

# Chapter 7

## Forecasting Software Vulnerabilities Using Time-Series Techniques

**Baidyanath Biswas**  
*IIM Lucknow, India*

**Dileep Kumar G.**  
*Adama Science and Technology University, Ethiopia*

### ABSTRACT

*This chapter discusses the concepts of time-series applications and forecasting in the context of information systems security. The primary objective in such formulation is the training of the models followed by efficient prediction. Although economic and financial forecasting problems extensively use time-series, predicting software vulnerabilities is a novel idea. The chapter also provides appropriate guidelines for the implementation and adaptation of univariate time-series for information security. To achieve this, the authors focus on the following techniques: autoregressive (AR), moving average (MA), autoregressive integrated moving average (ARIMA), and exponential smoothing. The analysis considers a unique data set consisting of the publicly exposed software vulnerabilities, available from the U.S. Dept. of Homeland Security. The problem is presented first, followed by a general framework to identify the problem, estimate the best-fit parameters of that model, and conclude with an illustrative example from the above dataset to familiarize readers with the business problem.*

DOI: 10.4018/978-1-5225-3534-8.ch007

## **1. LEARNING OBJECTIVES**

- This chapter will guide the readers to learn four advanced forecasting techniques – Autoregressive (AR), Moving Average (MA), Autoregressive Integrated Moving Average (ARIMA), and Exponential Smoothing.
- The academic context of this study is set in the background of cybersecurity and software vulnerability disclosures. It makes the chapter a unique topic for researchers.
- The chapter will aid students to solve forecasting problems dealing with various Information Security issues – virus attacks, malware, and data breaches.
- Throughout the chapter, we extensively use R to model the forecasting problems for our analysis. R is a freeware, open-source language and commonly used to analyze statistical research problems.
- In a real business scenario, the organization needs to minimize the IT risk of future software breaches and exploits. A proactive way of doing this involves the selection and use of software with common functionality but lowest vulnerability count among a set of available players in the market. Time-series forecasting methods proposed in this chapter will help academicians as well as top management in organizations to achieve this goal.

## **2. INTRODUCTION AND MOTIVATION**

Security vulnerabilities are weaknesses that are resident in each software through which attackers can intrude into the IT systems and cause severe damage to the users. Often hackers exploit publicly disclosed vulnerabilities by targeting information systems, which lead to severe financial losses in those organizations (Telang & Wattal, 2007). The Common Vulnerabilities and Exposures (CVE) Project jointly with the US National Vulnerability Database (NVD) acts as a ready repository for publicly reported software vulnerabilities. While there were only 900 publicly reported vulnerabilities till 1999, a drastic increase in their disclosure each year has led to more than 82,000 CVEs reported in the NVD till date<sup>1</sup>. The year 2017 is witnessing a massive rise in the number of vulnerability exploits with maximum exposure count of more than 5400 till May, compared to 6400 for the entire year of 2015 and 2016. Figure 1 reports the approximate annual number of security vulnerabilities till date.

39 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/forecasting-software-vulnerabilities-using-time-series-techniques/207383](http://www.igi-global.com/chapter/forecasting-software-vulnerabilities-using-time-series-techniques/207383)

## Related Content

---

### A Review of Data Governance Definitions and Emerging Perspectives

Uma G. Gupta and San Cannon (2020). *International Journal of Data Analytics* (pp. 30-47).

[www.irma-international.org/article/a-review-of-data-governance-definitions-and-emerging-perspectives/258919](http://www.irma-international.org/article/a-review-of-data-governance-definitions-and-emerging-perspectives/258919)

### Logistics for City and Regional Planning: Urban and Regional Planning Without Taking Into Account the Effects of Transport Logistics

Hermann Knoflacher (2018). *Intelligent Transportation and Planning: Breakthroughs in Research and Practice* (pp. 768-790).

[www.irma-international.org/chapter/logistics-for-city-and-regional-planning/197161](http://www.irma-international.org/chapter/logistics-for-city-and-regional-planning/197161)

### Measuring the Quality of Healthcare Services in Bangladesh

Fahima Khanam and Nayem Rahman (2019). *International Journal of Big Data and Analytics in Healthcare* (pp. 15-31).

[www.irma-international.org/article/measuring-the-quality-of-healthcare-services-in-bangladesh/232323](http://www.irma-international.org/article/measuring-the-quality-of-healthcare-services-in-bangladesh/232323)

### Statistics

(2017). *Comparative Approaches to Using R and Python for Statistical Data Analysis* (pp. 1-31).

[www.irma-international.org/chapter/statistics/175143](http://www.irma-international.org/chapter/statistics/175143)

### Unstructured Healthcare Data Archiving and Retrieval Using Hadoop and Drill

Hang Yue (2018). *International Journal of Big Data and Analytics in Healthcare* (pp. 28-44).

[www.irma-international.org/article/unstructured-healthcare-data-archiving-and-retrieval-using-hadoop-and-drill/223165](http://www.irma-international.org/article/unstructured-healthcare-data-archiving-and-retrieval-using-hadoop-and-drill/223165)