Politecnico di Torino

SmartData@PoliTO

SmartData

November 22nd, 2021, 5:30 PM CEST

SmartTalk: Covivio

Luca Gioacchini

Ph.D Student – Politecnico di Torino

DarkVec: Automatic Analysis of Darknet Traffic with Word Embeddings

ABSTRACT

Darknets are passive probes listening to traffic reaching IP addresses that host no services. Traffic reaching them is unsolicited by nature and often induced by scanners, malicious senders and misconfigured hosts. Its peculiar nature makes it a valuable source of information to learn about malicious activities. However, the massive amount of packets and sources that reach darknets makes it hard to extract meaningful insights. In particular, multiple senders contact the darknet while performing similar and coordinated tasks, which are often commanded by common controllers (botnets, crawlers, etc.). How to automatically identify and group such senders that share similar behaviors remains an open problem. We here introduce DarkVec, a methodology to identify clusters of senders (i.e., IP addresses) engaged in similar activities on darknets. DarkVec leverages word embedding techniques (e.g. Word2Vec) to capture the co-occurrence patterns of sources hitting the darknets. We extensively test DarkVec and explore its design space in a case study using one month of darknet data. We show that with a proper definition of service, the generated embeddings can be easily used to (i) associate unknown senders' IP addresses to the correct known labels (more than 96% accuracy), and (ii) identify new attack and scan groups of previously unknown senders.

SmartTalks Season 2