Chapter 2
Enhance Network Intrusion Detection System by Exploiting BR Algorithm as an Optimal Feature Selection

Soukaena Hassan Hashem
University of Technology, Iraq

ABSTRACT
This chapter aims to build a proposed Wire/Wireless Network Intrusion Detection System (WWNIDS) to detect intrusions and consider many of modern attacks which are not taken in account previously. The proposal WWNIDS treat intrusion detection with just intrinsic features but not all of them. The dataset of WWNIDS will consist of two parts; first part will be wire network dataset which has been constructed from KDD’99 that has 41 features with some modifications to produce the proposed dataset that called modern KDD and to be reliable in detecting intrusion by suggesting three additional features. The second part will be building wireless network dataset by collecting thousands of sessions (normal and intrusion); this proposed dataset is called Constructed Wireless Data Set (CWDS). The preprocessing process will be done on the two datasets (KDD & CWDS) to eliminate some problems that affect the detection of intrusion such as noise, missing values and duplication.

INTRODUCTION
Intrusion is any set of deliberate, unauthorized, inappropriate, and/or illegal activities by perpetrators either inside or outside an organization, which can be deemed a system penetration, that attempt to compromise the integrity, confidentiality or availability of a system resource (Hashem, 2013). An Intrusion Detection is a security mechanism that monitors and analyzes network or computer system events to provide real-time warnings for unauthorized access to system resources or to archive log and traffic information for later analysis. The detection of intrusion (intrusion attempts) operates on logs or other information available from the computer system or the network. ID is an important component of infrastructure of protection mechanisms (Majeed et al, 2013).
Intrusion Detection System is software, hardware or a combination of both that monitors and collects system and network information and analyzes it to determine if an intrusion has occurred. Snort is an open source IDS available to the general public. IDS may have different capabilities depending upon how complex and sophisticated the components are. Inevitably, the best intrusion prevention system will fail. Thus a system’s second line of defense is IDS, and this had Bee Ranker (BR) n which is the focus of much research in recent years (Majeed et al, 2013).

Data Mining-based ID techniques generally fall into two main categories: ‘misuse detection’ and ‘anomaly detection’. In misuse detection systems, patterns of well-known attacks are used to match and identify known intrusion. These techniques are able to automatically retrain ID models on different input data that include new types of attacks, as long as they have BR n labeled appropriately. Unlike signature-based IDSs, models of misuse are created automatically, and can be more sophisticated and precise than manually created signatures. A base stone of misuse detection techniques strength is their high degree of Precision in detecting known attacks and their variations. Misuse detection techniques in general are not effective contra new attacks that have no matched rules or models yet. Anomaly detection, on the other hand, builds models of normal behavior, and flags observed activities that deviate significantly from the established normal usage profiles as anomalies, that is, possible intrusions. Anomaly detection techniques thus identify new types of intrusions as diversions from usual usage. Anomaly detection techniques can be effective contra unknown or new attacks since no a priori knowledge about fixed intrusions are required. However, anomaly-based IDSs tend to generate more false alarms than misuse-based IDSs because an anomaly can just be a new normal behavior. Some IDSs use both anomaly and misuse detection techniques (Zhou & Zhao, 2013).

DARPA’99 “KDD’99 dataset” which represents the most widely used dataset for the evaluation of ID methods since 1999. This dataset is prepared by Stolfo et al. and is built based on the data captured in DARPA’98 IDS evaluation program. The most used is 10% KDD’99 dataset, available in Notepad format at (Bensefia and Ghoulami, 2011) and it consists of thousands connection session. Each session has 41 features and is labeled as either normal or an intrusion, with exactly one specific attack type. There are four attacks categories these are: Denial of Service Attack (DoS), User to Root Attack (U2R), Remote to Local Attack (R2L), and Probing Attack. KDD’99 features can be classified into three groups: Basic features, Content features, and Traffic features (Lee et al, 1999; The UCI, 1999).

Feature selection, also known as “subset selection” or “variable selection”, is an important pre-processing step used in data mining and machine learning because it treat huge no. of data with many attributes, where a subset of the features available from the original data are selected for posterior application of a learning algorithm. Feature selection is the most critical step in constructing intrusion detection models since it tend to reduce, if possible, number of features (attributes) and select the most intrinsic of these features in the classification decision, and hence to reduce the computation time of implementing the classification algorithms. Feature selection was proven to have a significant impact on the performance of the classifiers, since it can reduce the building and testing time of a classifier by reasonable percent that according many of previous experiments and researches (Majeed et al, 2013).

The Bee Ranker (BR) Algorithm is a new population-based search algorithm that mimics the food foraging behavior of swarms of honey BR s. In its basic version, the algorithm performs a kind of ‘neighborhood search’ combined with ‘random search’ and can be used for optimization problems (Hashem et al, 2013).