Chapter II
Statistical Analysis Software

INTRODUCTION

Statistical software and their corresponding computing environments are essential factors that will lead to the achievement of efficient and better research. If we think of computing and classifying algorithms as the roadmap to arrive at our final destination, a statistical package is the vehicle that is used to reach this point. Figure 2.1 shows a basic roadmap of the roles that statistical software packages play in network security.

One of the advantages of using a statistical package in network security is that it provides a fairly easy and quick way to explore data, test algorithms and evaluate models. Unfortunately, not every package is suitable for analyzing network traffic. Given the natural characteristics of the network traffic data (i.e., large size and the ability to change dynamically), several fundamental attributes are necessary for specific packages. First, the package should have good data management capacities, which include the capacity to read large data and output/save resulting files in different formats, the capability to merge and link processed data with other data sources, and the ability to create, modify and delete variables within data. Second, it should be able to process large amounts of data efficiently because statistical analyses in network security are usually based on dynamic online data, which requires the application to conduct analyses timely; this differs from areas such as healthcare, life science, and epidemiology where statistical analyses are conducted based on static data.
on static offline data. Third, it should support modern modeling procedures and methods, such as the Bayesian methods, hidden Markov model, hierarchical generalized linear model, etc. Finally, because usability is an important factor, we want the software to be both accessible and user-friendly. These attributes are particularly important during the development phase because they allow us to quickly test hypotheses and examine modeling strategies effectively.

Since many commercial and research-oriented software packages may not have all of the aforementioned attributes, we may need to implement multiple packages, such as packages for data management, for fitting a particular model, and for displaying results graphically. In the end, we may more likely use a general-purpose programming language, such as C, C++ or Java to create a customized application which we can later integrate with the other components of the intrusion detection or prevention system. The results obtained from the statistical software can be used as a gold-standard benchmark to validate the results from the customized application.

In this chapter, we will introduce several popular commercial and research-oriented packages that have been widely used in the statistical analysis, data mining, bioinformatics, and computer science communities. Specifically, we will discuss SAS\(^1\), Stata\(^2\) and R in Sections The SAS System, STATA and R, respectively; and briefly describe S-Plus\(^3\), WinBUGS, and MATLAB\(^4\) in Section Other Packages. The goal of this chapter is to provide a quick overview of these analytic software packages with some simple examples to
Use of Social Media for Women Entrepreneurship (With Special Reference to Bhopal)
www.igi-global.com/article/use-of-social-media-for-women-entrepreneurship-with-special-reference-to-bhopal/247141?camid=4v1a