Scalable machine learning-based intrusion detection system for IoT-enabled smart cities

Md Arafatur Rahman[a], A. Taufiq Asyhari[b], L. S. Leong[a], G. B. Satrya[c], M. Hai Tao[d], M. F. Zolkipli[a]

[a] Faculty of Computing, University Malaysia Pahang, Malaysia

[b] School of Computing and Digital Technology, Birmingham City University, Birmingham, B4 7XG, United Kingdom

[c] School of Applied Science, Telkom University, Bandung, Indonesia

[d] Department of Computer Science, Baoji University of Arts and Sciences, Shaanxi, China

ABSTRACT

Given a scale expansion of Internet of Things for sustainable resource management in smart cities, proper design of an intrusion detection system (IDS) is critical to safeguard the future network infrastructure from intruders. With the growth of connected things, the most-widely used centralized (cloud-based) IDS often suffers from high latency and network overhead, thereby resulting in unresponsiveness to attacks and slow detection of malicious users. In this paper, we address the limitation of centralized IDS for resource-constrained devices by proposing two methods, namely semi-distributed and distributed, that combine well-performing feature extraction and selection and exploit potential fog-edge coordinated analytics. In order to distribute the computational tasks, we individually develop parallel machine-learning models corresponding to a partitioned attack dataset. In the semi-distributed case, the parallel models, running on the edge side, are applied for side-by-side feature selections, which are then followed by a single multi-layer perceptron classification running on the fog side. In the distributed case, the parallel models individually perform both the feature selection and multi-layer perceptron classification after which the outputs are combined by a coordinating edge or fog for final decision making. Based on the comparative study of existing works, the numerical results demonstrate the promise of the proposed methods, giving a comparable detection accuracy to the superior centralized IDS as well as exemplify their inherent trade-offs between the accuracy and building time performance.

KEYWORDS

Feature extraction; Feature selection; IDS; Intrusion detection; IoT; Machine learning; Semi-distributed; Distributed

ACKNOWLEDGEMENT

This work was partially supported by a Research Grant (RDU192202), which is funded by University Malaysia Pahang.