FROM SENSING TO MAKING SENSE

Edge Computing For Global Genomics-based Threat Detection, Assessment, And Decision Support

Rebecca Hutchins GTRI **Frazier Baker** GTRI **Collin Thomas** GTRI **Michael Lavelle** GTRI

Genomic sequencing provides a pathogen agnostic threat detection capability. Advances in sample processing and sequencing methods have led to field-ready kits, allowing for the generation of sequencing data outside traditional laboratory infrastructure. This sequence data can be processed using bioinformatic tools on field-deployable computing equipment to identify or categorize biological threats and provide insights in real time. Although DNA sequencing of a single sample can generate gigabytes of data, a laptop computer can be used to process the data in the field without an internet connection. These technological advances in field based sample processing, field based DNA sequencing, and field based data analysis pose a new problem: how to store, manage, and transfer the field data to gain insights and support timely decision making.

Using these technological advances, field forward operators will be able to make a threat identification based on sequence data and have immediate offline access to threat assessment information (e.g., pathogen modes of transmission, symptoms of infection, effective medical interventions). However, field operators may be geographically isolated from analysts and operational-level decision makers who would benefit from quick access to these data. Additionally, as new reference databases are released for the bioinformatic taxonomic classification tools, field laptops would benefit from receiving these updates to ensure field operators have the most up to date information regarding potential biological threats. To address these issues we propose a system for managing the sharing of summary data, sequence data, and reference databases over a global network designed to support both field forward operators and centralized command in gaining actionable insights quickly.

The proposed system leverages NoSQL databases which are distributed across a global network of datacenters. The proximity of the data centers will allow for faster transfer speeds during synchronization with field laptops. Additionally, NoSQL databases support schemaless data structures which allows for more fluid data formats as new technologies, biological threats, and indicators emerge. The taxonomic classification information (biological threat identification) coupled with key pieces of metadata (e.g., geographical location, date/time stamp, reference database version) makes up the summary data file. The summary data will be stored separately from the raw sequence data, enabling quick upload of the summary data containing key information for operational-level decision makers. The raw sequence data upload will follow when the field laptop can establish a high-bandwidth connection with a data center. After uploading the sequence information, the laptop can download the latest version of the reference sequence library before being redeployed to the field. Using the proposed process, biointelligence data analysis and insights can move to the edge with the warfighter while still sending valuable information back to commanders and analysts at the speed of relevance.