MODELING AND CUSTOM DATASET GENERATION OF THE DNP3 SCADA PROTOCOL

BY

ARTURO CUEVAS

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois Urbana-Champaign, 2023

Urbana, Illinois

Adviser:

Professor David M. Nicol

ABSTRACT

DNOP is

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my thesis advisor Professor David M. Nicol for his invaluable guidance and support in my graduate studies. Professor Nicol displayed immense patience, consideration, and kindness to me during the completion of my degree that I hope I can repay to the future generation of researchers.

As well I would to thank the support of my fellow research group members at the Information Trust Institute: David Emmerich, Matthew Needham, and Logan Marlow, for their support and expertise in the completion of my thesis. I also would like to thank the endless support from my family and friends who made this possible.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

The critical infrastructure of modern society is heavily dependent on Supervisory Control and Data Acquisition (SCADA) systems. SCADA protocols, such as Distributed Network Protocol 3 (DNP3), play a significant role in a variety of sectors including utilities, water treatment, and oil and gas. Due to the increased reliance on SCADA systems and the potential implications of disruptions, it is crucial to ensure the robustness of these systems against possible cyber threats.

The study of network traffic, specifically the DNP3 protocol, is an essential part of understanding, predicting, and mitigating potential attacks on these systems. Synthetic traffic generation has been employed extensively as a tool for this purpose. However, creating realistic and representative DNP3 traffic for such analyses poses a significant challenge.

This thesis aims to explore and compare two distinct methods for creating synthetic yet realistic DNP3 traffic: cycle detection and Discrete Time Markov Chain (DTMC). These methods were implemented using a custom-built library in Scapy, a powerful Python-based interactive packet manipulation program and library. The research objectives of this study are two-fold. First, we aim to gain a deeper understanding of the normal traffic patterns of the DNP3 protocol. Second, we aim to evaluate the efficacy of the cycle detection and DTMC methods in generating synthetic traffic that is representative of these normal patterns.

The paper is structured as follows. Chapter 2 provides an overview of the existing literature on the DNP3 protocol, synthetic traffic generation, and the two methods under consideration. Chapter 3 details the methodology employed in customizing Scapy for DNP3 traffic analysis and implementing the two methods. Chapter 4 discusses the design and implementation of the traffic generation process. Chapter 5 presents the results of our

experiments, including a comparison of the two methods based on several metrics. The findings are then discussed and interpreted in Chapter 6, and Chapter 7 provides a conclusion and suggestions for future research.

Chapter 2

BACKGROUND

2.1 DNP3 Protocol
2.2 Previous Methods for Traffic Generation
2.3 Cycle Detection
2.4 DTMC
2.5 Role of Scapy in Traffic Analysis

Chapter 3

MOTIVATION

3.1 Importance of DNP3 Traffic Analysis
3.2 Current Challenges in DNP3 Traffic Analysis
3.3 Proposed Methodology for Analysis

Chapter 4

CUSTOM DNP3 LIBRARY

4.1 Building the Custom Scapy Library
4.2 Utility of Scapy in DNP3 Traffic Analysis
4.3 Benefits and Limitations of the Custom Library

Chapter 5

SYNTHETIC TRAFFIC GENERATION

5.1 Cycle Detect Method for Traffic Generation
5.2 DTMC Method for Traffic Generation
5.3 Comparative Analysis of the Two Methods

Chapter 6

CASE STUDY:

6.1 Evaluation Criteria
6.2 Performance Analysis
6.1 Experiment Setup
6.2 DNP3 Traffic Analysis using Cycle Detection
6.3 DNP3 Traffic Analysis using DTMC
6.4 Comparative Analysis of Both Methods
6.3 Discussion on Results

Chapter 7

CASE STUDY/DISCUSSION

Chapter 8

CONCLUSION

8.1 Summary of Findings
8.2 Implications and Applications
8.3 Recommendations for Future Research

RERFERENCES