistics, as well as Table 2 from our paper. To run the script, first install the prerequisite libraries (the Python Pandas library) from the *requirements.txt* file, and then execute it as follows: `python3 parse_dataset.py`