Where are you?

Real-time tracking and prediction of user movement in computer networks

Introduction

In today's computer networks there is a lot of diverse traffic and if we analyze it, we can infer user's behavior, monitor the network functionality and detect malicious activity. The user of the computer network can either be a person or a computer.

Related work

Currently there are specialized research fields for malicious activity detection, user activity inference and network monitoring. All these three fields have seen significant development in recent years.

The analysis of user behavior ranges from simple statistics, like classification of traffic, distribution of users over a day and over different access points, to more complex machine learning algorithms. They analyze user movement from their log-in credentials or network addresses of user devices, however, to best of our knowledge no complex analysis of user movement in computer networks has been conducted.

Idea

Current solutions inspect network traffic independent from distinguishing between users and omit the cases, where users change their network access point or new users start using the same access point.

We propose a deeper analysis of user activity with emphasis on their spatial movements and temporal patterns over multiple access points.

This analysis could create additional knowledge, with which we will be able to detect, describe, predict and prevent advance, complex and higher level malicious activity, social behavior and monitoring of the network itself. The knowledge acquired through this kind of analysis would bring new competitive edge to enterprises.

Our approach

We propose a framework for analysis of user activity in enterprise networks that will use user activity matching in addition to network addresses and log-in credentials. It will be based on incremental clustering with added concept drift adaptation and cluster re-initialization.

We will present each cluster with a vector of attributes, that we store outside of the clustering algorithm. We will then use nearest neighbor search algorithm to match the user activity to their previous activity independently of the time difference and network access point. To solve issues with large volume and velocity of traffic, we will use flows to aggregate our input data and for storage and computation, we will use on-line distributed system.

Empirical evaluation of our solution will be done in real-time on our university computer network.

References:
[1] Silva, J. a, Faria, E. R., Barros, R. C., Hruschka, E. R., Carvalho, A. C. P. L. F. De, & Gama, J. (2013). Data stream clustering. ACM Computing Surveys, 46(1), 1–31. http://doi.org/10.1145/2522968.2522981
[2] Liao, H.-J., Richard Lin, C.-H., Lin, Y.-C., & Tung, K.-Y. (2013). Intrusion detection system: A comprehensive review. Journal of Network and Computer Applications, 36(1), 16–24. http://doi.org/10.1016/j.jnca.2012.09.004
[3] Papadopouli, M., Shen, H., & Spanakis, M. (2005). Modeling client arrivals at access points in wireless campus-wide networks. 14th IEEE Workshop on Local and Metropolitan Area Networks, LANMAN 2005, 1–7. http://doi.org/10.1109/LANMAN.2005.1541514

Author: As. Aleks Huč, MSc
aleks.huc@fri.uni-lj.si
Mentor: Prof. Denis Trček, PhD
Laboratory for e-media
Faculty of Computer and Information Science
University of Ljubljana