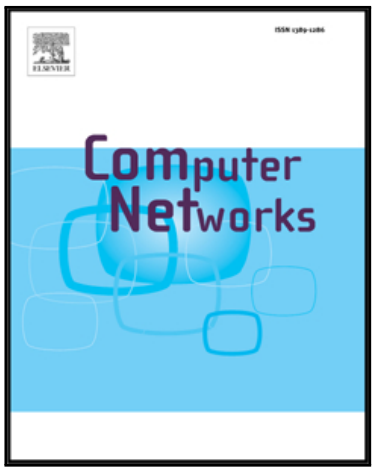

BotCluster: A Session-based P2P Botnet Clustering System on NetFlow

Chun-Yu Wang , Chi-Lung Ou , Yu-En Zhang , Feng-Min Cho , Pin-Hao Chen , Jyh-Biau Chang , Ce-Kuen Shieh

PII: S1389-1286(18)30835-1
DOI: https://doi.org/10.1016/j.comnet.2018.08.014
Reference: COMPNW 6573

To appear in: *Computer Networks*

Received date: 23 April 2018
Revised date: 3 August 2018
Accepted date: 30 August 2018

Please cite this article as: Chun-Yu Wang , Chi-Lung Ou , Yu-En Zhang , Feng-Min Cho , Pin-Hao Chen , Jyh-Biau Chang , Ce-Kuen Shieh , BotCluster: A Session-based P2P Botnet Clustering System on NetFlow , *Computer Networks* (2018), doi: https://doi.org/10.1016/j.comnet.2018.08.014

BotCluster: A Session-based P2P Botnet Clustering System on NetFlow

Chun-Yu Wang[1], Chi-Lung Ou[1], Yu-En Zhang[1], Feng-Min Cho[1], Pin-Hao Chen[1], Jyh-Biau Chang[3], Ce-Kuen Shieh[1,2]

[1]Institute of Computer and Communication Engineering, Department of Electrical Engineering, National Cheng Kung University,Tainan, Taiwan

[2]National Center for High-Performance Computing, Taiwan

[3]Department of Digital Applications, University of Kang Ning, Tainan, Taiwan

wicanr2@gmail.com, ou0peter@gmail.com, q36054316@mail.ncku.edu.tw, tgbv41@gmail.com, q36051164@mail.ncku.edu.tw, andrew@ukn.edu.tw, shieh@ee.ncku.edu.tw

Abstract

This study presents a Session-based P2P Botnet Clustering system implemented on MapReduce for aggregating malicious hosts within NetFlow traffic logs. The proposed botnet detection system, designated as BotCluster, merges the unidirectional records of NetFlow into bi-directional sessions and then utilizes a 3-level grouping to cluster similar sessions into groups with a like behavior. Besides, BotCluster would eliminate unrelated sessions and keep the large irregular sessions using the similarity and regularity of Botnets in their communication nature. The clustered groups can be considered as malicious behavioral collections because only man-made malware would generate the large of the similar pattern in network traces. The performance of BotCluster is evaluated using real-world NetFlow traffic logs collected from two university campuses in Taiwan (i.e., NCKU and CCU). The datasets have sizes of 239 GB and 137 GB, respectively, and contain a total of approximately 2.4 billion flows and a total of approximately 18 million IP address. The precision of the BotCluster detection results is evaluated using the VirusTotal blacklist service. It is shown that BotCluster achieves a detection precision of 96.23% and 86.62% for the NCKU and CCU datasets, respectively. Finally, when applied to a combined dataset containing the NetFlow logs of both campuses, BotCluster achieves an average precision of 97.58%. In other words, given sufficient observation duration, BotCluster provides the ability to detect even stealthy and concealed bots with a high degree of reliability.

Keywords *— P2P Botnet Detection; Botnet Detection; Netflow; MapReduce;*