*SSH DICTIONARY ATTACK DETECTION* (ca. 6-9 months)

About QSight IT:

With it's 160 employees QSightIT is one of the largest Cyber Security companies in the Netherlands with an absolute focus on security, technology integration, consultancy and maintenance of vital parts of our customer's IT infrastructures. Our customer's are among others in the fields of: finance, government, commercial enterprises and health. We co-operate with technology partners who are recognized by the market as world leaders.

SSH dictionary attacks:

Within our R&D departement we currently have an assignment for an enthusiastic person in our Data Science team concerned with the deployment of data science as a means to protect our customer's

SSH means Secure SHell. It is an application, that lets the user interact (remotely) with the a system. The most well-known examples are bash, zsh or tsh, which are frequently seen on Linux systems. SSH attacks are fraudulent attempts to compromise a system through a SSH connection. These attacks are a concern for network managers, because of the potentially large risk of a successful attack. A specific category of this type of attacks is the so-called "dictionary attack". This involves trying a large number of username-password combinations on the (remote) system to gain access to it. It is shown that this type of attack can be successfully recognized from analyzing network data.

An algorithm based on network data patterns for this purpose was published by the University of. This algorithm uses network (meta) data in the NetFlow format, described on Wikipedia.

This algorithm employs some fixed-value thresholds, to distinguish between an attack and normal traffic. This "one size fits all" approach is not practicable for QSight IT, since QSight IT deals with a wide variety of large networks. Therefore, QSight IT is looking to improve upon the algorithm by making it "learn" these fixed-value parameters.

The phases of a dictionary attack:

These attacks usually take about an hour and are generally divided into three phases:

Phase 1: Scanning

The attacker scans an IP address block to find hosts that run a SSH daemon.

* Many /small/ flows to a /large/ number of hosts

·A /large/ number of connection attempts in a /short/**time interval

Phase 2: Brute-force

The attacker tries to login to one or more hosts, using a large number of username-password combinations

* The number of packets per flow falls into a specific /range/

·The number of flows is /larger/ than a certain value per /time interval/ per attacker
Phase 3: Die off

The attacker has succeeded to login and is executing commands on the (now) compromised system.

* The number of packets per flow will either be /less than /or /more than/ a certain number

The project:

The project will start with an orientation phase, in which the candidate should understand the published algorithm. The algorithm should be implemented and tested. The test results will be presented to the Data Science team as well as to the operations department, which will take the product in production. This first phase is expected to last about a month. This includes coaching of the candidate as well as some tutoring on basic data science best practices and necessary scripting languages, if needed.

If the results are satisfactory, then a first product version be deployed in a limited scope. This deployment phase is an important part of the assignment.

Further development of the project will take place in an iterative manner. Each iteration should lead to an improvement of the deployed product. The iterations will be made as small as possible, which greatly enhances the chances for measurable success. It is expected that the project may be finished in 6-9 months.

Tooling:

QSight IT has a Hortonworks Hadoop state-of-the-art big data platform. Net flow data is ingested in real-time and stored on HDFS. Spark (with Python bindings) is used to perform real-time and batch analyses of the data. Machine learning and numerical libraries include Keras, scikit-learn and numpy as well as MLib for Spark.

QSight IT develops software using SCRUM. The software management suite is provided by Atlassian. Research and development tasks are managed using JIRA. Stash is used to access the GIT repository and perform code reviews. Bamboo is used as build and deployment server.

Development takes place on a virtual box for each individual developer. QSight IT has near-identical acceptance and production environments, which are continuously monitored by devops personnel.

Note:

Background screening may be part of the procedure.