Users in organizations regularly access various internal and external computational resources. Such user activity is logged as events by various devices (firewalls, DLP systems, IDSs, routers, antivirus, VPN, servers, etc.) These events are then collected by Security Information and Event Management (SIEM) systems for further processing and analysis in an attempt to detect cyber attacks. Specifically, users’ behavioral profiles can be derived based on the collected events in order to detect anomalies or malicious activity.

Previous studies proposed and evaluated methods for intelligent data analysis, specifically profiling users’ activity, in order to identify abnormal behavior. However, existing methods are incapable of dealing with advanced attacks that are able to “stay below the radar” and hide malicious activity within legitimate activity and thus evade such detection mechanisms.

In this research we develop a method that is based on machine learning techniques combined with statistical analysis for deriving users’ behavioral profiles based on the collected events in order to detect long term trends and anomalies (low amplitude anomalies).

While the collected events are always logged with the source IP address they are not always logged with the relevant username (used as the identifier) and therefore, many of the collected events are not directly linked with the appropriate user. In this research we also describe a method for associating an IP address with an actual username based on a set of logged events. This is a crucial precondition for generating an accurate user’s profile. The proposed method was evaluated using real large datasets (logs) and showed 88% accuracy in the identification of usernames.