Write Up Of Project

In today's highly connected world where more and more organizations are setting up online and connecting to their customers over the internet, the data generated and stored by them is growing at an exponential rate. The data collected by the organization is critical for its day to day operations and future decision-making policies, This makes data a crucial asset for the organization. Therefore, the organization must take steps to ensure the integrity, confidence, and availability of the systems used to store the data.

This issue is further enhanced by the increase in web applications and information systems that increases the risk of exposure of the databases, therefore, database security is more crucial today than ever before. Through this project, I tried to develop a system to ensure database security for various organizations, like banks and hospitals.

It is also vital to address the increasingly important issue of insider threats to the organization. In general insider threats are more harmful and dangerous than outside threats like hacking, malware, etc. This is because an insider already has the proper authorization to access the database and knows the nuances of the security systems put in place. This increases difficulty to identify malicious transactions made by an insider in comparison to outside threats. Through this project, I intended to develop an effective technique to identify malicious transactions from both inside and outside threats.

In this project, I utilized various data mining techniques like Sequential Pattern Mining(SPM) and Association Rule Mining(ARM) to develop data dependencies among the various attributes in the database. These data dependencies can then be used to check if a new transaction follows them, depending on which the transaction may be classified as either a malicious or safe transaction. In addition to this, I also made use of anomaly detection techniques to find patterns in the transactions that can be used to identify whether the transaction is malicious or safe. In the end, I combined the result obtained from the two systems described above to finally assign an anomaly score to the transaction and label it as either malicious or safe.

Sequential Pattern Mining or SPM is a topic of data mining concerned with finding statistically relevant patterns between data examples where the values are delivered in a sequence. In this, all the sub-sequences that satisfy the minimum support in a set of sequence patterns are selected.

Association Rule Mining or ARM is a rule-based machine learning method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measure of interestingness. It is meant to find frequent patterns, correlations, associations in data sets that are present in the database.

The SPM and ARM algorithms together can then be used to identify existing patterns in the operations carried out in any given transaction. This can then be used to define data dependencies that can be used to check if a given transaction is malicious. Also, since each attribute in any given database may have a varying degree of importance or sensitivity, we can assign each of these attributes different weights. For example, in a banking system, the account number and debit card number are more important and sensitive in comparison to the names and other personal details of the account holder. Thus it is more important to track the malicious transactions according to their sensitivities.

I am currently working to use the improved versions of studied algorithms and integrate different concepts mentioned to develop robust IDS that can be deployed in organizations to safeguard its database.