|  |  |  |
| --- | --- | --- |
| CSIS 486 | Riley Dorough  Kayla Echols  Julia Wilkins  Brett Wolff | Real Time AI and ML Processing for Initial Cyber Threat Intelligence Validation Testing |
|  | **April 9th , 2022** | **Phase 2 Step 5** |

Title [Kayla Echols]

* The goal for this lab is to get live data into the existing Knime pipeline, prove that it works correctly, and document the results.
* Another goal of this lab it to take steps of showing initial correlation between the OT data and how it could be potentially helpful information in the future.

Abstract [Kayla Echols]

The team was responsible for getting live data from the MS SQL that contains the database for the OT data and also live data for the IT data into Knime. The team also took steps towards initial correlation by using nodes provided in Knime to correlate the data that is in our pipelines. The correlation nodes that were used and documented in this lab show potentially relevant information that could be used in the future to predict anomalies. More research and documentation must be conducted for the correlation to be proven useful and helpful.

Introduction [Kayla Echols]

The purpose of this lab is to successfully establish real time data in our Knime environment using both IT and OT data. After researching, designing, and testing, our team was able to successfully get live data from our data sources to the Knime environment using a series of nodes. The team used a node to get live data from the MS SQL database from our Windows machine which must be refreshed to get the live data. After the team was able to successfully get live data, we then began researching how to correlate the data within Knime. We also used a node for this requirement of the lab there were two nodes that were tested including the Linear Correlation and Rank Correlation. The team also studied and discovered a way to flag anomalies using the Rule Engine to take a step forward in anomaly detection. Our team was able to successfully get live data and take more steps for future correlation for this project. In order to complete this lab our team needed access to Knime, the Liberty VPN, and the virtual machines in VSphere.

Method [Riley Dorough]

A screenshot of a computer

Description automatically generated with medium confidence

**Changes to Operational Nodes & Pipelines**

* Added/ Configured SQL Server Connector node
* Added/ Configured SQL Query Reader node
* Modified string manipulation as direct SQL queries do not convert Boolean values to string data type
* Added live/pseudo live data retrieval

**In Progress Pipelines, Nodes, and Experimental Functionality**

* Added initial Rule Engine nodes & initial weighted flagged anomaly reporting pipeline
* Added initial correlation matrix nodes for R&D into time-based correlation

Graphical user interface, text, application, email

Description automatically generated

* The SQL Server Connector node establishes a connection to our data historian located on another device within our network
* This must be connected to a data request node to recover data from the connected data historian

Graphical user interface, text, application

Description automatically generated

* The SQL Query Reader node Simply sends requests to the SQL Server Connector node to be passed to our data historian
* The data historian returns data to the SQL Server Connector node which is processed to the beginning of the pipeline where data manipulation begins
* This allows for live/ pseudo live data depending on how one defines the process. At the reset of the pipeline, all data is up to date to the most current reading from the OPTO systems delayed by the propagation of information to the SQL server and to Knime

Graphical user interface, text, application

Description automatically generated

* Error code assigned to flagged weighted values has begun to be experimented with
* The Rule Engine node currently assigns values to different anomalies to track where the anomalies occur
* Testing will continue to attempt to produce more functionality and usability

A picture containing chart

Description automatically generatedChart, scatter chart

Description automatically generated with medium confidence

* Various correlation matrix nodes have begun to be experimented with to produce useable metrics for validating correlations between different data types
* Effort will continue to test direct data correlation, anomaly correlation, and rate of change of data value and anomaly state for possible methods of delineating possible true correlation from coincidence

Results [Julia Wilkins]

This lab Team 1 had two goals. Get live data into Knime and set up the basics of a flagging system for anomaly detection. The first major issue encountered was trying to get live data into Knime. For OT data, we used a Knime node that would connect to the MS SQL database hosted on the windows box that hosts OPTO. After some help from Team 5 and solving some issues with the fire wall, Team 1 now has access to somewhat live data in Knime. The node does not actively stream data into Knime, but rather when refreshed will send a query to the database and update with new data. IT data works similarly. The PowerShell script is running as a service on the Knime box, and when the node is refreshed will update that table with new date.

This task of establishing anomaly flagging is a step forward toward in the overall goal of having a flagging system for anomalies in the pipeline that will eventually all compile into a report. So, for this deliverable, a basic flagging system for anomalist data was established. If data was in a bin that we as a team consider an anomaly, then a 1 is added to a flag column. The goal is that when compiling reports, it will make it easier to find data entries that contain potentially anomalous data. Next we will attempt to calculate the rate of change for each of the data points in the timeline. This will give a new facet to anomaly detection and all for more accurate flagging.

Discussion [Brett Wolff]

Conclusion [Brett Wolff]