Advanced AI solution for enterprise malware detection

An application that runs on an enterprise's datacentre storage systems (SAN) and does real time monitoring of the SAN I/O activity. It employs an advanced AI solution for anomaly detection that allows blocking malware from corrupting enterprise data.

Overview _

Client:
ProLion GmbH
Cybersecurity, anti-malware solutions
Vienna, AT

Business case:
+ Detect and prevent ransomware attacks
+ Monitor the storage usage

Industry _

+ IT Services
+ Data centers
+ Storage
+ Cybersecurity

Services _

+ Custom software development
+ Product development

Project type _

+ Web
+ Distributed backend

Technology _

+ Java
+ NetApp Clustered Data ONTAP
+ Hazelcast
+ REST endpoints
+ AWS virtualisation
+ Machine learning

Description _

An application that runs on an enterprise's datacentre storage systems (SAN) and does real time monitoring of the SAN I/O activity. It employs an advanced AI solution for anomaly detection that allows blocking malware from corrupting enterprise data.

Challenges _

Since malware can hit in many different forms and have a heavy impact on the final user, we have to:

–

+ Provide a powerful solution that protects against all threats (both known and new / unknown).
+ Ensure the best malware detection accuracy, while keeping false positives at minimum (or zero).
+ Deliver real-time detection and protection that spans across the whole SAN network.
+ Keep SAN performance unaffected.

Solutions _

We met client's high expectations with a series of cross-technology solutions:

–

+ AI anomaly detection techniques that determine what is "normal" traffic and allow it to pass while "suspicious" traffic is blocked.
+ Model training and evaluation with extensive real data, collected from production SAN logs.
+ Processing and enhancement of collected data set, to obtain an even greater synthetic "real-like" dataset.
+ Setting up of simulated SAN environments; and release of malware to collect footprints.
+ Model parameters tweaking, to ensure highest precision and recall scores.
+ Implementation of distributed architecture, with sensors on each SAN node and dedicated processing nodes to run the detection model.
+ Development of a home-grown decision tree variant that is both accurate and lightweight enough for the use case.
+ Hyperparameter tuning to minimize the model while maintaining the accuracy.