Real-Time Monitoring of Denial of Service Attacks

Stefan Scholz

2020-07-22

Goals

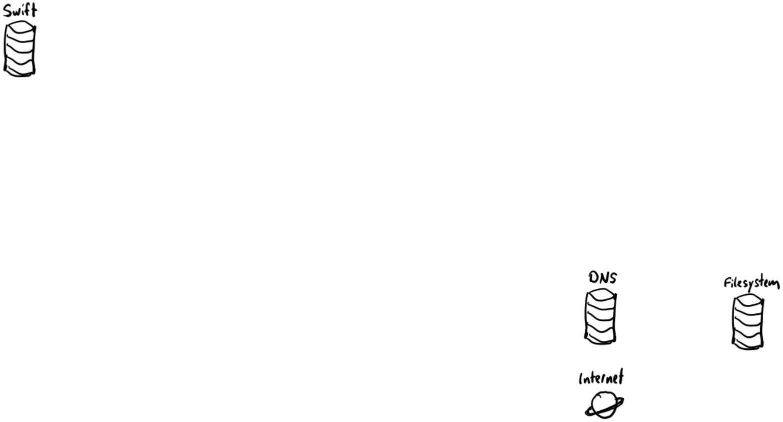
1. Track targets
2. Resolve hosts
3. Crawl hosts

System

- Platforms
  - Apache Kafka: distributed streaming platform
  - Faust: stream processing library, porting the ideas from Apache Kafka to Python
- Components
  - Externals: other systems, interacted outside this system
  - Records: objects, used within the system
  - Agents: processors, process messages in the system
  - Topics: queues, used to send and receive messages
  - Tables: dictionaries, store records in memory
- Settings

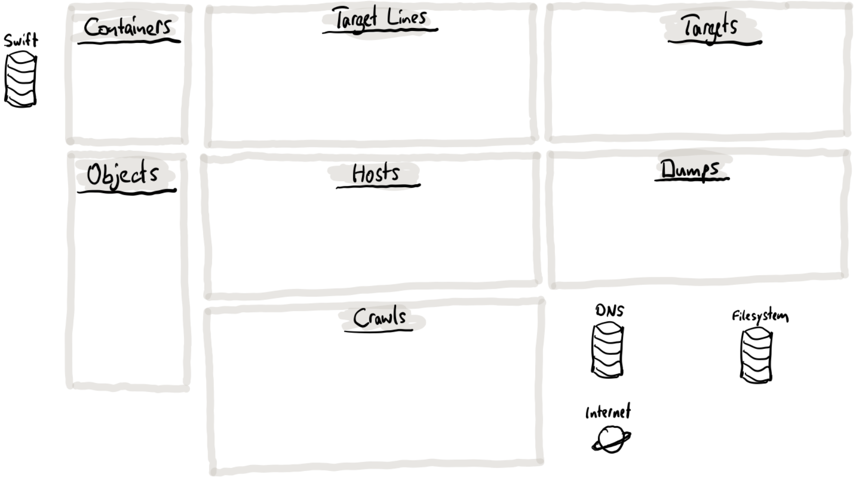
System

Externals



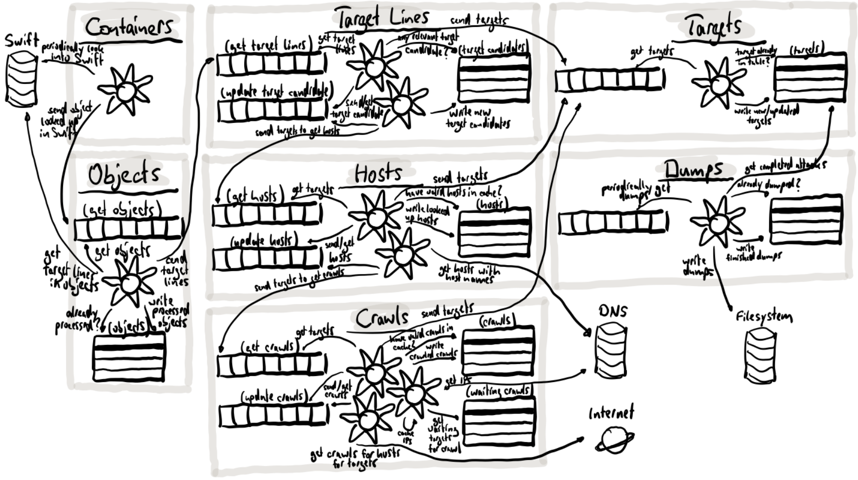
System

Records



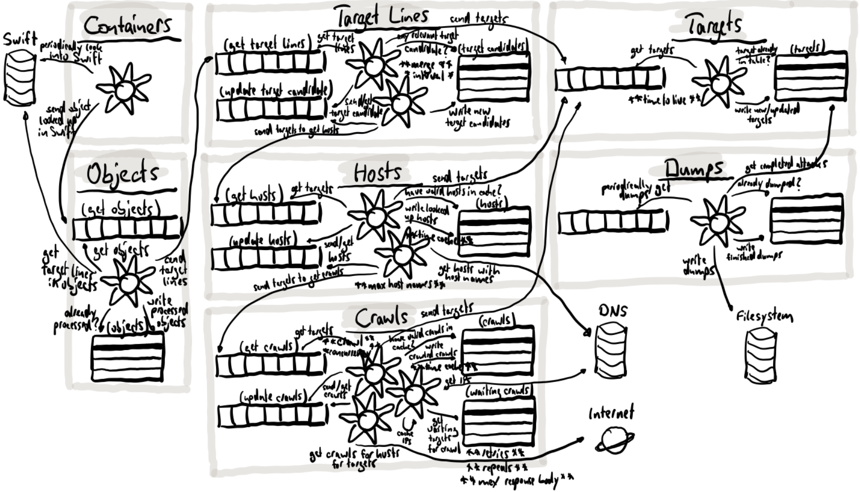
System

Agents, Topics and Tables



System

Settings



System

Settings

- Settings and questions 1
  - Target merge interval: what is maximum difference between start time of target line and latest time of target line that they are merged as one target?
  - Target TTL: how long shall a target live in the system, i.e. be crawled?
  - Host cache interval: how long can host names for IP address be cached?
  - Host maximum names: how many host names shall be possible for IP address?

System

Settings

- Settings and questions 2
  - Crawl request header: shall default header for request be changed?
  - Crawl response timeout: for how long shall be waited for response of host?
  - Crawl maximum body size: at which size is body of response from crawl truncated?
  - Crawl cache interval: for how long shall successful crawl be cached?
  - Crawl retries: how often retry crawl on failed crawl?
  - Crawl retries backoff: which coefficient shall be used for exponential backoff between retries?
  - Crawl repeat interval: in which intervals shall crawls be repeated?
  - Crawl concurrency: how many hosts can be crawled concurrently?

Data

- Format
  - General: data dumps, saved as JSON files, compressed with Gzip
  - Jupyter notebook
- Examples
  - Test data: randomly generated times and targets, one data dump, data-telescope-crawler-dos-202007201318.json.gz
  - Real data: recorded in testing on 2020-07-22

Analysis

- Attacks, targets, hosts and crawls
  - Jupyter notebook

Outlook

- Track targets: now target lines available in Kafka topic, rebuild system
- Resolve hosts: implement own reverse DNS resolution
- Crawl hosts: record additional information in crawls
- Settings: define values for settings