Interview Task

Satellite AIS message position tagging

Spire's technology stack makes extensive use of Python, and good Python knowledge and development experience is essential to be successful in this role. With that in mind, we would like to set the following task to assess your current experience and approaches to building a service in Python.

Background

Satellites orbiting Earth are tracked using a common data format known as a TLE (link), which contains information that can be used to determine the satellites' positions as they progress along their orbital paths. The TLEs are updated periodically to consider small adjustments in the satellite orbits. TLEs are identified by their satellite catalog number, also known as NORAD ID.

Most of Spire's satellites monitor AIS (link) traffic, listening to vessels broadcasting their position and other voyage data in messages of different types. It is useful to determine the satellite's location at the time an AIS message was received, for example to perform quality control: if the reported position in the AIS message broadcasted by the vessel is outside of the footprint of the satellite, that is a red flag and might for example be indicative of deliberate position spoofing. In this context, the satellite footprint can be approximated as a circle with a 3000 km surface distance radius around the sub-satellite point.

To represent AIS messages on the customer-facing API, Spire uses NMEA 0183 (link) sentences, with prepended metadata tags to encode for example receiver source and reception timestamp. Messages that are received from satellites are identified as such and tagged with the satellite's NORAD ID. You can read more about the details and find sample data at under the "Understanding the AIS message stream format" and "Sample TCP Stream results" headings here.

Task

Provide the implementation for a Python-based AWS Lambda function that:

- Is designed to be invoked periodically,
- upon invocation scans an AWS S3 bucket for unprocessed text files containing AIS messages according to the format also used on Spire's TCP stream, as described above,
- for each satellite AIS message in each of the files, o   determines the longitude and latitude of the sub-satellite point when the message was received,
    - o *[**non-mandatory stretch goal**] raises a flag if the AIS message contains the vessel's position and that position is outside of the satellite's footprint,*
- writes all processed satellite AIS messages to a file in another S3 bucket, augmented with the satellite longitude and latitude *and any red flag (non-mandatory).*

Constraints and assumptions:

- It is encouraged to use third party Python packages, for example: o   https://pypi.org/project/pypredict/ to determine satellite location from a TLE, and o   https://pypi.org/project/simpleais/ to extract position data from AIS messages in NMEA sentences.
- About 100k satellite AIS messages will need to be processed daily.
- Input data files can be deleted after processing.
- Use Spire's TLE service to get the TLE for a NORAD ID by sending a HTTP GET request for the URL http://tle.spire.com/<norad_id>. The service will respond with the requested TLE in the body of the HTTP response.
- Document and justify any additional assumptions, rationales, or improvement suggestions.