ScienceDirect

Network Security
Volume 2020, Issue 3, March 2020, Pages 8-14

Feature

Advanced malicious beaconing detection through AI

Yessine Borchani

Show more

Share Cite

https://doi.org/10.1016/S1353-4858(20)30030-1
Get rights and content

As efforts to more securely protect the world's privacy and data continue to improve, with the introduction of stricter compliance regulations and the deployment of increasingly complex network infrastructures, so too has enterprise adoption of and reliance on encryption. Cryptographic encryption protocols, namely secure sockets layer (SSL) and its successor, transport layer security (TLS), were estimated by Gartner to be implemented across 80% of enterprise web traffic in 2019, and the Ponemon Institute found that 43% of organisations had a consistent, enterprise-wide encryption strategy in place in 2018.1, 2

As efforts to more securely protect the world's privacy and data continue to improve, so too has enterprise adoption of and reliance on encryption.

However, threat actors have also started to leverage encryption and are hiding their nefarious activities among regular encrypted traffic, making malicious packets nearly impossible to detect. Fortunately, by using AI, enterprises can achieve a quicker and more accurate analysis of all traffic on their network to check for beaconing behaviour, explains Yessine Borchani of Barac.

Section snippets

Malicious beaconing

These encryption-based attacks take many forms. One of the most concerning types of malware where hackers are using encryption to hide their activities is malicious beaconing – or, more specifically, command-and-control (C&C) malware. Cyber criminals are using this method to infiltrate protected networks, commanding their malware to perform anything from stealing data from the local system to initiating distributed denial of service (DDoS) attacks. Gartner predicts that over 70% of malware …

Detection problems

Beacon analysis is an arduous and often inaccurate process, a problem that will only be compounded as levels of encryption continue to grow. Collecting and reviewing enough data to detect beaconing behaviour – such as the frequent calls home to the C&C server and the regular beaconing packet sizes – as well as sorting out the false positives from legitimate threats, is extremely challenging. As such, many tools struggle to keep up, leaving networks vulnerable for extended periods of time.

Threat …

New protocol

A further, recent development that impedes the decryption method of incoming data is the introduction of the new TLS 1.3 protocol. Ratified by the IETF in August 2018, and replacing its predecessor TLS 1.2, TLS 1.3 delivers security and speed improvements by eliminating unnecessary handshake steps and by adopting newer encryption methods.

On one hand, this new protocol boosts privacy through stronger encryption protocols and streamlined authentication processes in order to strengthen the privacy …

How AI solves the problem

Traditional algorithms are easy targets for sophisticated cyber criminals and cannot keep pace with attackers who can code their way around such defences. Every time an attacker develops a new evasion tactic, the entire algorithm becomes obsolete and ineffective and has to be rewritten. This is a costly and time-consuming endeavour, and one that runs a risk of not being able to detect beaconing behaviour at all.

Instead, organisations need to leverage the benefits of unsupervised machine …

The process

Starting with raw data (Table 1), the rows are grouped into source/destination couples, the source being the IP address of the client and the destination being either an IP address or a DNS name. When using the DNS name, it must be taken into consideration that a server, potentially an attacker, can change its IP address to avoid detection. Then, for each group, a generic feature is selected from the raw data, for example, the mean of fragment sizes or time intervals. Table 2 illustrates the …

Number of clusters

For the spatial model, based on the k-means algorithm, an automation approach for the elbow method is used**.** The elbow method is defined as follows: for each value of k, the SSE (sum of squared errors) is calculated and then plotted. When the variability of the SSE no longer appears significant compared to the previous value, the ideal number of clusters has been reached. In Figure 2, the ideal number of clusters can be considered as six.

The automation of this process is rather simple. The SSE …

Meaning of the clusters

To the human eye, it may seem clear which clusters are infected and which are clean, something that can be judged by the mean values of each cluster or centroid, and by the scatter plot of the points constituting the dataset. To a machine, all of these variants are values, so a method needs to be introduced that can determine the meaning of each cluster and merge similar clusters to ultimately reveal whether the beaconing is severe, moderate or simply normal traffic.

Here, the hierarchical …

The results:

Three types of malicious beaconing behaviour can be detecting using this algorithm:

- **Beaconing with distortion:** In this case, time intervals' values vary ever so slightly, caused by the addition of a random number of milliseconds each time by the server. For example, the following array of time intervals can be detected: [284, 300, 320, 296, 307]. All values appear to be around 300, with an arbitrary value added each time. …
- **Beaconing with skipped exchanges:** In this instance, an array of time …

…

AI: the future

As the use of malware continues to grow and threat actors find new ways of exploiting encrypted data flows, enterprises need to look to new technologies to be able to more effectively protect against the looming and dangerous threat of malicious beaconing.

By using AI – unsupervised machine learning clustering algorithms in particular – enterprises can achieve a quicker and more accurate analysis of all traffic on their network to check for beaconing behaviour, looking for variability in time …

About the author

Yessine Borchani is a data scientist at Barac (www.barac.io ↗) and is also studying for his engineer's degree at Insat in Tunis, Tunisia. A software engineer specialising in data science and data engineering, he is currently working on automating the detection of beaconing in encrypted TLS traffic. Previously, he has worked on genetic algorithms for software refactoring, deep learning for computer vision and unsupervised machine learning for malicious behaviour detection in …

…

…

Recommended articles

References (11)

‘Protecting from growing attack vector: encrypted attacks’
(2016)

‘2018 Global Encryption Trends Study’
(April 2018)

‘Transparency Report: HTTPS encryption on the web’
(Nov 2019)

‘Guide to GDPR: Encryption’. ICO

‘Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025’. Statista

There are more references available in the full text version of this article.

Cited by (5)

APT beaconing detection: A systematic review

2022, Computers and Security

> *Citation Excerpt :*
> …They developed a classifier to extract feature information and store data in the database, such as DLL file names and numbers, special API function names and numbers, special registered information and system paths in the process memory. Moreover, some of the studies (Ghafir et al., 2017; Borchani, 2020; Liu et al., 2012, 2013; Rass et al., 2017; Ghafir et al., 2018; Balduzzi et al., 2013), used Bro passive, which is an open-source software for analyzing traffic, along with other tools to capture and monitor traffic. These tools are typically used to monitor security by carefully examining all traffic on a given link for signals of suspicious activity.…

Show abstract

Global Analysis with Aggregation-based Beaconing Detection across Large Campus Networks ↗

2023, ACM International Conference Proceeding Series

Comprehensive Analysis of IoT Malware Evasion Techniques ↗

2021, Engineering, Technology and Applied Science Research

Using deep packet inspection in CyberTraffic analysis ↗

2021, Proceedings of the 2021 IEEE International Conference on Cyber Security and Resilience, CSR 2021

A Review of Intrusion Detection Systems in RPL Routing Protocol Based on Machine Learning for Internet of Things Applications ↗

2021, Wireless Communications and Mobile Computing

About the author

Yessine Borchani is a data scientist at Barac (www.barac.io ↗) and is also studying for his engineer's degree at Insat in Tunis, Tunisia. A software engineer specialising in data science and data engineering, he is currently working on automating the detection of beaconing in encrypted TLS traffic. Previously, he has worked on genetic algorithms for software refactoring, deep learning for computer vision and unsupervised machine learning for malicious behaviour detection in encrypted network traffic.

View full text