Evaluating Reputation of IP Addresses

Václav Bartoš
CESNET a.l.e.
Zikova 4, 160 00 Prague, Czech Republic
bartos@cesnet.cz

Keywords. Network security, alert sharing, reputation, blacklists, machine learning

Abstract

Security monitoring tools, such as honeypots, IDS, behavioral analysis or anomaly detection systems, generate large amounts of security events or alerts. These alerts are often shared within some communities using various alert sharing systems (such as Warden, AbuseHelper, n6, MISP, *etc.*). Number of alerts processed by such sharing systems may go up to millions per day [1].

We build a large database of IP addresses which combines information from such an alert sharing system with information from many other data sources.The database, called *Network Entity Reputation Database (NERD)*, maintains a record for each IP address reported as malicious, stores meta-data about all the alerts and enriches it with information abouts IP address' autonomous system, geolocation, reverse DNS record, its presence on blacklists, connection type, *etc*. All this data is available to users in full detail and can be used, for example, to get information about IP addresses during incident investigation. However, for many use-cases such detailed and diverse data are not suitable. A summarization of all the data in a record into a single number (or a small set of numbers) is needed, as it allows to get quick overview of an IP address or to rank the addresses and create top lists.

We propose such a summarization called *reputation score*. It evaluates the level of threat the IP address poses, *i.e.* the probability of future attacks originated by the address and severity of such attacks. It is therefore based on a prediction of future behavior of IP addresses. The prediction is based on IP address' past behavior and all the other data known about it.

Such scoring can then be used, among other things, to create lists of the most dangerous IP addresses and block their traffic in a network, which should result in preventing large portion of incoming attacks with a low risk of blocking legitimate traffic. Blocking based on the reputation score should provide better results than simply blocking all IP addresses detected in recent past, since the reputation score expresses the expected future behavior and may take into account things like dynamic IP addresses (which have lower probability of recurring attacks since the host behind the address may change often) or TOR exit nodes (which are frequent source of some attacks).

In order to capture all such relations in data, even those not obvious to a human expert, the prediction algorithm is to be learned by machine learning methods. This is currently work in progress. We experiment with various types of neural networks and try to predict for each IP address the expected number of alerts in the next 24 hours for each of several attack types. The preliminary results indicate that such a prediction is possible with acceptable error rates and therefore that the reputation score can indeed be effectively used for creating blocklists.

References

[1] Bartoš, V.: Analysis of alerts reported to Warden. Tech. Rep. 1/2016, CESNET