Machine Learning
Technical Brief

Overview

In the current cybersecurity landscape, threat actors are always probing and constantly switching tactics, making companies susceptible to malware incidents and outbreaks, business disruption and data breaches. Bitdefender GravityZone Endpoint Security Platform defends your endpoints from the full range of sophisticated cyber-attacks with high efficacy, low end-user impact and low administrative overhead. It consists of multiple layers of defense that erect obstacles for the bad guys to make sure they keep stumbling. Each layer is designed to stop specific types of threats, tools, or techniques, covering multiple stages of the attacks.

Bitdefender leverages machine learning across its entire portfolio. Scanning engine, HyperDetect, Sandbox Analyzer, Content Control, Global Protective Network are a few examples of Bitdefender technology that makes use of machine learning. This document primarily focuses on Machine Learning based Threat Detection (scanning engine). Bitdefender Machine Learning based Threat Detection is part of the GravityZone Endpoint Security Platform. It provides protection against zero-day threats in the pre-execution stage.

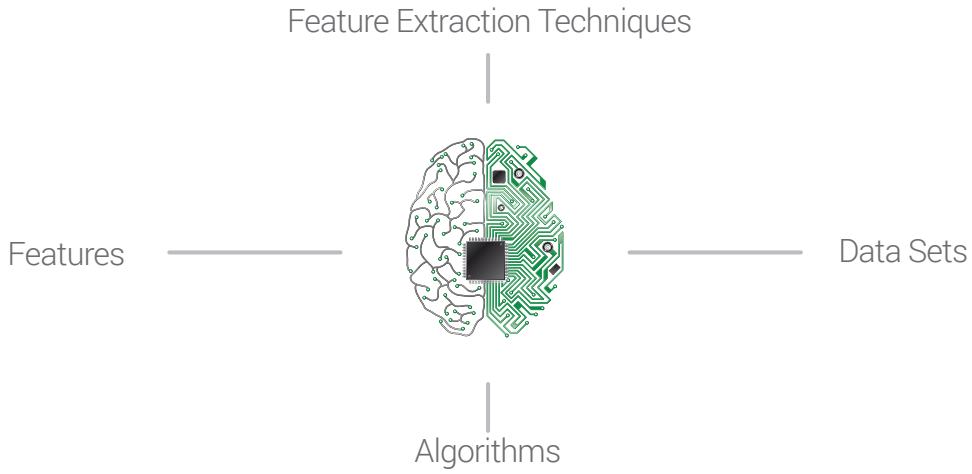
Detection Stage	Technology Type	Threat Coverage
Pre-Execution	Machine learning	File-based malware, Trojans, Password stealers, Exploits, Obfuscated malware, Targeted attacks, Script-based attacks, Mutating and Polymorphic malware, Ransomware

Understanding the importance of Machine Learning technology

Machine learning is the ability of computer programs to analyze big data, extract information automatically, and learn from it. In cyber security, it can play an important role because it can predict an object's (e.g. a file or URL) malicious intent without any previous knowledge of the object. Bitdefender's patented machine-learning technology uses well-trained algorithms - some of them specialized in specific forms of attacks and others more generic, to predict, detect and block zero-day threats.

The key ingredients of Bitdefender's Machine Learning technology:

- **Features**: A feature in Machine Learning is an individual measurable property of a phenomenon being observed. Bitdefender extracts both static and dynamic features of files and URLs. Bitdefender's deep understanding of malware behavior allows it to identify the right set of features.
- **Feature Extraction Techniques**: Bitdefender's uses purpose-built emulator, unpacking, de-obfuscating techniques to extract static and dynamic features of files and URLs.
- **Machine Learning algorithms**: An algorithm in machine learning is a program that derives insight from the data. Bitdefender leverages several different algorithms. These algorithms also have a level of overlapping functions that make them more resilient against advanced attacks. It also includes custom machine learning algorithms to enhance accuracy of detection.
- **Data Sets**: In machine learning, data sets are very important in training and testing of machine learning models. Bitdefender possesses one of the industry's largest databases of clean and malicious file samples to train and test machine learning models, dramatically improving detection efficacy and accuracy.



Features

- Local Machine Learning models as well as Cloud Machine Learning models for malicious file and URL detection.
- Multiple Machine Learning algorithms with more than 75,000 models, including: Perceptrons , Binary Decision Trees, Restricted Boltzmann Machines, Genetic Algorithms, Support Vector Machines, Artificial Neural Networks, Custom algorithms for false positives mitigation, More than 40,000 static and dynamic features

Some examples of features Bitdefender extracts from a file:

- The unpack code contains strings that might indicate system persistence
- File is packed with an unknown packer
- File is obfuscated (unknown packer, unknown compiler)
- Abnormal use of different assembly instructions (like call, jump, etc.)
- Multiple Feature Extraction Techniques: **Emulator**: Emulates the code (assembly instructions), see what it does, its intent and extracts features; **Unpacker routine**: Purpose-built unpacker routine that can extract dynamic features such as strings, code, html scripts injected, URLs etc.; **Cryptographic filters**: Applies cryptographic filters to extract features from encrypted data;
- Comprehensive data sets for training and testing machine learning models: Fresh samples; Varied malware; Representative malware; Unsupervised machine learning in the Cloud

Benefits

- Detect advanced attacks early and prevent breaches, reduce incident response costs and efforts
- Reduce threat-hunting burden
- Machine Learning greatly increases the detection rate of zero-day threats in pre-execution stage including file-based malware, trojans, password stealers, exploits, obfuscated malware, targeted attacks, script-based attacks, mutating and polymorphic malware, ransomware
- Local machine learning model ensures protection of offline devices.
- It is part of a single, integrated endpoint security agent and central management platform, greatly reducing administrative burden. Customers don't need to deploy a mixture of endpoint security solutions

BD-NGZ-Dec05.20.17-IT-#
Bitdefender-2017-TechnicalBrief-MachineLearning-creat2103-A4-en_EN-2