1. **Application: Bluedon AI Firewall**

According to the design principle of modulation, we wrap and regard the AI-based maldoc detector as one independent detection module that can be easily used in our security product lines. This module can then be easily integrated into our security products such as next-generation firewall. An interesting question here is: How Bluedon manages to apply AI technology seamlessly to a 30 years old security product?

In the current development of Network and Gateway Security products. The capability of performing malicious file scanning effectively and efficiently at layer7 (the network application layer) is the international standard. The industry has strict demand on this product feature. A good detection module should have (1) millisecond latency for single file detection; (2) 99% accuracy while maintaining FP rate to be less than 0.01%.

The reason behind the demand of low-latency is obvious: The module has sequentially been placed into the working pipeline, high-latency will lead to the increase of Packet Drop Rate(PDR) and occasionally data loss. This is strictly forbidden for security devices. In the past few years, as the rapid development of malwares, the formal industry best practice – the pattern matching engine has gradually fade away from the mainstream recently. Two main reasons can be roughly seen: (1) In order to meet the requirement of having high detection accuracy, a large number of security analysts are needed for pattern writing but this manual process is not scale at all; (2) As the fast growing size of the core database, the time for core operation – Pattern Matching (PM) grows exponentially. The above two strong pieces of evidence inspire us to discover better engine rooted from AI.

By the year of 2018, we have managed to integrate our AI maldoc detector into firewall inline, in the hope of replacing the old engine. Although both of the 2 engines belong to the category of static analysis based engine, the improvement from shifting from the old to the new AI engine is tremendous. On one side, the AI engine do NOT need to update frequently, since it can detect unseen malwares effectively for years. According to our experiment results, the average updating frequency for our AI engine is half a year as compared to 2 weeks for the old pattern matching engine; On the other side, AI engine enjoys low resource consumption when execution. According to our study, during the phase of model prediction, AI engine can ONLY take up as much as 1/3 of CPU and 50% of memory consumption. The portion from CPU is mainly due to the computations such as feature extraction and confidence score computation. The portion from memory is mostly due to the fact that AI model is needed to be sited entirely in main memory when for prediction.

In the context of firewall, different actions are triggered based on the probability and reasons output from the AI maldoc detection module. For instance, if the output probability is great than a certain threshold say 0.9, this indicates the AI module has high confidence that this document is malicious, a blocking operation is triggered, connection is dropped and an alert is raised for further investigation; If the output probability is less than a certain threshold say 0.1, this indicates that the AI module has high confidence that this document is benign, we will by default allow and monitor this connection as normal.

The truly interesting part lies when the output probability is in the range from 0.1 to 0.9. If this happens, we will by default upload the samples to our Threat Intelligence Cloud where multiple dynamic analysis will be performed with Sandbox, Threat Intelligence and Security Team. According to our heuristics in a typical usage scenario, 95% of the files are being processed inline while about 5% of the files are uploaded. Dynamic analysis from our TI cloud plays a great complementary role to the static analysis method inline. By means of combination between the two, we can now completely offer the end users the more advanced AI-enabled security solution. We make our cloud service a subscription service and freely open the research community.