NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

DISSERTATION

EXPLORATORY DATA ANALYSIS OF DEFENSIVE CYBER DECEPTION EXPERIMENTATION

by

Michael Senft

June 2023

Dissertation Supervisor: James B. Michael

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC, 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE June 2023	3. REPORT TYPE AND DATES COVERED Dissertation

4. TITLE AND SUBTITLE EXPLORATORY DATA ANALYSIS OF DEFENSIVE CYBER DECEPTION EXPERIMENTATION	5. FUNDING NUMBERS
6. AUTHOR(S) Michael Senft	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. This report has supplemental(s).

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.	12b. DISTRIBUTION CODE A

13. ABSTRACT (maximum 200 words)

The use of cyber deception is a powerful but underutilized practice for spoiling cyber-attacker activity from initial reconnaissance to actions on an objective. This dissertation applies exploratory data analysis techniques to network traffic data collected during the Tularosa study, the largest experiment of defensive cyber deception involving human subjects to date. Using a network-graph–based information technology artifact developed for this research temporal variables of latency, frequency, and duration for system interactions were extracted from over 200 gigabytes of network traffic collected during the Tularosa study.

Analysis of the data bolsters previous assertions that decoy-based deception is significantly more effective against cyber-attackers when presence of deception is known and suggests this impact is enduring. Distinct temporal patterns for aggregated network traffic for each experimental group were also identified. This research also provides the first application of factor analysis to data collected during the Tularosa study, which highlighted the key role of technical experience in explaining observed variance. Finally, quantum probability theory is explored as a potential model to explain variations in latency observed between the control and experimental groups which may be the result of interference effect and/or order effect phenomena. Cyber-defenders can apply findings from this research to impede cyber-attacker activity.

14. SUBJECT TERMS cyber deception, deception, Tularosa study, decoy, network graph, network traffic, cyber, cybersecurity, quantum, quantum probability theory, exploratory data analysis, factor analysis	15. NUMBER OF PAGES 309
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

EXPLORATORY DATA ANALYSIS OF DEFENSIVE CYBER DECEPTION EXPERIMENTATION

Michael Senft
Lieutenant Colonel, United States Army
BE, Virginia Polytechnic Institute and State University, 2003
MEng, Washington State University, 2007
MS, Computer Science, Naval Postgraduate School, 2015
MS, Network Operations and Technology, Naval Postgraduate School, 2015

Submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY IN INFORMATION SCIENCES

from the

NAVAL POSTGRADUATE SCHOOL
June 2023

Approved by:

James B. Michael
Department of
Computer Science
Dissertation Supervisor

Raymond R. Buettner
Department of
Information Sciences
Dissertation Chair

Anthony M. Canan
Department of
Information Sciences

Mollie R. McGuire
Department of
Information Sciences

Ross J. Schuchard
Department of
Operations Research

Kimberly Ferguson-Walter
Laboratory for Advanced
Cybersecurity Research

Approved by:

Alex Bordetsky
Chair, Department of Information Sciences

Joseph P. Hooper
Vice Provost of Academic Affairs

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The use of cyber deception is a powerful but underutilized practice for spoiling cyber-attacker activity from initial reconnaissance to actions on an objective. This dissertation applies exploratory data analysis techniques to network traffic data collected during the Tularosa study, the largest experiment of defensive cyber deception involving human subjects to date. Using a network-graph–based information technology artifact developed for this research temporal variables of latency, frequency, and duration for system interactions were extracted from over 200 gigabytes of network traffic collected during the Tularosa study.

Analysis of the data bolsters previous assertions that decoy-based deception is significantly more effective against cyber-attackers when presence of deception is known and suggests this impact is enduring. Distinct temporal patterns for aggregated network traffic for each experimental group were also identified. This research also provides the first application of factor analysis to data collected during the Tularosa study, which highlighted the key role of technical experience in explaining observed variance. Finally, quantum probability theory is explored as a potential model to explain variations in latency observed between the control and experimental groups which may be the result of interference effect and/or order effect phenomena. Cyber-defenders can apply findings from this research to impede cyber-attacker activity.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I. INTRODUCTION 1
- A. CYBER THREAT LANDSCAPE 1
- B. CYBER ATTACK FRAMEWORKS 3
- C. RESEARCH PROBLEM 4
  - 1. Research Purpose 4
  - 2. Research Questions 5
  - 3. Research Approach 5
- D. CONTRIBUTIONS 6
- E. CONNECTION WITH FIELD OF INFORMATION SCIENCES 6
- F. DISSERTATION ORGANIZATION 6

II. LITERATURE REVIEW 9
- A. THEORY FRAMEWORK 9
  - 1. General Deception Theory 10
  - 2. Cue-Based and Content-Based Deception Theory 10
  - 3. Cyber Deception Theory 11
  - 4. Information Theory 12
  - 5. Decision-Making Under Uncertainty 13
- B. PRIOR CYBER DECEPTION EXPERIMENTATION INVOLVING HUMAN SUBJECTS 16
- C. OPERATIONAL DEFINITIONS 17

III. RESEARCH DESIGN 21
- A. INTRODUCTION 21
- B. RESEARCH DATA SOURCE 21
  - 1. Tularosa Study Overview 21
  - 2. Tularosa Study Data Collection 24
  - 3. Tularosa Study Experimental Design 25
- C. RESEARCH METHODOLOGY 28
  - 1. Subject Network Activity Timeline Reconstruction and Temporal Variable Extraction 28
  - 2. Exploratory Data Analysis of Subject Network Activity 29
  - 3. Factor Analysis of Subject Activity 32
- D. JUSTIFICATION FOR SECONDARY ANALYSIS OF PREVIOUSLY COLLECTED EXPERIMENTAL DATA 35

E. VALIDITY 35
1. Threats to Internal Validity 35
2. Threats to External Validity 36

IV. DATA EXTRACTION AND TRANSFORMATION 37
A. BACKGROUND 37
B. DATA EXTRACTION AND TRANSFORMATION PROCESS 39
1. PCAP Transformation Overview 39
2. PCAP Transformation Validation Check 43
C. SUBJECT ACTIVITY EXPORT 44
1. Temporal Variables 44
2. Subject Activity Per Minute 45
3. Enrichment of Extracted Subject Activity 46
D. TECHNICAL ARTIFACTS 47

V. EXPLORATORY DATA ANALYSIS 49
A. ANALYTIC FRAMEWORK 49
B. EXPERIMENT SUBJECT SELECTION 51
C. DAY 1 EXPERIMENT DATA ANALYSIS 53
1. Real System Interactions For All Groups (AI, AU, PI, PU) 53
2. Decoy System Interactions for Cyber Deception Groups (PI, PU) 86
3. Real and Decoy System Interactions for Cyber Deception Groups (PI, PU) 92
D. DAY 2 EXPERIMENT DATA ANALYSIS 99
1. Real System Interactions Across All Groups (AIAU, AUPU, PIAU, PUAU) 99
2. Decoy System Interactions for Cyber Deception Group (AUPU) 115
E. DAY 1 AND DAY 2 EXPERIMENT DATA COMPARISON ANALYSIS 117
1. Real System Interactions for All Groups (AI, AU, PI, PU and AIAU, AUPU, PIAU, PUAU) 117
2. Decoy System Interactions for Cyber Deception Groups (PI, PU and AUPU) 118

VI. FACTOR ANALYSIS 121
A. PRINCIPAL COMPONENT ANALYSIS 121
1. Day 1 Analysis 122

2. Day 2 Analysis 132
B. EXPLORATORY FACTORY ANALYSIS 141
1. Day 1 Analysis 142
2. Day 2 Analysis 148

VII. THEORETICAL MODELS OF DEFENSIVE CYBER DECEPTION 155
A. VECTOR-BASED CONCEPTUAL MODELS 155
1. OODA Loop-based Model 155
2. Typology of Misperception Based-Model 159
B. DEMPSTER-SHAFER THEORY APPLICATION 164
C. QUANTUM PROBABILITY THEORY-BASED MODEL 166
1. Introduction to Quantum Probability Theory 166
2. Interference Effect 167
3. Order Effect 169
4. QPT State Vector Representation 173

VIII. CONCLUSION 179
A. KEY FINDINGS 179
B. REASSESSMENT OF TULAROSA STUDY HYPOTHESES 180
C. MEASURES OF SUCCESS 182
1. Day 1 Analysis 183
2. Day 2 Analysis 187
D. CONTRIBUTIONS 189
E. FUTURE WORK 190
1. Quantum Random Walk Theory-Based Cyber-Attacker Decision-Making Model 190
2. Decoy Effectiveness Improvement using Threat and Capability Coevolutionary Wargame Model 190
3. Further Factor Analysis 191
4. Additional Tularosa Study Data Integration 192

APPENDIX A. PROGRAMMING CODE 193
A. PCAP TRANSFORMATION 193
1. Split Oversize PCAP (250+ MB) into 200 MB files 193
2. PCAP to CSV Conversion 193
B. PYTHON CODE USED TO IMPORT CSV PCAP DATA INTO NEO4J 193

APPENDIX B. FACTOR ANALYSIS 225

A. DAY 1 ANALYSIS 225
1. Absent-Informed 225
2. Absent-Uninformed 229
3. Present-Informed 233
4. Present-Uninformed 238
5. Combined Day 1 242
B. DAY 2 ANALYSIS 246
1. Absent-Informed / Absent-Uninformed 246
2. Absent-Uninformed/Present-Informed 251
3. Present-Informed/Absent-Informed 255
4. Present-Uninformed/Absent-Uninformed 259
5. Combined Day 2 264

SUPPLEMENTAL 269

LIST OF REFERENCES 271

INITIAL DISTRIBUTION LIST 281

LIST OF FIGURES

Figure 1. Tularosa Study Experimental Conditions 22

Figure 2. Tularosa Task Instructions for "Not Informed" Condition. Source: Ferguson-Walter et al. (2021). 23

Figure 3. Source: Tularosa Task Instructions for "Informed" Condition. Source: Ferguson-Walter et al. (2021). 23

Figure 4. Tularosa Study Experiment Cyber Range 26

Figure 5. Deception Absent Cyber Range Configuration 27

Figure 6. Deception Present Cyber Range Configuration 27

Figure 7. Subject Network Activity Timeline Reconstruction Overview 28

Figure 8. Visual Depiction of Latency Measurement 31

Figure 9. Visual Depiction of Duration Measurement 31

Figure 10. Visual Depiction of Frequency Measurement 32

Figure 11. PCA and EFA Variance Comparison. Adapted from Schreiber (2021). 33

Figure 12. Reflective and Formative Perspectives. Source: Coltman et al. (2008). 34

Figure 13. Wireshark Screenshot of PCAP data 38

Figure 14. PCAP Data Extracted to CSV Format 40

Figure 15. Data Model Schema for Tularosa Network Traffic Data 41

Figure 16. Time Relationships and Properties as Represented in Neo4j 42

Figure 17. Extracted Subject Activity by System 45

Figure 18. Subject Network Activity Per Minute 46

Figure 19. Real System Subject Activity Enrichment 47

Figure 20. Decoy System Subject Activity Enrichment 47

Figure 21. Day 1 Real System Latency Histogram 55

Figure 22. Day 1 Real Systems Latency Box-and-Whisker Plot 56

Figure 23. Day 1 Real System Frequency Histogram. 58

Figure 24. Day 1 Real System Frequency Box-and-Whisker Plot 59

Figure 25. Day 1 Real System Duration Histogram. 61

Figure 26. Day 1 Real System Duration Box-and-Whisker Plot. 62

Figure 27. Day 1 Real System Average Packets Per Minute. 64

Figure 28. Day 1 Real System Median Packets Per Minute. 65

Figure 29. Day 1 Real System Standard Deviation Per Minute 65

Figure 30. Day 1 Real System Variance Per Minute. 66

Figure 31. Day 1 Real System Subject Activity Threshold – 2 Packets Per Minute 67

Figure 32. Day 1 Real System Subject Activity Threshold – 7 Packets Per Minute 67

Figure 33. Day 1 Real System Subject Activity Threshold – 20 Packets Per Minute 68

Figure 34. Day 1 Real System Subject Activity Threshold – 54 Packets Per Minute 68

Figure 35. Day 1 Real System Subject Activity Threshold – 148 Packets Per Minute 69

Figure 36. Day 1 Real System Subject Activity Threshold – 403 Packets Per Minute. 69

Figure 37. Day 1 Real System Interaction Timeline by Group 70

Figure 38. Day 1 Real System Latency 90% Winsorization Box-and-Whisker 76

Figure 39. Day 1 Real System Frequency 90% Winsorization Box-and-Whisker Plot. 78

Figure 40. Day 1 Real System Duration 90% Winsorization Box-and-Whisker Plot 80

Figure 41. Day 1 Real System Latency 80% Winsorization Box-and-Whisker Plot 82

Figure 42. Day 1 Real System Frequency 80% Winsorization Box-and-Whisker Plot 84

Figure 43. Day 1 Real System Duration 80% Winsorization Box-and-Whisker Plot 86

Figure 44. Elapsed Day 1 Decoy System Interaction Percentages By Groups 91

Figure 45. Elapsed Day 1 Real and Decoy System Interaction Percentages 97

Figure 46. Elapsed Day 1 Split Real and Decoy System Interaction Percentages..... 98

Figure 47. Day 2 Real System Latency Histogram 100

Figure 48. Day 2 Real System Latency Box-and-Whisker Plot..... 101

Figure 49. Day 2 Real System Frequency Histogram 103

Figure 50. Day 2 Real System Frequency Box-and-Whisker Plot..... 104

Figure 51. Day 2 Real System Duration Histogram..... 106

Figure 52. Day 2 Real System Duration Box-and-Whisker Plot 107

Figure 53. Day 2 Real System Average Packets Per Minute 109

Figure 54. Day 2 Real System Median Packets Per Minute 110

Figure 55. Day 2 Real System Standard Deviation Per Minute..... 110

Figure 56. Day 2 Real System Variance Per Minute 111

Figure 57. Day 2 Real System Activity Threshold – 2 Packets Per Minute 112

Figure 58. Day 2 Real System Activity Threshold – 7 Packets Per Minute 112

Figure 59. Day 2 Real System Activity Threshold – 20 Packets Per Minute 113

Figure 60. Day 2 Real System Activity Threshold – 54 Packets Per Minute 113

Figure 61. Day 2 Real System Activity Threshold – 148 Packets Per Minute 114

Figure 62. Day 2 Real System Activity Threshold – 403 Packets Per Minute 114

Figure 63. Day 2 Real System Elapsed System Interaction Percentages..... 115

Figure 64. Day 1 and 2 Elapsed Real Systems Interaction Percentages 118

Figure 65. Day 1 and 2 Elapsed Decoy Systems Interaction Percentages 119

Figure 66. Day 1 Absent-Informed PCA Summary Plots 123

Figure 67. Day 1 Absent-Informed PCA Scree Plot 123

Figure 68. Day 1 Absent-Uninformed PCA Summary Plots 125

Figure 69. Day 1 Absent-Uninformed PCA Scree Plot..... 125

Figure 70. Day 1 Present-Informed PCA Summary Plots..... 127

Figure 71. Day 1 Present-Informed PCA Scree Plot..... 127

Figure 72. Day 1 Present-Uninformed PCA Summary Plots..... 129

Figure 73. Day 1 Present-Uninformed PCA Scree Plot 129

Figure 74. Day 1 Combined PCA Summary Plot 131

Figure 75. Day 1 Combined PCA Scree Plot 131

Figure 76. Day 2 Absent-Informed/Absent-Uninformed PCA Summary Plots 133

Figure 77. Day 2 Absent-Informed/Absent-Uninformed PCA Scree Plot..... 133

Figure 78. Day 2 Absent-Uninformed/Present-Uninformed PCA Summary Plots..... 135

Figure 79. Day 2 Absent-Uninformed/Present-Uninformed PCA Scree Plot..... 135

Figure 80. Day 2 Present-Informed/Absent-Uninformed PCA Summary Plots 137

Figure 81. Day 2 Absent-Informed/Absent-Uninformed PCA Scree Plot..... 137

Figure 82. Day 2 Present-Uninformed/Absent-Uninformed PCA Summary Plots..... 139

Figure 83. Day 2 Present-Uninformed/Absent-Uninformed PCA Scree Plot..... 139

Figure 84. Day 2 Combined PCA Summary Plots 140

Figure 85. Day 2 Combined PCA Scree Plot..... 141

Figure 86. Absent-Uninformed OODA Loop Model. Adapted from Boyd (1995). 156

Figure 87. Absent-Informed OODA Loop Model. Adapted from Boyd (1995). 157

Figure 88. Present-Uninformed OODA Loop Model. Adapted from Boyd (1995). 158

Figure 89. Present-Informed OODA Loop Model. Adapted from Boyd (1995). 159

Figure 90. A Typology of Misperception. Source: Whaley (1982). 161

Figure 91. Absent-Uninformed Perception Components 162

Figure 92. Absent-Informed Perception Components 163

Figure 93. Present-Uninformed Perception Components 163

Figure 94. Present-Informed Perception Components 164

Figure 95. Day 1 Real System Average Packets Per Minute Activity Peaks and Troughs 165

Figure 96. Illustration of the categorization-decision interference effect. Source: Busemeyer & Wang (2015). 168

Figure 97. Presumed Information Order for AI and PI Groups 170

Figure 98. Presumed Information Order for AU and PU Groups 171

Figure 99. Belief State Vector for Medical Decision-Making Task. Source: Trueblood and Busemeyer (2011, p. 1523). 172

Figure 100. Belief State Vector for Tularosa Study Decision-Making Task. Adapted from Trueblood & Busemeyer (2011). 173

Figure 101. QPT Representation of Sending Packet Without Uncertainty 174

Figure 102. Absent-Uninformed QPT Representation 175

Figure 103. Absent-Informed QPT Representation 176

Figure 104. Present-Uninformed QPT Representation 177

Figure 105. Present-Informed QPT Representation 178

Figure 106. Reconnaissance Progress Results Summary. Adapted from Ferguson-Walter et al. (2021). 182

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

Table 1. Day 1 Real System Latency Descriptive Statistics 54

Table 2. Day 1 Real Systems Latency Shapiro-Wilk-Royston Test for Normality 56

Table 3. Day 1 Real Systems Latency Levene's Test for Equal Variance 57

Table 4. Day 1 Real Systems Latency Kruskal-Wallis Test 57

Table 5. Day 1 Real Systems Latency Dunn's Post Hoc Test with Benjamini-Hochberg Correction 57

Table 6. Day 1 Real System Frequency Descriptive Statistics 58

Table 7. Day 1 Real System Frequency Shapiro-Wilk-Royston Test for Normality 59

Table 8. Day 1 Real System Frequency Levene's Test for Equal Variance 60

Table 9. Day 1 Real System Frequency Kruskal-Wallis Test 60

Table 10. Day 1 Real System Frequency Dunn's Post Hoc Test with Benjamini-Hochberg Correction 60

Table 11. Day 1 Real System Duration Descriptive Statistics 61

Table 12. Day 1 Real System Duration Shapiro-Wilk-Royston Test for Normality 62

Table 13. Day 1 Real System Duration Levene's Test for Equal Variance 63

Table 14. Day 1 Real System Duration Kruskal-Wallis Test 63

Table 15. Day 1 Real System Duration Dunn's Post Hoc Test with Benjamini-Hochberg Correction 63

Table 16. Day 1 Real System Log Transformed Latency Shapiro-Wilk-Royston Test for Normality 71

Table 17. Day 1 Real System Log Transformed Frequency Shapiro-Wilk-Royston Test for Normality 71

Table 18. Day 1 Real System Log Transformed Duration Shapiro-Wilk-Royston Test for Normality 72

Table 19. Day 1 Real System Cube Root Transformed Latency Shapiro-Wilk-Royston Test for Normality 72

Table 20. Day 1 Real System Cube Root Transformed Frequency Shapiro-Wilk-Royston Test for Normality 73

Table 21. Day 1 Real System Cube Root Transformed Duration Shapiro-Wilk-Royston Test for Normality 73

Table 22. Day 1 Real System Box-Cox Transformed Latency Shapiro-Wilk-Royston Test for Normality 74

Table 23. Day 1 Real System Box-Cox Transformed Frequency Shapiro-Wilk-Royston Test for Normality 74

Table 24. Day 1 Real System Box-Cox Transformed Duration Shapiro-Wilk-Royston Test for Normality 74

Table 25. Day 1 Real System Latency 90% Winsorization Descriptive Statistics 75

Table 26. Day 1 Real System Frequency 90% Winsorization Descriptive Statistics 77

Table 27. Day 1 Real System Duration 90% Winsorization Descriptive Statistics 79

Table 28. Day 1 Real System Latency 80% Winsorization Descriptive Statistics 81

Table 29. Day 1 Real System Frequency 80% Winsorization Descriptive Statistics 83

Table 30. Day 1 Real System Duration 80% Winsorization Descriptive Statistics 85

Table 31. Day 1 Decoy System Latency Descriptive Statistics 87

Table 32. Day 1 Decoy System Latency Shapiro-Wilk-Royston Test for Normality 87

Table 33. Day 1 Decoy System Latency Levene's Test for Equal Variance 88

Table 34. Day 1 Decoy System Latency Kruskal-Wallis Test 88

Table 35. Day 1 Decoy System Frequency Descriptive Statistics 88

Table 36. Day 1 Decoy System Frequency Shapiro-Wilk-Royston Test for Normality 89

Table 37. Day 1 Decoy System Frequency Levene's Test for Equal Variance 89

Table 38. Day 1 Decoy System Frequency Kruskal-Wallis Test 89

Table 39. Day 1 Decoy System Duration Descriptive Statistics 90

Table 40. Day 1 Decoy System Duration Shapiro-Wilk-Royston Test for Normality 90

Table 41. Day 1 Decoy System Duration Levene's Test for Equal Variance 90

Table 42. Day 1 Decoy System Duration Kruskal-Wallis Test 90

Table 43. Day 1 Real and Decoy System Latency Descriptive Statistics 92

Table 44. Day 1 Real and Decoy System Latency Shapiro-Wilk-Royston Test for Normality 93

Table 45. Day 1 Real and Decoy System Latency Levene's Test for Equal Variance 93

Table 46. Day 1 Real and Decoy System Latency Kruskal-Wallis Test 93

Table 47. Day 1 Real and Decoy System Frequency Descriptive Statistics 94

Table 48. Day 1 Real and Decoy System Frequency Shapiro-Wilk-Royston Test for Normality 94

Table 49. Day 1 Real and Decoy System Frequency Levene's Test for Equal Variance 94

Table 50. Day 1 Real and Decoy System Frequency Kruskal-Wallis Test 94

Table 51. Day 1 Real and Decoy System Duration Descriptive Statistics 95

Table 52. Day 1 Real and Decoy System Duration Shapiro-Wilk-Royston Test for Normality 95

Table 53. Day 1 Real and Decoy System Duration Levene's Test for Equal Variance 96

Table 54. Day 1 Real and Decoy System Duration Kruskal-Wallis Test 96

Table 55. Day 2 Real System Latency Descriptive Statistics 99

Table 56. Day 2 Real System Latency Shapiro-Wilk-Royston Test for Normality 101

Table 57. Day 2 Real System Latency Levene's Test for Equal Variance 102

Table 58. Day 2 Real System Latency Kruskal-Wallis Test 102

Table 59. Day 2 Real System Latency Dunn's Post Hoc Test with Benjamini-Hochberg Correction 102

Table 60. Day 2 Real System Frequency Descriptive Statistics 103

Table 61. Day 2 Real System Frequency Shapiro-Wilk-Royston Test for Normality 104

Table 62. Day 2 Real System Frequency Levene's Test for Equal Variance 105

Table 63. Day 2 Real System Frequency Kruskal-Wallis Test 105

Table 64. Day 2 Real System Frequency Dunn's Post Hoc Test with Benjamini-Hochberg Correction 105

Table 65. Day 2 Real System Duration Descriptive Statistics 106

Table 66. Day 2 Real System Duration Shapiro-Wilk-Royston Test for Normality 107

Table 67. Day 2 Real System Duration Levene's Test for Equal Variance 108

Table 68. Day 2 Real System Duration Kruskal-Wallis Test 108

Table 69. Day 2 Real System Duration Dunn's Post Hoc Test with Benjamini-Hochberg Correction 108

Table 70. Day 2 Decoy System Latency Descriptive Statistics 116

Table 71. Day 2 Decoy System Frequency Descriptive Statistics 116

Table 72. Day 2 Decoy System Duration Descriptive Statistics 116

Table 73. Day 1 Absent-Informed PCA 123

Table 74. Day 1 Absent-Uninformed PCA 126

Table 75. Day 1 Present-Informed PCA 128

Table 76. Day 1 Present-Uninformed PCA 130

Table 77. Day 1 Combined PCA 131
Table 78. Day 2 Absent-Informed/Absent-Uninformed PCA 133
Table 79. Day 2 Absent-Uninformed/Present-Uninformed PCA 135
Table 80. Day 2 Present-Informed/Absent-Uninformed PCA 137
Table 81. Day 2 Present-Uninformed/Absent-Uninformed PCA 140
Table 82. Day 2 Combined PCA 141
Table 83. Day 1 Absent-Informed EFA 143
Table 84. Day 1 Absent-Uninformed EFA 144
Table 85. Day 1 Present-Informed EFA 145
Table 86. Day 1 Present-Uninformed EFA 146
Table 87. Day 1 Combined EFA 147
Table 88. Day 2 Absent-Informed/Absent-Uninformed EFA 148
Table 89. Day 2 Absent-Uninformed/Present-Uninformed EFA 150
Table 90. Day 2 Present-Informed/Absent-Uninformed EFA 151
Table 91. Day 2 Present-Uninformed/Absent-Uninformed EFA 152
Table 92. Day 2 Combined EFA 153
Table 93. Day 1 Summary of Statistical Analysis Results 180

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AI	Absent-Informed
AU	Absent-Uninformed
CPT	Classical Probability Theory
DST	Dempster-Shafer Theory
EFA	Exploratory Factor Analysis
IBLT	Instance-Based Learning Theory
IDS	Intrusion Detection System
IMT2	Information Manipulation Theory 2
IRB	Institutional Review Board
IT	Information Technology
JCS	Joint Chiefs of Staff
ML	Machine Learning
OS	Operating System
OT	Operational Technology
PCA	Principal Component Analysis
PI	Present-Informed
PU	Present-Uninformed
QPT	Quantum Probability Theory
TDT	Truth-Default Theory
TSQ	Task-Specific Questionnaires
UTC	Coordinated Universal Time

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I would like to thank my wife, Elizabeth, for her love, support, and patience, and my sons, Shelton and Jackson, for the joy they have given me. You are my motivation and inspiration. Second, I want to thank my parents, Mark and Connie, for their sacrifices in providing an enriching environment for me to thrive in. Finally, I want to thank my dissertation committee, Dr. Bret Michael, Dr. Ray Buettner, Dr. Kimberly Ferguson-Walter, Dr. Mustafa Canan, Dr. Mollie McGuire, and LTC Ross Schuchard, Ph.D., for their guidance and support on my epistemic journey. A special thanks to Dr. Kimberly Ferguson-Walter and her colleagues for their work on the Tularosa study and continued efforts to improve cybersecurity through deception and other cyberpsychology-informed defenses.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Cyber deception has been highlighted by many researchers as a promising tool to improve defenses against cyber attackers, yet a significant gap exists in extant literature exploring the impact of defensive cyber deception on the timing of cyber attacker behavior. The impact of deception on the timing of target behavior has been studied across a diverse range of environments, from conversational settings to sporting fields, generating new insights on useful measurements and predictive variables (Boltz et al., 2010; Brault et al., 2012). However limited research has been conducted on the impact of defensive cyber deception on the timing of cyber attacker behavior within the environment of cyberspace. In research on deception across a wide range of disciplines and application, Dr. Barton Whaley observed "every deception operation is a timed (chronological or moment to moment) process," making the incomplete understanding of the impact of defensive cyber deception on the timing of cyber attacker behavior a significant limitation in development of both cyber deception theory and practice (Whaley, 2007, p. 55). The motivation behind this research is to generate better understanding of the impact of timing through application of quantitative methods to both improve the effectiveness of defensive cyber deception and promote its use to increase the security of critical systems and networks.

A. CYBER THREAT LANDSCAPE

In 2022 the cyber threat landscape increased in scale, impact, and destructiveness driven by enhanced capabilities possessed by cyber threat actors and continued emergence of novel threats (European Union Agency for Cybersecurity [ENISA], 2022). Notable cyber incidents in 2022 include destructive cyber-attacks against power grids and critical infrastructure in Ukraine, disruption of the United Kingdom National Health Service following an attack on a Managed Service Provider, and compromise of over 60 Red Cross organizations (Zoltan, 2022). Threats from ransomware, malware, social engineering and other cyber-attacks to critical government, commercial and military networks by state and non-state actors continue to rise along with intensified incidents

impacting supply chains (ENISA, 2022). State-sponsored actors have increased exploitation of critical vulnerabilities and zero-day exploits, with a record number of zero-day exploits disclosed (ENISA, 2022). Cyber criminals also pose a progressively greater threat with ransomware attacks on critical infrastructure along with disruptive attacks against Operational Technology (OT) systems (ENISA, 2022). Hacking-for-hire and hacktivism also present an increasingly concerning threat with advanced capabilities used against a wide range of targets for financial and political motivations (ENISA, 2022). While there is no shortage of recommendations to defend against these threats, it is clear a multi-faceted approach is needed (ENSIA, 2022). Wang and Lu offer that current cyber defense practices largely focus on reactive response instead of confronting cyber-attackers during their reconnaissance activities (Wang & Lu, 2018). They emphasize if an attacker has sufficient information about systems and networks, they can likely prepare a short duration attack to achieve their objectives, further complicating detection and mitigation of an attack (Wang & Lu, 2018). Wang and Lu draw attention to research on cyber deception and Moving Target Defense (MTD) as promising capabilities to disrupt hacker reconnaissance and attack planning (Wang & Lu, 2018). Cyber-defenders possess the tremendous benefit of home-field advantage to thwart and detect cyber-attackers, and deception plays a key role in effective utilization of this advantage (Senft, 2019).

The utility of cyber deception has been highlighted by Heckman et al., Jajodia et al. and Schuh as a promising vehicle to reverse current asymmetric advantages in cyberspace, which favors attack over defense, by increasing attacker uncertainty and providing new opportunities for defenders to observe and capture attacker methods and techniques (Heckman et al., 2015; Jajodia et al., 2016; Schuh, 2016). In shifting the advantage to cyber defense, Al-Shaer et al. outline four benefits that cyber deception provides defenders: 1) deflect adversaries from their goals, 2) distort adversaries perception by introducing doubt, 3) deplete resources, and 4) discover new tactics, techniques, and procedures of adversaries (Al-Shaer et al., 2019). Incorporation of temporal reasoning into cyber deception systems has also been identified as essential future work by multiple researchers (Ashenden et al., 2021; Nunes et al., 2015; Wang &

Lu, 2019). Much work is needed to both harness the benefits that cyber deception provides defenders and to demonstrate the effectiveness of cyber deception in achieving these benefits in a consistent and cost-effective manner. Structure is needed to frame the threats posed by cyber-attackers given the complexity of cyberspace, which is provided by cyber-attack frameworks.

B. CYBER ATTACK FRAMEWORKS

Multiple frameworks for understanding cyber-attacker tactics and techniques are available including the Lockheed Martin Cyber Kill Chain® and MITRE ATT&CK® (Lockheed Martin, 2022; MITRE, 2022). For brevity, future references will include only the model's name, Cyber Kill Chain® and ATT&CK®. ATT&CK® provides a comprehensive knowledge base containing real-world techniques, tactics and procedures used by adversaries to compromise information technology systems (MITRE, 2022).

Despite different areas of focus, both frameworks share a common goal of improving understanding of cyber-attacker tactics, techniques and procedures and both begin with the same first step required for attackers to achieve their goals, reconnaissance. The Cyber Kill Chain® defines reconnaissance as "research, identification and selection of targets," while ATT&CK® defines it as "techniques that involve adversaries actively or passively gathering information that can be used to support targeting" (Hutchins et al., 2011, p. 4; MITRE, 2022, Tactics section). ATT&CK® further details reconnaissance into 10 techniques each with multiple sub-techniques. During the Tularosa study, techniques of Active Scanning (T1595), Gather Victim Host Information (T1592), and Gather Victim Network Information (T1590) were utilized by subjects in completion of the task assigned for the experimental trials. Integration of MITRE ATT&CK® with cyber deception was introduced by Stech et al. in 2016, outlining a concept to use ATT&CK™ tactics in adversary modeling to enhance effective cyber deception (Stech et al., 2016). Their work also highlights the importance of coordinated deception in conjunction with other defensive cyber operations to maximize its effectiveness (Stech et al., 2016).

C. RESEARCH PROBLEM

1. Research Purpose

The purpose of this study is to explore the impact of defensive cyber deception on the timing of cyber attacker behavior by focusing on initial attacker interaction with target systems (latency), frequency of their system interactions and duration of these system interactions. Limited information is available regarding the impact of defensive cyber deception on the timing of cyber attacker behavior, despite timing being a key construct of deception (Whaley, 2007). This is a problem for cyber deception theory because this knowledge gap limits the ability to accurately explain and predict outcomes on the timing of cyber-attacker behavior when defensive cyber deception approaches are used. This is a problem for the application of cyber deception because incomplete knowledge of the impact of defensive cyber deception on the timing on cyber-attacker behavior results in ambiguity for assessing the effectiveness of its employment. This study is important because it addresses a knowledge gap for theory and application of defensive cyber deception on its impact on the timing of cyber attacker behavior using network traffic data readily available for collection in both experimental and operational settings.

This research resulted in the development of a tailorable Information Technology (IT) artifact for the reconstruction of subject activity timelines from multiple disparate cyber/IT datasets. This research also generated contributions to cyber deception practice through investigation of useful measurements that are correlated with observed effects resulting from the employment of cyber deception. Tukey highlighted the importance of learning the right things to measure in exploring phenomena, which remains a significant gap for cyber deception research (Tukey, 1969; Wang & Lu, 2019). As observed by Whaley and Aykroyd (2007), "It is much more difficult to find suitable heuristic models for deception, precisely because deception is not made of concrete elements but is, rather, composed of intangibles that cannot be directly measured or mapped" (p. 68). Finally, this research identified challenges using temporal data from cyber/IT sources to help inform Artificial Intelligence (AI) /Machine Learning (ML) efforts to improve cyber defense.

2. Research Questions

This research was guided by the following questions:

What role do variables captured during the Tularosa study have in explaining variance of cyber-attack behavior observed under varying experimental conditions of cyber deception?

What impact did varying experimental conditions of defensive cyber deception have on the timing of cyber-attack behavior as observed through temporal variables extracted from network traffic captured during the Tularosa study?

Sub-Questions:

1. *What impact was observed on initial interaction with real and decoy systems?*
2. *What impact was observed on overall frequency of interactions with real and decoy systems?*
3. *What impact was observed on overall duration of interactions with real and decoy systems?*

3. Research Approach

Quantitative multivariate exploratory data analysis techniques were applied to cyber deception experimentation data previously collected during the Tularosa study. This analysis was performed using a post-positivist epistemic stance, which offers that an objective reality exists with measurable properties, while acknowledging biases and flaws are inherent in all methods of acquiring knowledge. Recognizing these biases and flaws, multiple analytic techniques and perspectives were used in analysis of the data.

Exploring the impact of defensive cyber deception on the timing of cyber attacker behavior requires the identification of variables associated with their behavior. For this research, a temporal variable is defined as a "property that takes on different values" which includes an element of time (Kerlinger & Lee, 2000, p. 40). Temporal variables generated by attacker behavior as observed by network defenders may also serve as useful measurements and predictors for cyber deception, which could also be compared with findings from deception experimentation in other environments.

D. CONTRIBUTIONS

The findings of this research illuminate the impact of defensive cyber deception on the timing of cyber-attacker behavior using human generated network traffic from the most comprehensive dataset available for cyber deception experimentation involving human subjects.

E. CONNECTION WITH FIELD OF INFORMATION SCIENCES

The original definition offered by Borko for information science, a subdiscipline of the broader field of information sciences, is as "a discipline that investigates the properties and behavior of information" (Borko, 1968, p. 4). He further elaborates that information science also explores "the forces governing the flow of information" (Borko, 1968, p. 3). The process of deception, regardless of medium, involves manipulation of the behavior and/or flow of information, presenting a false version of reality to change a target's action to benefit of the deceiver, which connects the field of information science with field of deception (Custy & Rowe, 2008). Deception framed as information manipulation across multiple dimensions is also a key concept for several theories on deception. Furthermore, information theory provides useful concepts for analyzing the relationship between deception, information, and communication. As presented by Custy and Rowe, a Z-channel communication model can be used to describe mathematically the information content and level of uncertainty generated by deception (Custy & Rowe, 2008). The multidisciplinary nature of this research, integrating both technical and behavioral approaches to exploring the phenomenon of cyber deception, as captured by the Tularosa study, is reflective of research typified by the field information sciences.

F. DISSERTATION ORGANIZATION

Chapter II of this dissertation provides an overview of literature pertinent to general deception, cyber deception and deception experimentation involving human subjects. Chapter III covers the research design and methodology of the exploratory data analysis performed for this dissertation including an overview of the Tularosa study which generated the data used. Chapter IV discusses the Information Technology artifact

created to extract and transform the network traffic captured during the Tularosa study into a structure. Chapter V provides the statistical and graphical assessment of the temporal variables extracted from the exploratory data analysis performed. Chapter VI covers the Principal Component Analysis (PCA) and Exploratory Factor Analysis (EFA) techniques used to investigate the data. Chapter VII provides two theoretical models of defensive cyber deception for consideration. Chapter VIII concludes with key findings, contributions, limitations, and recommendations for future research.

THIS PAGE INTENTIONALLY LEFT BLANK

II. LITERATURE REVIEW

This chapter provides background information on the complex phenomenon of cyber deception, beginning with the medium through which cyber deception occurs, cyberspace, then exploring classical deception theory. Many definitions for "cyberspace" are available, with the definition provided by Joint Publication 3-12 Cyberspace Operations being commonly referenced as "A global domain within the information environment consisting of the interdependent networks of information technology infrastructures and resident data, including the Internet, telecommunications networks, computer systems, and embedded processors and controllers" (Joint Chiefs of Staff [JCS], 2018, p. GL-4). Perhaps a more enduring definition is provided by a 1993 book on Supercomputing which offers cyberspace as "a new information reality, parallel to but distinct from our well-known physical reality" (Kaufmann & Smarr, 1993, p. n12). Regardless of definition, information plays a key role in cyberspace and is used as focal point throughout this dissertation for exploring cyber deception.

A. THEORY FRAMEWORK

With the complexity of the phenomena of deception, a wide range of theory on deception and related constructs has been explored. Given the focus on cyber deception as a defensive measure to prevent cyber attackers from achieving their goals, emphasis has been placed on exploring deception theory related to military conflict. The interdisciplinary nature of cyber deception presents several challenges in conducting a thorough literature review across multiple academic literature data repositories. An IT artifact was developed for use as a computational literature review tool to conduct the comprehensive literature review across a wide range of literature types and sources needed to identify works related to deception, computer deception and cyber deception. This tool is documented in "A Network-Graph Based IT Artifact Aiding the Theory Building Process" published in the *Proceedings of the 55th Hawaii International Conference on System Sciences* and is publicly available at on NPS's GitLab site https://gitlab.nps.edu/michael.senft/graph-based-literature-review-tool (Senft & Corrado, 2022).

1. General Deception Theory

In 1982, Barton Whaley offered a preliminary general theory of deception based on his work exploring deception across fields where deception is prevalent, magic and war (Whaley, 1982). This general theory of deception provides both a prescriptive and descriptive framework for the phenomena of deception (Whaley, 1982). Whaley also asserts "deception is the same regardless of whatever field it appears in. It is not a function of technology. All deceptions are applied psychology — the psychology of misperception" (Whaley, 1982, p. 179). Contemporary work by Donald Daniel and Katherine Herbig provides additional insights into the phenomena of deception with development of models and concepts of the deception process (Daniel & Herbig, 1982). J. Bowyer Bell further extends the work on developing deception theory in "Toward a Theory of Deception," which also explores the dynamics of counter-deception (Bell, 2003). A thorough synopsis of these theories and conceptual models is provided by Bennett and Waltz (2007).

2. Cue-Based and Content-Based Deception Theory

While deception in its many forms has been extensively researched, the bulk of research and related experimentation has been focused on observation of physical behavior for leakage of verbal or non-verbal cues (Cranford et al., 2021). In cyberspace however, physical cues are unavailable, making detection of deception challenging (Cranford et al., 2021). Conversation, where a sender and receiver exchange information over some medium, is used frequently in deception research to generate verbal and non-verbal cues for analysis and prediction (Burgoon et al., 1996; Burgoon et al., 2003; Newman et al., 2003; Hancock et al., 2004a; Hancock et al., 2004b). The extensive research on cue-based deception by Buller, Burgoon, Vrij and other researchers was reviewed, but considered as outside of the scope of the dissertation given this body of work is largely based on physical proximity between deceiver and individual being deceived. More recently, two new theories of deception focused on communication and information content have been offered – Truth-Default Theory (TDT) and Information Manipulation Theory 2 (IMT2) (Levine & McCornack, 2014). These theories differ from previously offered deception theories including Interpersonal Deception Theory, which seek to identify cues, such as

cognitive effort, that distinguish truth from deception (Levine & McCornack, 2014). TDT offers that in human communication, humans typically operate with a default presumption that the receiver of the communication is generally honest (Levine, 2014). IMT2 frames deception as the covert manipulation of information across multiple dimensions (McCornack et al., 2014). While TDT and IMT2 do not specifically address cyber deception, they provide a useful perspective on the larger phenomena of deception, suggesting humans presume honesty as a starting point in communication and describing impacts of covert information manipulation.

Instance-Based Learning Theory (IBLT) offered by Gonzalez et al. (2003) has been applied to generate cognitive models for the design of dynamic deception (Gonzalez et al., 2020). IBLT posits that human decision-making in dynamic tasks is a function of memory recall involving a process of recognition, judgement, choice, execution, and feedback (Gonzalez et al., 2003). IBLT has been utilized to explore both framing effects on the decision-making of cyber-attackers in an insider attack scenario and decision-making of human cyber-defenders (Cranford et al., 2020). While cognitive models developed for cyber-attackers and cyber-defenders using IBLT provide useful insights into the role of previous experience on performance of dynamic tasks, further experimentation is needed to evaluate the utility of IBLT for explaining and predicting cyber-attacker behavior under varying conditions of deception (Du et al. 2022).

3. Cyber Deception Theory

No theory of cyber deception is currently offered in peer reviewed literature, however use of deception in support of computer security was first documented in 1989 by Clifford Stoll in *The Cuckoo's Egg*. This book provides a firsthand account of Stoll's effort to seek out a cyber-attacker who gained access to computer systems at the Lawrence Berkeley National Laboratory, details how a honeypot with false information was used to assist in tracking a cyber attacker (Stoll, 1989). The role of deception for information protection was also explored by Fred Cohen in work on a Deception ToolKit in 1998 (Cohen, 1998). A 1999 paper on strategic cyber defense identified the goal of deception for cyber defense as "to create as much confusion and uncertainty as possible for potential

attackers, regarding the value and location of critical information systems and resources" (Tirenin & Faatz, 1999, p. 461). It is worth noting even when deception is not present "the cyber attacker can become confused, even if the defender does not do anything deliberately, simply because of the inherent complexity and uncertainty involved in accessing and understanding remote computer network targets" (Gartzke & Lindsay, 2015, p. 336).

The deception process provided by Whaley was adapted for computer security uses by Jim Yuill in a systematic exploration on the application of deception within information technology environments (Yuill, 2006). In this work Yuill suggested "Deception can be used to attack hackers' decision-making processes; thus deception provides an offensive security measure—something computer security defenders sorely lack." (Yuill, 2006, p. 2). Since 2006 there has been a dramatic increase in cyber deception research, which is heavily focused on development of systems and techniques for defensive deception. The application of game theory to cyber deception is also a popular topic within the field, which may reflect the importance of the attacker versus defender dynamic present in the phenomena of cyber deception.

In 2019, Pawlick et al. published "A Game-theoretic Taxonomy and Survey of Deception for Cybersecurity and Privacy." This work, which surveyed 24 articles from 2008 to 2018, provides a taxonomy of deception based on game-theoretic concepts, defining six types of deception in cyberspace: moving target defense, obfuscation, perturbation, mixing, honey-x and attacker engagement (Pawlick et al., 2019). In 2021 Cranford et al., published work on development of a cognitive theory of cyber deception based on signaling as a deception technique, manipulating information on the actual state of a network (Cranford et al., 2021). They offer that situations where cyber-attackers are aware that cyber deception may be present require decision making under uncertainty given that cues to detect cyber deception are only available through exploration (Cranford et al., 2021).

4. Information Theory

Concepts from information theory have also been previously applied to explore relationships between deception and communication. In a 2008 paper, Custy & Rowe offer

conditional entropy as a measure of effectiveness for a given deception technique based on the probability of the deception's success combined with the average uncertainty possessed by a subject about the state of their environment (Custy & Rowe, 2008). In "A Note on the Role of Deception in Information Protection" Cohen also highlighted that with the use of deception "the attacker's level of uncertainty rises" (Cohen, 1998, Theoretical Issues section). Deception succeeds when a subject is unable to correctly determine the state of their environment and fails when they can correctly determine it (Custy & Rowe, 2008). While much of traditional deception literature focuses on the success or failure of deception, perhaps evaluation of uncertainty generated in a given environment provides a better measure of deception effectiveness in the domain of cyberspace, given both its complex and dynamic nature, and that targeted systems are potentially under constant observation. Custy and Rowe do not directly address psychological deception, but their insights on analyzing deception effectiveness through quantification of uncertainty and their definition of deception based on presentation of a "specific false version of reality" provide useful intuition for investigation of the complex phenomenon of cyber deception in its entirety (2008, p. 6).

5. Decision-Making Under Uncertainty

While no theory of cyber deception is currently available, defining cyberspace as an information reality and classifying the goal of deception as increasing uncertainty through manipulation of information connects two key constructs with the reconnaissance task completed by subjects during the Tularosa study for the characterization of effective cyber deception. The basis for the reconnaissance task given to subjects during the Tularosa study was to collect information to reduce uncertainty for future operations. This task, focused on direct observation of the target network, is described in greater detail in Chapter III. Yuill et al. provide a detailed overview of a discovery methodology used by cyber-attackers to sense and receive information from computer targets using an iterative process of hypothesis generation and evidence accumulation (Yuill et al., 2006). In carrying out the reconnaissance task to reduce uncertainty, cyber-attackers and subjects during the Tularosa study are faced with making choices under uncertainty. Busemeyer denotes a distinction between decision making under uncertainty and decision making under risk

(Busemeyer, 1985). He offers decision making under uncertainty involves "outcomes probabilities initially unknown and learned through experience" whereas decision-making under risk involves "perfect knowledge of outcome probabilities" (Busemeyer, 1985, pp. 560–561). As subjects during the Tularosa study lack knowledge of outcome probabilities, emphasis is placed on decision-making under uncertainty.

Many theories for choice under uncertainty exist, but Tversky & Kahneman offer that they "commonly specify 1) the objects of choice, 2) a valuation rule, and 3) the characteristics of the functions that map uncertain events and possible outcomes into their subjective counterparts" (Tversky & Kahneman, 1992, p. 316). Furthermore, they suggest "Theories of choice are at best approximate and incomplete" as choice is "a constructive and contingent process" (Tversky & Kahneman, 1992, p. 317). In lieu of investigating theories of choice under uncertainty to the reconnaissance task given the subjects in the Tularosa study, focus is instead placed on exploring the nature of uncertainty faced by these subjects.

At a high-level, cyber-attackers are faced with three types of uncertainty in conducting reconnaissance activities against a target network, aleatory, epistemic, and model. As offered by Kozyreva et al. "epistemic uncertainty refers to incompleteness of knowledge, whereas aleatory uncertainty stems from the statistical properties of the environment that exist independent of a person's knowledge" (Kozyreva et al., 2019, p. 349). Both the terms "aleatory" and "ontic" have been used to describe uncertainty created by variability, but for consistency the term "aleatory" will be used to describe this form of uncertainty as the term "ontic uncertainty" has a specific connotation when describing quantum systems. (Walker et al., 2003; Kvam et al., 2021). The benefit of delineating between aleatory and epistemic uncertainty is highlighted by Kiureghian and Ditlevsen by identifying "which uncertainties can be reduced and which uncertainties are less prone to reduction" (Kiureghian & Ditlevsen, 2009, p. 109). Model uncertainty is a characterization of the sufficiency of the conceptual model used to understand a system and the implementation of the model (Walker et al., 2003). The combination of these uncertainties reflects the considerable potential for confusion of cyber-attackers even when deception is not present, as highlighted by Gartzke and Lindsay (Gartzke & Lindsay, 2015).

Swiler and Giunta provide more detailed definitions for both aleatory and epistemic uncertainty, defining aleatory uncertainty as characterization of "the inherent randomness in the behavior of the system under study" and epistemic uncertainty as characterization of the "the lack of knowledge about the appropriate value to use for a quantity that is assumed to have a fixed value in the context of a specific application" (Swiler & Giunta, 2007, Uncertainty Quantification section). Walker et al. further distills model uncertainty into two parts, model structure uncertainty and model technical uncertainty (Walker et al., 2003). They describe model structure resulting from "lack of sufficient understanding of the system (past, present, or future)" that is under study (Walker et al., 2003, p. 9). Model structure uncertainty includes "uncertainty associated with the relationships between inputs and variables, among variables, and between variables and output, and pertains to the system boundary, functional forms, definitions of variables and parameters, equations, assumptions and mathematical algorithms" (Walker et al., 2003, p. 10). Walker et al. offer that model technical uncertainty is uncertainty generated by computer errors in the implementation of this model, but in the context of reconnaissance model technical uncertainty could be considered as ambiguity created in application of the conceptual model of the target to the network environment presented using the technical tools available to the cyber-attacker.

In the Tularosa study, subjects experienced aleatory uncertainty as they lacked knowledge of the number and type of real or decoy systems present on the cyber range network. The statistical distribution of system types and number of systems on the cyber range network environment configuration were unknown to the subjects. These subjects experienced epistemic uncertainty due to an initial lack of knowledge about these systems present on the cyber range network. Epistemic uncertainty about these systems was reduced over time as subjects gathered more information about these systems. Model uncertainty is more challenging to quantify, but the conceptual model used by subjects during the Tularosa study to understand the cyber range network is generated by the collective education and experience unique to each individual. Changes in network activity patterns may suggest an impact from model uncertainty.

B. PRIOR CYBER DECEPTION EXPERIMENTATION INVOLVING HUMAN SUBJECTS

While controlled experimentation exploring the phenomenon of deception through the medium of cyberspace involving human subjects was first documented in 2001, such research is rare (Cohen et al., 2001). An extensive review of extant publicly available literature revealed five additional archetypes, with work by Rowe et al. in 2007, Heckman et al. in 2013, Ferguson-Walter et al. in 2020, Shade et al. in 2020 and most recently by Aggarwal et al. in 2021. Also noteworthy is the relatively small number of subjects participating in these experiments. For "Red Teaming Experiments with Deception Technologies" by Cohen et al. (2001), 15 subjects participated in the experiment. For "Fake Honeypots: A Defensive Tactic for Cyberspace" by Rowe et al. (2006), 14 subjects participated in the experiment. For "Active Cyber Defense with Denial and Deception: A Cyber-wargame Experiment" by Heckman et al. (2013), four teams participated in the experiment but the number of subjects per team was not specified. For "The Tularosa Study: An Experimental Design and Implementation to Quantify the Effectiveness of Cyber Deception" by Ferguson-Walter (2020), 139 subjects participated in the experiment. For "The Moonraker Study: An Experimental Evaluation of Host-Based Deception." by Shade et al. (2020), 59 subjects participated in the experiment. For "Decoys in Cybersecurity: An Exploratory Study to Test the Effectiveness of 2-sided Deception" by Aggarwal et al. (2021), 48 subjects participated in the experiment.

Of these experimental studies, only the Tularosa study conducted by Ferguson-Walter et al. recruited professional penetration testers providing subject matter expertise that closely matches unauthorized hacker activity (Ferguson-Walter, 2020). The Tularosa study recruited 139 professional red teamers, individuals trained to emulate cyber-attacks for security testing, to perform network penetration tasks on a controlled cyber range over two-day periods (Ferguson-Walter et al., 2019; Ferguson-Walter, 2020). The focus of the Tularosa study on recruiting a broad, random sample of professional penetration testers providing subject matter expertise that closely matches unauthorized hacker activity makes it an exemplar for the exploration of the phenomenon of cyber deception (Ferguson-Walter, 2020). While analysis of the Tularosa study data has already yielded empirical evidence on

the effectiveness of cyber deception for impeding attackers, many aspects of the wealth of data collected during this study have yet to be thoroughly evaluated providing opportunities to generate new insights from the data (Ferguson-Walter et al. 2021).

C. OPERATIONAL DEFINITIONS

The following operational definitions will be used for terms frequently referenced in this research:

- **Deception**: "the presentation of a specific false version of reality by a deceiver to a target for the purpose of changing the target's actions in a specific way that benefits the deceiver" (Custy & Rowe, 2008, p. 6).

- **Cyber/Computer-Security Deception**: "planned actions taken to mislead attackers and to thereby cause them to take (or not take) specific actions that aid computer-security defenses" (Yuill, 2006, p. 2).

- **Cyberpsychology-based Deception**: Actions taken to influence the mental model of a cyber attacker through manipulation of information about the possibility of cyber deception technology (Ferguson-Walter et al. 2021). The Tularosa study utilized cyberpsychology-based deception in the "Informed" condition (Ferguson-Walter et al. 2021).

- **Cyber Deception Technology**: Implementation of technical artifacts or measures within a network or computer system specifically designed to deceive a cyber attacker (Adapted from Ferguson-Walter et al. 2021). The Tularosa study utilized decoy systems as cyber deception technology in the "Present" condition (Ferguson-Walter et al. 2021).

- **Decoy System**: A physical or virtual device placed on a network whose primary purpose is to "obfuscate the true network assets and confuse the attacker about the true network topology" (Ferguson-Walter, 2020, p. 10).

- **Temporal Variable**: "property that takes on different values," which includes an element of time (Kerlinger & Lee, 2000, p.40).

- **Latency**: Measurement of time delay. For this research, latency is considered as the time between experiment trial start and event observation.

- **Duration**: "the time during which something exists or lasts" (Merriam-Webster, 2022, Duration section). For this research, duration is considered as the time between first and last occurrence of an event observation.

- **Frequency**: "the number of repetitions of a periodic process in a unit of time"(Merriam-Webster, 2022, Frequency section). For this research, frequency is considered as the number of event observations that occur over a given time period.

- **Aleatory Uncertainty**: characterization of "the inherent randomness in the behavior of the system under study" (Swiler & Guinta, 2007, Uncertainty Quantification section). "Aleatory uncertainties are characterized by frequency distributions" (Swiler & Guinta, 2007, Uncertainty Quantification section). Also termed "ontic uncertainty" (Walker et al., 2003).

- **Epistemic Uncertainty**: characterization of "the lack of knowledge about the appropriate value to use for a quantity that is assumed to have a fixed value in the context of a specific application" (Swiler & Guinta, 2007, Uncertainty Quantification section). "Epistemic uncertainties are characterized degrees of "belief" which can be improved through accumulation of additional information (Swiler & Guinta, 2007, Uncertainty Quantification section).

- **Model Uncertainty**: characterization of the sufficiency of the conceptual model used to understand a system and the implementation of the model (Walker et al., 2003).

- **Reconnaissance**: "techniques that involve adversaries actively or passively gathering information that can be used to support targeting" (MITRE, 2022, Tactics section).

THIS PAGE INTENTIONALLY LEFT BLANK

III. RESEARCH DESIGN

A. INTRODUCTION

Research for this dissertation was conducted in three phases: (1) Subject Network Activity Timeline Reconstruction, (2) Exploratory Data Analysis, and (3) Factor Analysis. The first phase, Subject Network Activity Timeline Reconstruction, focused on the reconstruction of subject activity timelines from network data captured during the Tularosa study. This step required development of an IT artifact to extract the temporal variables of latency, frequency, and duration for use in the next step, Exploratory Data Analysis (EDA). EDA of subject network activity consisted of assessing these temporal variables guided by a previously generated set of hypotheses used for analysis of data from Tularosa study. Factor Analysis of subject activity using both Principal Component Analysis and Exploratory Factor Analysis provides additional exploration of relationships within the data by investigating variance across the variables examined.

Prior to the start of research for this dissertation, an Institutional Review Board (IRB) Human Subject Research determination request was submitted to the NPS IRB due to the data having been generated by human subject activity. The NPS IRB determined that this research did not involve human subject research based on de-identification procedures applied to the data by the data owner. A copy of the determination is available upon request to the author.

B. RESEARCH DATA SOURCE

1. Tularosa Study Overview

The Tularosa study is an exemplar in cyber deception experimentation for both its focus on recruiting subjects with expertise that closely matches unauthorized hacker activity interest and the breadth and depth of technical and non-technical data collected (Ferguson-Walter, 2020). The Tularosa study recruited over 130 subjects conducting network penetration tasks on a controlled cyber range under varying conditions of technical deception (presence) and psychological deception (awareness). The conditions of deception presence and awareness were manipulated for subjects performing network

penetration tasks on the established cyber range in separate 7-hour blocks, which included a 30-minute break for lunch. These conditions are depicted in Figure 1 along with the specific dependent variables of interest previously described of latency, frequency and duration for real and decoy systems based on the experimental conditions.

		Deception Awareness	
		Not Informed	**Informed**
Deception Presence	**Absent**	**Real System** Latency Duration Frequency	**Real System** Latency Duration Frequency
	Present	**Real System & Decoy System** Latency Duration Frequency	**Real System & Decoy System** Latency Duration Frequency

Figure 1. Tularosa Study Experimental Conditions

At the beginning of each experimental trial, subjects were presented with a task briefing that outlined purpose and objectives for the penetration testing exercises. The task briefing provided for the "Not Informed" condition is depicted in Figure 2 and the task briefing provided for the "Informed" condition shown in Figure 3. The only difference between the two task briefings is the inclusion of the phrase "There may be deception on the network" for participants in the "Informed" condition.

"You represent an APT group attempting to gather information....

You have achieved an initial foothold on the company network, and now must discover as much as you can about potentially valuable targets on the network. You will conduct recon on the network and locate vulnerable services, misconfigurations, and working exploits....

Your objective is to collect as much relevant information about the target network as you can in the allotted time without compromising future network operations...

When you learn potentially useful information about target systems on this network you will immediately report this information to your team"

Figure 2. Tularosa Task Instructions for "Not Informed" Condition. Source: Ferguson-Walter et al. (2021).

"You represent an APT group attempting to gather information....

You have achieved an initial foothold on the company network, and now must discover as much as you can about potentially valuable targets on the network. You will conduct recon on the network and locate vulnerable services, misconfigurations, and working exploits....

Your objective is to collect as much relevant information about the target network as you can in the allotted time without compromising future network operations...

When you learn potentially useful information about target systems on this network you will immediately report this information to your team

There may be deception on the network"

Figure 3. Source: Tularosa Task Instructions for "Informed" Condition. Source: Ferguson-Walter et al. (2021).

For the conditions of deception presence, the absent condition included only real systems (50) on the cyber range with an even split between Windows and Linux Operating Systems (OS), whereas the present condition included both real systems (50) with the same even split between Windows and Linux OSs and decoy systems (50) with half of the decoys emulating Windows OSs and half emulating Linux OSs on the cyber range. The term

Absent-Uninformed, abbreviated AU, is used to denote the experimental condition where decoys were not present and subjects were not provided with information about deception on the network. The term Absent-Informed, abbreviated AI, is used to denote the experimental condition where decoys were not present and subjects were provided with information that there may be deception on the network. The term Present-Uninformed, abbreviated PU, is used to denote the experimental condition where decoys were present and subjects were not provided with information about deception on the network. Finally, the term Present-Informed, abbreviated PI, is used to denote the experimental condition where decoys were present and subjects were provided with information that there may be deception on the network.

2. Tularosa Study Data Collection

Over 516 gigabytes of data capturing subject activity across more than a dozen file formats was collected (Ferguson-Walter, 2020). The data includes keylog data, network traffic, Intrusion Detection System (IDS) alerts, participant logs, decoy alerts and survey data. (Ferguson-Walter, 2020). The type and quantity of data collected ranges across more than a dozen file formats presents a considerable challenge for analysis, which is frequently encountered by researchers studying human interaction with IT systems and networks. Several sources of the data collected during the Tularosa study included timestamped information which would enable reconstruction of subject activity timelines (Ferguson-Walter, 2020). A Network Time Protocol (NTP) server was deployed on the cyber range networks to synchronize timing across systems within the experiment environment. Several considerable technical challenges impeded this reconstruction until analysis conducted for this dissertation including multiple data/time formats used across different sensor and log data, date/time data captured using multiple time zones, and a wide range of file types used to capture the data. Development of a robust approach to analyze the volume, velocity and variety of data generated during the Tularosa study was required to reconstruct detailed by-subject timelines of captured activity to extract the temporal variables for further analysis.

3. Tularosa Study Experimental Design

Communication between computer systems could be considered a type of conversation, with senders and receivers transmitting information via digital communications over a network. For the Tularosa study, nearly all communication between computer systems within the cyber range environment was the result of human interaction with a single computer system provided to the subject for use during each experiment trial. Network traffic captured during the experiment is therefore reflective of human subject behavior and serves as a rough analog to a traditional conversation, with packets generated and transmitted between systems serving a similar role as words and other verbal utterances. Temporal aspects of conversation have been used widely in research across a wide range of fields, from detecting dementia to detecting deception, which highlights the utility of temporal-based variables for exploring the phenomenon of cyber deception (Meilán et al., 2014; Boltz et al., 2010).

Unlike traditional conversation, where the sender is typically aware of all potential receivers, subjects in the Tularosa study were only provided with a network range (x.x.x.x/24), which contained potentially up to 251 other computer systems. Only 251 addresses were available out of the 254 assignable addresses because three IP addresses were already assigned, one for the router network interface, one for the network timing server, and one for the subject's computer system. Subjects were specifically informed not to perform attacks on the network timing server. The experimental setting experienced by subjects during the Tularosa study is depicted in Figure 4. The subject's initial understanding of the network environment consists of these three previously identified devices and an unknown number of additional systems within unknown OS types. All subjects started each experiment trial at the same initial state of uncertainty about other systems on the network.

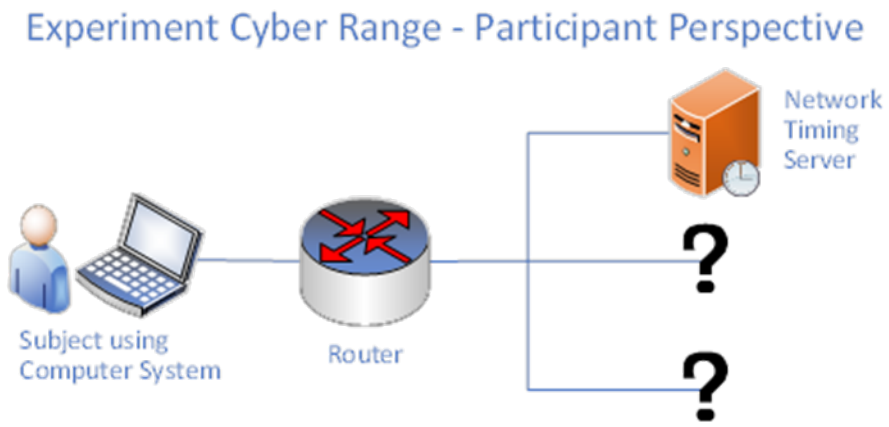


Figure 4. Tularosa Study Experiment Cyber Range

For subjects to initiate communication with any of the real or decoy systems on the network, they needed to identify system IP addresses using either passive or active reconnaissance using manual and/or automated tools. Once system IP addresses were identified, subjects could conduct further interaction with these systems to locate vulnerable services, misconfigurations, and valid exploits. Based on technical requirements for system interactions, variables of latency, duration, and frequency of communication between the subject's system and other systems on the network were identified as measurements that could be assessed to explore the impact of varying conditions of cyber and psychological deception on cyber-attacker behavior. The experiment cyber range conditions for the absence and presence of cyber deception are depicted in Figures 5 and 6 respectively.

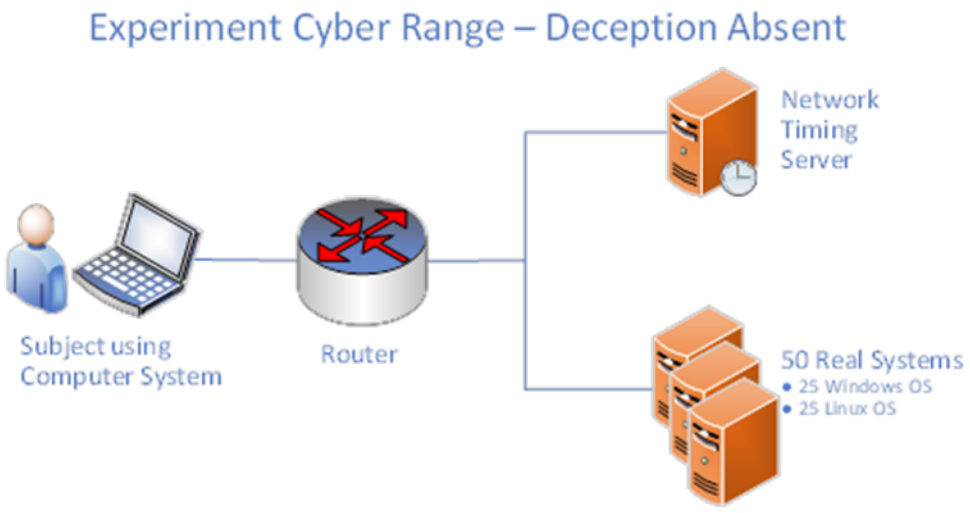


Figure 5. Deception Absent Cyber Range Configuration

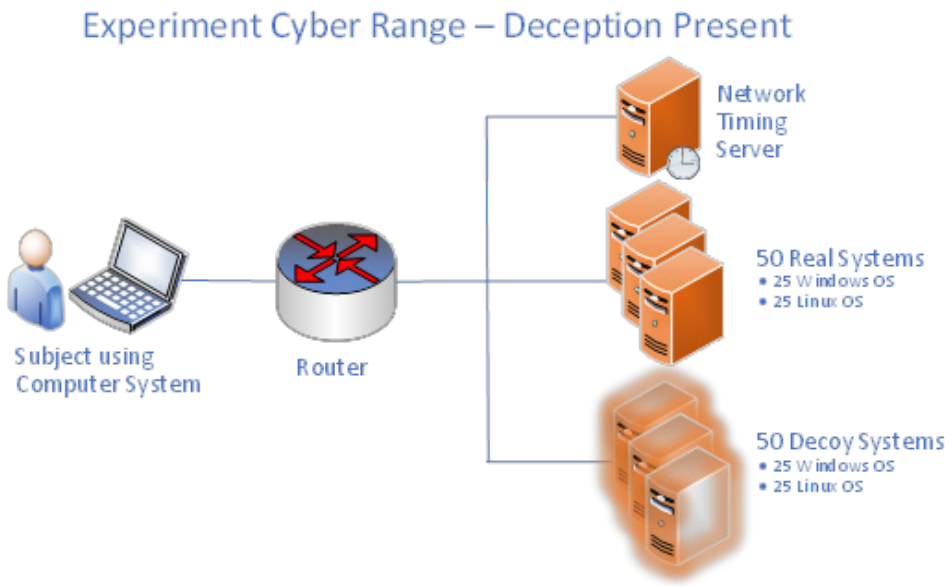


Figure 6. Deception Present Cyber Range Configuration

C. RESEARCH METHODOLOGY

1. Subject Network Activity Timeline Reconstruction and Temporal Variable Extraction

The Tularosa study captured a wide range of data including technical cyber data from computer systems and networks, questionnaire responses, cognitive battery assessments and physiological data as listed in Figure 7. The host and network data consisting of system logs, user keystrokes, full packet capture data, intrusion alerts and decoy alert logs contain referential timestamp information, which could be used to reconstruct a detailed timeline of user activity captured each day from their activity on the cyber range.

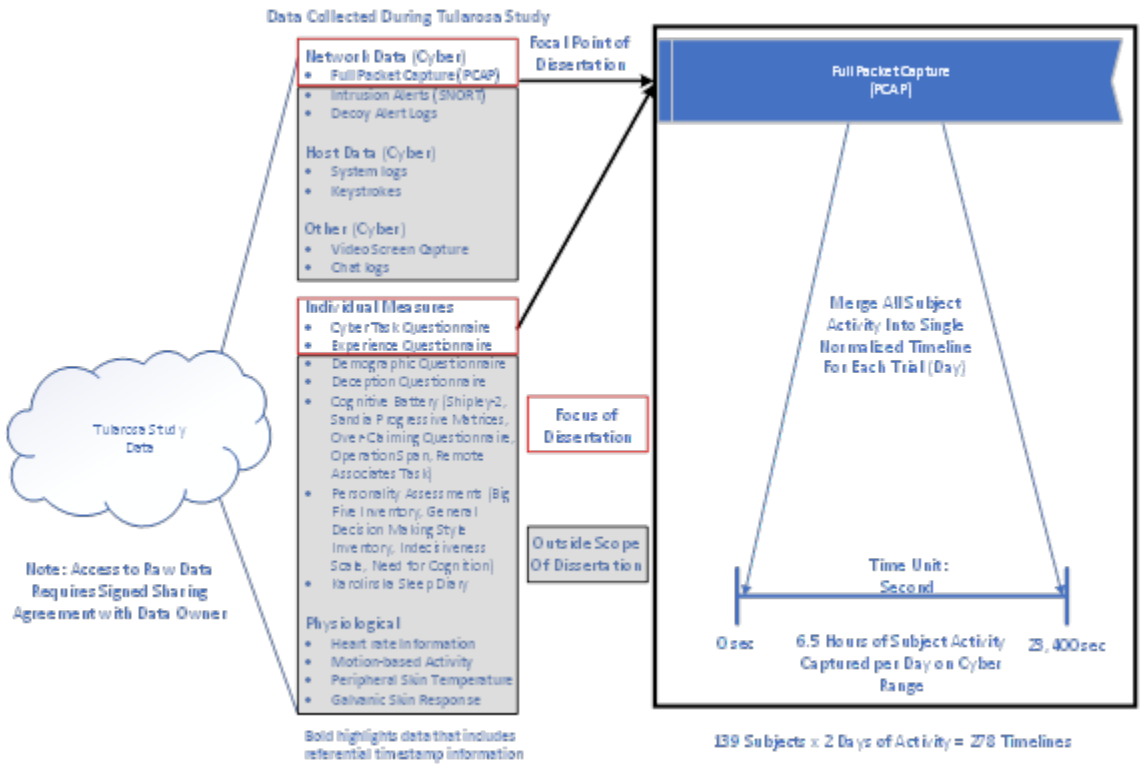


Figure 7. Subject Network Activity Timeline Reconstruction Overview

Each subject conducted network penetration activity under specific conditions of cyber deception presence and awareness on each day, which resulted in a total of 278 timelines potentially available for reconstruction. From these timelines of subject attack

behavior, as could be observed by a network defender, it was possible to extract temporal variables measuring latency, duration, and frequency of this activity.

2. Exploratory Data Analysis of Subject Network Activity

Extracting these temporal variables from the Tularosa study network traffic data enables assessment of these variables against previously offered hypotheses used to evaluate other data collected during the Tularosa study. In her dissertation, *An Empirical Assessment of the Effectiveness of Deception for Cyber Defense*, the Tularosa study lead Dr. Kimberly Ferguson-Walter offered five hypotheses that were evaluated from the original data collection. The fifth hypothesis concerned cognitive biases, which are outside the scope of research conducted for this dissertation. In a more recent work, "Examining the Efficacy of Decoy-based and Psychological Cyber Deception" Ferguson-Walter et al. (2021) offered four hypotheses further refined from Ferguson-Walter's original work, which are used in this research to assess the impact of varying experimental conditions of defensive cyber deception have on the timing of cyber-attack behavior using the temporal variables extracted from the network traffic captured during the Tularosa study

- **Hypothesis H1**: Defensive cyber tools and psychological deception impede attackers who seek to penetrate computer systems and exfiltrate information.
  Comparison Groups: Control (AU) versus Experimental (AI, PU, PI)
- **Hypothesis H2**: Defensive deception tools are effective even if an attacker is aware of their use.
  Comparison Groups: Uninformed (AU, PU) versus Informed (PI)
- **Hypothesis H3**: Cyber deception is effective if the attacker merely believes it may be in use, even if it is not.
  Comparison Groups: Control (AU) versus Psychological Deception (AI)
- **Hypothesis H4**: Cyber and psychological deception affects an attacker's cognitive and emotional state.
  Comparison Groups: Control (AU) versus Experimental (AI, PU, PI)

The first three of these hypotheses can be directly assessed using the temporal variables extracted from the Tularosa study network traffic. The fourth hypothesis can only be assessed indirectly using these variables.

Previous assessment of these hypotheses, with statistically significant differences between experimental groups have been performed using a wide range of dependent variables including keystroke count, commands with real IP addresses, bytes to real IP addresses, self-reported EternalBlue exploit attempts, self-reported exploit successes, Snort alert counts, time to first decoy alerts, decoy touch alerts, decoy scan alerts, decoy probe alerts, number of intrusion alerts, number of all decoy alerts, self-reported number of vulnerabilities identified, self-reported levels of surprise, confusion, frustration and suspicion (Ferguson-Walter, 2020, Ferguson-Walter et al., 2021). One item to note from this list is that most of the dependent variables assessed were derived from information only available from an attacker's vantage point. While obtaining information from an attacker is possible during experimentation, it is not possible in an operational setting, where attackers go to great lengths to remain hidden from network defenders. Only the Snort and decoy alerts previously explored would be available to network defenders in an operational setting.

In contrast, the latency, duration, and frequency of communication between the cyber-attacker's system and other systems on the network are readily available to network defenders. Reconstruction of subject activity timelines enables the extraction of these dependent variables for each system on the cyber range. For conditions involving the absence of deception, only 50 real systems were present. For conditions involving the presence of deception, 50 real systems and 50 decoys systems were present. Visual depictions of the latency, duration and frequency are illustrated in Figures 8, 9 and 10 respectively with detailed definitions for each variable available in Chapter II, Section C.

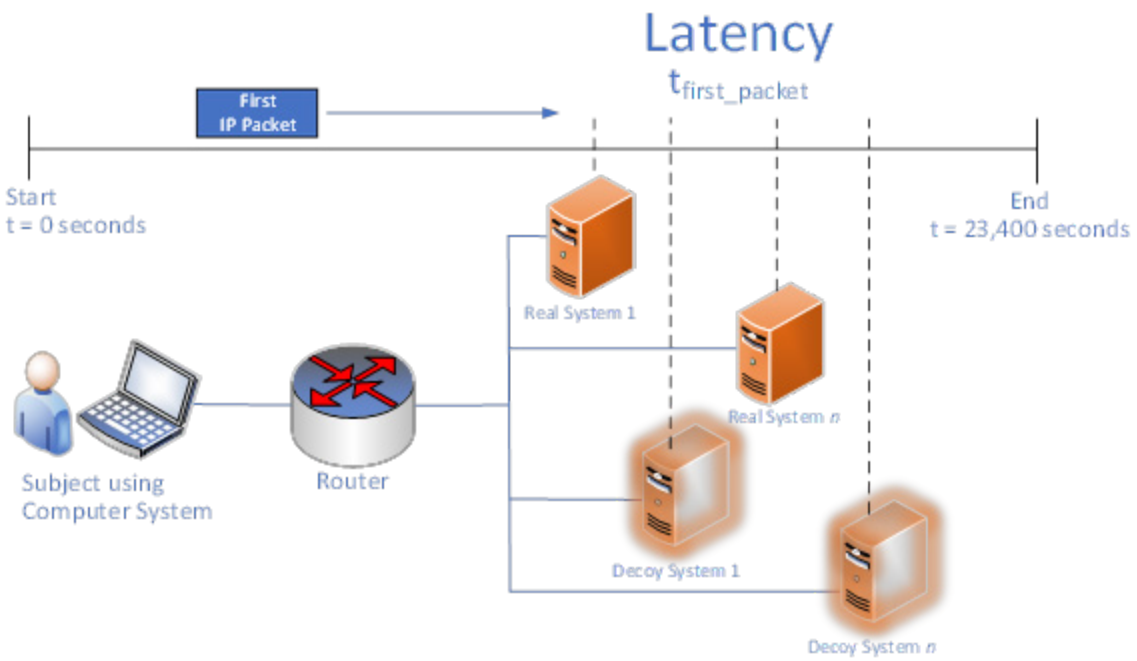


Figure 8. Visual Depiction of Latency Measurement

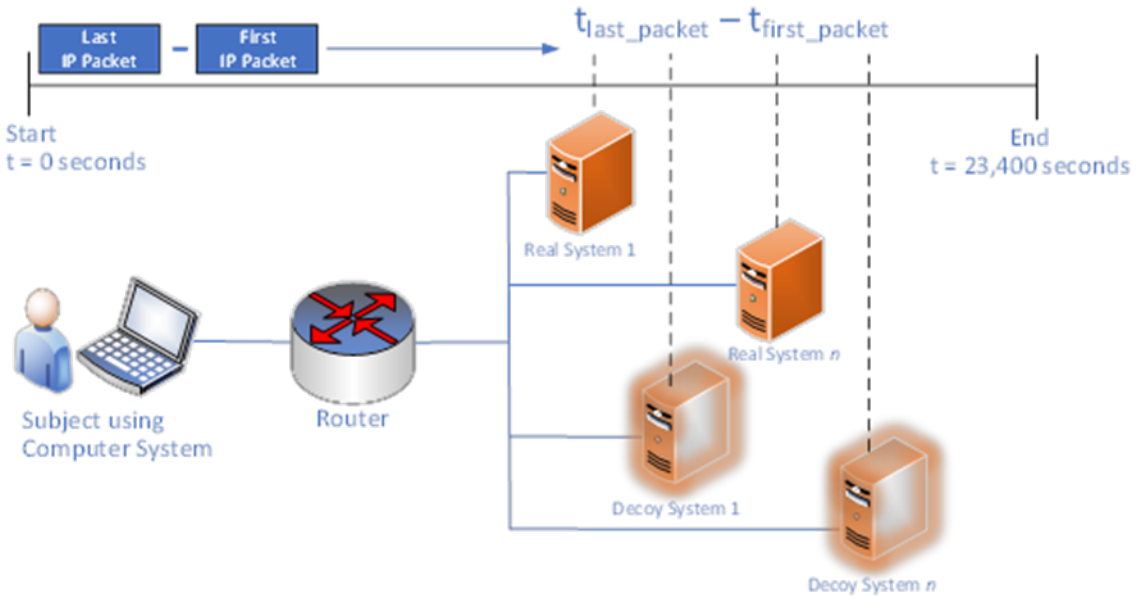


Figure 9. Visual Depiction of Duration Measurement

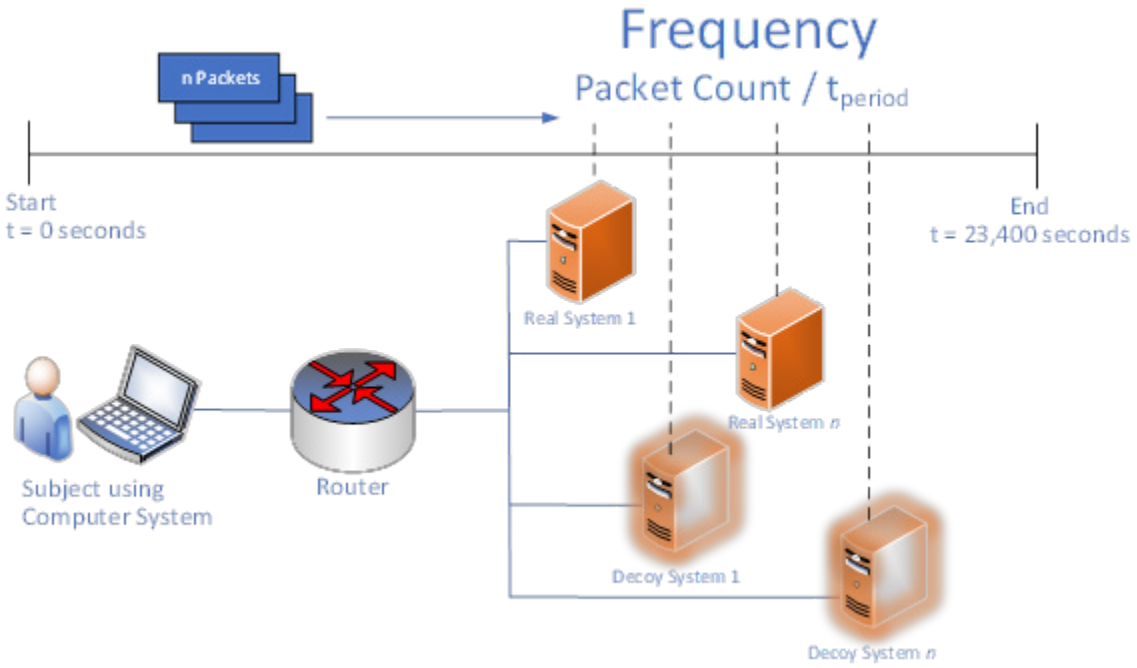


Figure 10. Visual Depiction of Frequency Measurement

Evaluation of previously offered hypotheses using this new set of temporal variables enables comparison with previous analysis for strength and direction of correlation. Comparison of analyses may also yield new insights into specific measurements with correlations that strongly support the offered hypotheses. The reconstruction of subject activity timelines creates additional opportunities for further analysis within and across experimental groups.

3. Factor Analysis of Subject Activity

Factor Analysis techniques are available to both generate a smaller set of representative variables from multiple correlated variables and explore the underlying data structure for the observed variables. Principal Components Analysis (PCA) and Exploratory Factor Analysis (EFA) are the most used Factor Analysis for multiple correlated variables but generate fundamentally different outputs. Separate analyses using PCA and Exploratory Factor Analysis EFA were used to generate insights from the Tularosa study using these data techniques to identify principal components and latent

constructs (Tabachnick & Fidell, 2013). The use of two related statistical techniques on the same dataset provides both reflective and formative perspectives on the data to explore the phenomenon of cyber deception, which is needed given the nascent state of current theory (Coltman et al., 2008). PCA considers the total variance of the observed variable, whereas EFA takes into account only the common variance of an observed variable illustrated in Figure 11.

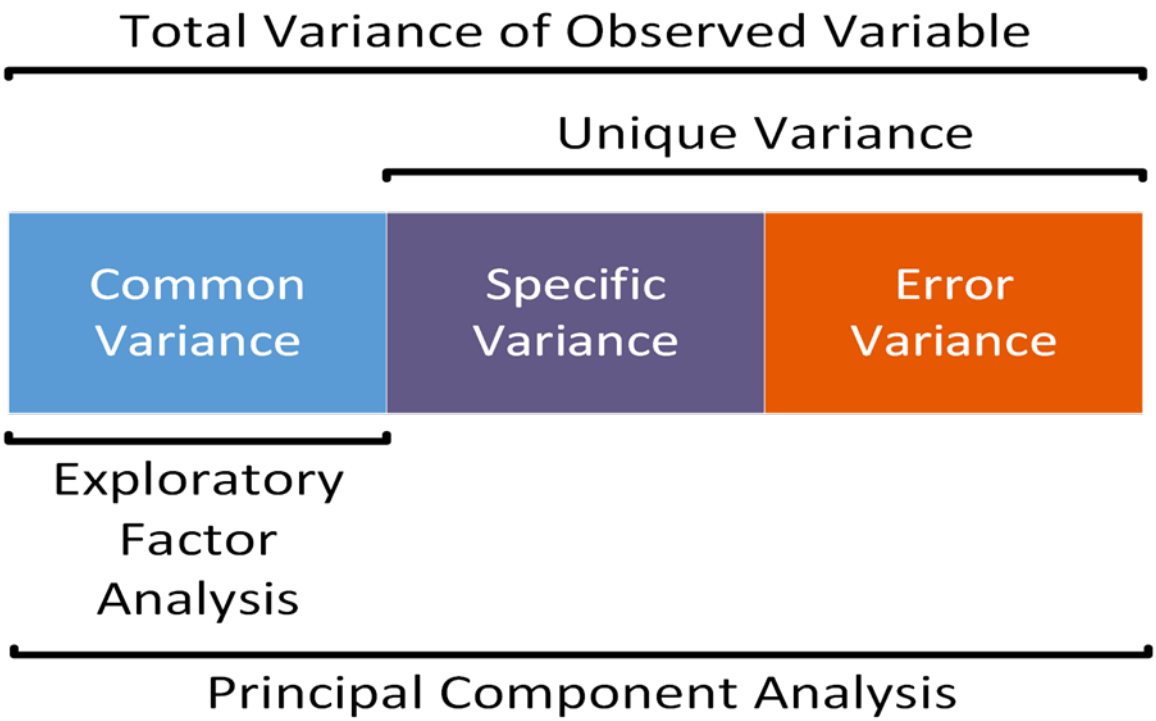


Figure 11. PCA and EFA Variance Comparison. Adapted from Schreiber (2021).

As highlighted by Joliffe "Despite their different formulations and objectives, it can be informative to look at the results of both techniques on the same data set. Each technique gives different insights into the data structure, with PCA concentrating on explaining the diagonal elements, and factor analysis the off-diagonal elements, of the covariance matrix, and both may be useful" (Joliffe, 1986, p. 128). The difference between the reflective

perspective generated by EFA and the formative perspective generated by PCA is illustrated in Figure 12.

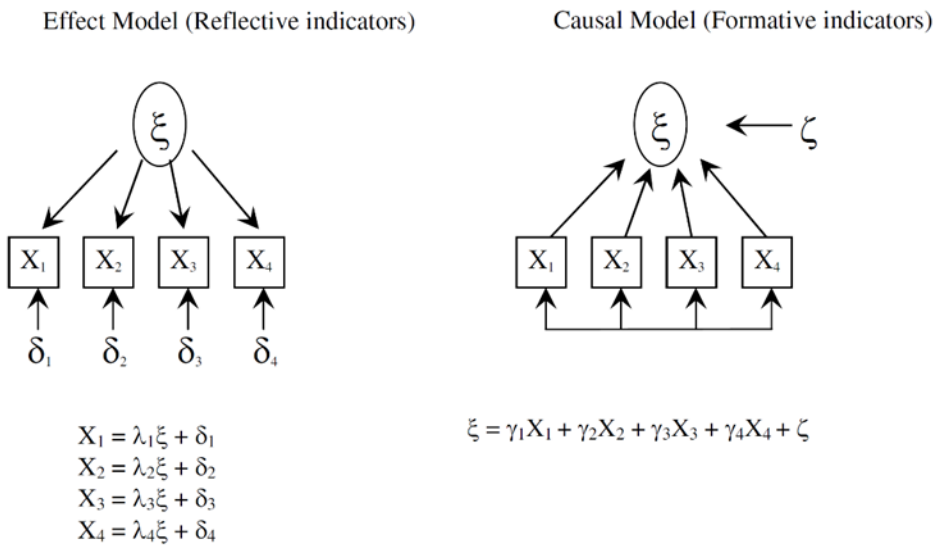


Figure 12. Reflective and Formative Perspectives. Source: Coltman et al. (2008).

PCA reduces measured variables into a smaller number of components which summarize their variance, while EFA identifies the number and nature of latent variables that account for variation between measured variables (Tabachnick & Fidell, 2013). PCA generates a formative model, which assumes the component identified is dependent on the measured variables, while EFA generates a reflective model, which assumes the latent constructs identified are independent of the measured variables. (Coltman et al., 2008). Typically, only one technique is applied against a given dataset, but the nascent state of cyber deception theory has little to offer in suggesting what role components or latent variables contribute to the phenomenon. Factor analysis has a rich history in application to research involving latent variables, making it a good fit in addressing the challenge described by Whaley (Haig, 2005). The use of both PCA and EFA on data gathered during the Tularosa study provides multiple perspectives on the phenomenon of cyber deception, which are useful for comprehensive investigation of the complex phenomenon of cyber deception.

D. JUSTIFICATION FOR SECONDARY ANALYSIS OF PREVIOUSLY COLLECTED EXPERIMENTAL DATA

Secondary analysis of existing data provides an opportunity to make significant contributions to both theory and practice through focused investigation of experimental data collected at both significant effort and cost. Costs to conduct cyber-related experimentation involving the number of human subjects recruited for the Tularosa study could easily exceed a million dollars, not to mention considerable lead times needed to execute an approved study. Secondary analysis does, however, have several drawbacks, most notably the lack of a comprehensive understanding of the data collection process and any issues experienced that may impact the accuracy and validity of the data. This limitation is partially mitigated through direct collaboration with members of the Tularosa study who were involved with the original collection of the data.

E. VALIDITY

In exploring the complex phenomenon of cyber deception, validity is a critical concern. As cautioned by Kerlinger and Lee, "Are we measuring what we think we are measuring?" (Kerlinger & Lee, 2000, p. 666). Campbell and Stanley highlighted a distinction between internal validity, "did in fact the experimental treatments make a difference" and external validity, "to what populations, settings, treatment variables, and measurements can this effect be generalized" (Campbell & Stanley, 1963, p. 5). While this research involves secondary analysis of previously collected experimental data where additional controls cannot be applied to the experiment design used, internal validity concerns for the original experiment design should be mentioned.

1. Threats to Internal Validity

From the internal validity concerns described by Campbell and Stanley, three sources stand out: (a) maturation; (b) testing; and (c) instrumentation (Campbell & Stanley, 1963). Given the experiment occurs over a 6.5-hour period, subject activity will potentially be impacted by fatigue and/or boredom making observed behaviors likely influenced by the passage of time. Subject behavior was also likely impacted depending on whether activity was captured on the first day or the second day of the experiment due to the

influence on behavior from repeated testing. The complexity of the cyber range used to collect data also raises concerns on the measurement instrumentation. While collection of data from digital sensors provides high fidelity measurements, minute changes to any of the device configurations on the cyber range could create cascading effects that could impact calibration of the measurements captured.

2. Threats to External Validity

The generalizability of this research must also be addressed. From the original experiment design, use of a broad, random sample of a population of professional red teamers as experimentation subjects increases the generalizability of conclusions offered by this research. Ferguson-Walter does highlight that the cyber range environment used during the Tularosa study is not representative of "in the wild" cyber environments, which impacts external validity (Ferguson-Walter, 2020, p. 50). The primary statistical validity threat comes from the low statistical power of non-parametric tests, which were needed due to the non-normal data analyzed. This threat was partially addressed by the analysis of a large number of temporal variables across different experimental conditions.

Selection bias is a significant concern for external validity from the use of exploratory data analysis. From the volume of previously analyzed data available, a large number of dependent variables could potentially be extracted. Justification for variables extracted and analyzed to evaluate the previously offered hypotheses was needed to address selection bias concerns. Finally, construct validity exploration was a primary focus for Factor Analysis of subject activity of this research. Given the nascent state of cyber deception theory, identification of constructs or factors that account for the effectiveness, or lack of effectiveness, of cyber deception has not yet been empirically assessed. Use of PCA and EFA to evaluate constructs and factors that help explain total observed variance and observed common variance captured within the data set supported the selection of variables analyzed.

IV. DATA EXTRACTION AND TRANSFORMATION

This section provides an overview of processes used to aggregate data collected during the Tularosa study for both extraction of the temporal variables of latency, frequency, and duration and by minute subject network activity for analysis of cyber-attack behavior captured during the Tularosa study

A. BACKGROUND

Analysis of the Tularosa study data focused on network traffic generated by human subjects using a computer system running Kali Linux connected to the network, which simulated a cyber-attacker's remote access to a target network. This network traffic data was captured using standard PCAP (Packet CAPture) format. PCAP provides a standardized data format containing information about network communication for Layers 2–7 of the OSI model. This highly compressed format requires command-line/text-based or graphical user interface (GUI) based tools to view and parse the data. *Tcpdump* is a commonly used command-line/text tool with *Wireshark* being a commonly used GUI tool.

A key challenge for analysis of the PCAP data containing subject activity captured during the Tularosa study was the volume of data captured with over 200 GB of PCAP data containing Day 1 and Day 2 subject activity. GUI tools used for analysis of PCAP data are typically limited to analysis of approximately 500 MB or less of PCAP data due to significant processing and memory resources required for large files. Figure 13 is a screenshot from *Wireshark* displaying PCAP data captured during subject activity in the Tularosa study. This limitation made the use of GUI-based tools for analysis of the Tularosa study network data infeasible due as subject activity routinely exceeded 2 GB of captured network activity in a single experiment trial. Command-line/text-based tools for PCAP analysis can easily handle large volumes of PCAP data (in excess of 2GB), but integration of other data sources containing additional context information with these tools creates a new set of challenges.

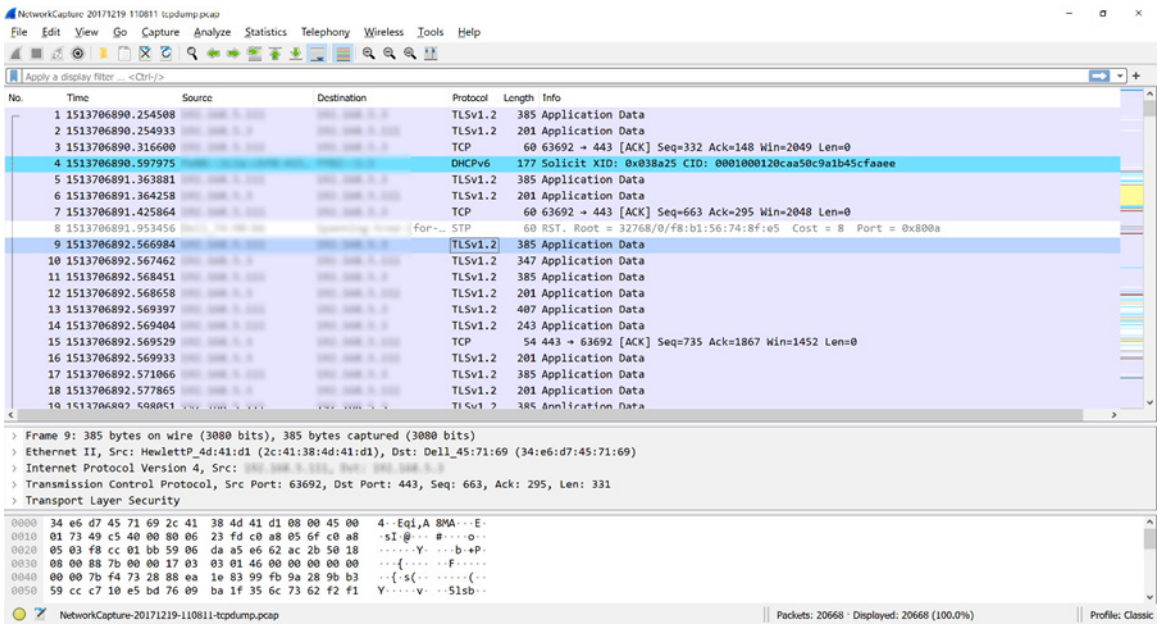


Figure 13. Wireshark Screenshot of PCAP data

To overcome these limitations, a network graph database was used for reconstruction of subject network activity. A network graph approach was used to efficiently import data and create relationships across multiple heterogeneous datasets. A time tree was created to link packets with timestamp data in Unix time based on actual time of capture to the relative time of capture from the start of each experiment trial. While use of a relational database was considered, the flexible data model schema offered by network graph databases combined with the ability for quick modification and integration of new sources of information was a driving factor in selection of a network graph-based approach. A flexible approach was needed for the initial exploration of the data given the breadth and depth of data captured during the Tularosa study. A secondary consideration for the use of a network graph database was the availability of open-source resources for increased accessibility and reuse of the IT artifacts developed during this research. Neo4j was selected as the network graph database platform on account of availability of an open-source version for non-commercial applications and ability to quickly ingest data in tabular/Comma Separate Value (CSV), JavaScript Object Notation (JSON) or eXtensible Markup Language (XML) formats (Neo4j, 2021). The first step to importing the PCAP

data into Neo4j was conversion of the highly compressed PCAP data into one of these formats. CSV was selected for simplicity in conversion of the PCAP data.

B. DATA EXTRACTION AND TRANSFORMATION PROCESS

1. PCAP Transformation Overview

Conversion of the highly compressed PCAP format to CSV with relevant data fields resulted in output file sizes between 1.5 to 2 times of the original PCAP file owing to the decompression of the highly compressed PCAP file. In the conversion of PCAP network traffic to CSV format, a total of 3692 files were created containing network activity for all 139 subjects. The *Pseudo-Code 1* block provides pseudo-code for the process of transforming network packet data (PCAP) traffic captured during the Tularosa study into a Comma Separated Value (CSV) formatted spreadsheet containing 18 information fields relevant to exploring attacker behavior. The full Linux script used can be found in Appendix A.

Pseudo-Code 1: PCAP file conversion to CSV Format

1: Identify PCAP files over 200 MB. PCAP files over 200 MB incur processing limitations
2: Split PCAP files over 200 MB into segments 200 MB or less
3: Use tshark to extract relevant information fields within PCAP files into CSV format

A lesson learned from the conversion of PCAP files to CSV format was in the selection of information fields within the PCAP files. Initially, the '_ws.col.Info' field containing additional protocol information for each packet was extracted, however, processing this field created several issues during the data import process. This field frequently contained characters in non-standard encodings causing errors during data import into Neo4j. Despite extensive troubleshooting, automated resolution of these errors could not be achieved, resulting in the data field being excluded from data import into Neo4j.

An example of the PCAP data extracted into CSV format is provided in Figure 14. These CSV spreadsheets were then imported into separate Neo4j databases for each subject.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	frame.number	frame.time_epoch	frame.len	eth.src	eth.dst	ip.src	ip.dst	ip.proto	_ws.col.Protocol	tcp.dstport	tcp.stream	udp.stream	tcp.srcport	udp.dstport	udp.srcport
2	1	1510010059	60	f8:b1:56:74:9c:52	01:80:c2:00:00:00				STP						
3	2	1510010060	42	34:e6:d7:45:71:8c	2c:41:38:4d:41:d1	[redacted]	[redacted]	17	UDP			0		35350	49652
4	3	1510010060	70	2c:41:38:4d:41:d1	34:e6:d7:45:71:8c	[redacted]	[redacted]	1	ICMP			0		35350	49652
5	4	1510010061	42	34:e6:d7:45:71:8c	2c:41:38:4d:41:d1	[redacted]	[redacted]	17	UDP			1		4236	49652
6	5	1510010061	70	2c:41:38:4d:41:d1	34:e6:d7:45:71:8c	[redacted]	[redacted]	1	ICMP			1		4236	49652
7	6	1510010061	42	34:e6:d7:45:71:8c	d0:57:7b:d0:cb:5a				ARP						
8	7	1510010061	60	d0:57:7b:d0:cb:5a	34:e6:d7:45:71:8c				ARP						
9	8	1510010061	60	f8:b1:56:74:9c:52	01:80:c2:00:00:00				STP						
10	9	1510010062	42	34:e6:d7:45:71:8c	2c:41:38:4d:41:d1	[redacted]	[redacted]	17	UDP			2		2555	49652
11	10	1510010062	70	2c:41:38:4d:41:d1	34:e6:d7:45:71:8c	[redacted]	[redacted]	1	ICMP			2		2555	49652
12	11	1510010062	42	34:e6:d7:45:71:8c	2c:41:38:4d:41:d1				ARP						
13	12	1510010062	60	2c:41:38:4d:41:d1	34:e6:d7:45:71:8c				ARP						
14	13	1510010063	42	34:e6:d7:45:71:8c	2c:41:38:4d:41:d1	[redacted]	[redacted]	17	UDP			3		49856	49652
15	14	1510010063	70	2c:41:38:4d:41:d1	34:e6:d7:45:71:8c	[redacted]	[redacted]	1	ICMP			3		49856	49652
16	15	1510010063	60	f8:b1:56:74:9c:52	01:80:c2:00:00:00				STP						
17	16	1510010064	42	34:e6:d7:45:71:8c	2c:41:38:4d:41:d1	[redacted]	[redacted]	17	UDP			4		46882	49652
18	17	1510010064	70	2c:41:38:4d:41:d1	34:e6:d7:45:71:8c	[redacted]	[redacted]	1	ICMP			4		46882	49652
19	18	1510010064	60	d0:57:7b:d0:cb:5a	ff:ff:ff:ff:ff:ff				ARP						
20	19	1510010065	42	34:e6:d7:45:71:8c	2c:41:38:4d:41:d1	[redacted]	[redacted]	17	UDP			5		44521	49652
21	20	1510010065	70	2c:41:38:4d:41:d1	34:e6:d7:45:71:8c	[redacted]	[redacted]	1	ICMP			5		44521	49652
22	21	1510010065	60	f8:b1:56:74:9c:52	01:80:c2:00:00:00				STP						
23	22	1510010066	42	34:e6:d7:45:71:8c	2c:41:38:4d:41:d1	[redacted]	[redacted]	17	UDP			6		41564	49652
24	23	1510010066	70	2c:41:38:4d:41:d1	34:e6:d7:45:71:8c	[redacted]	[redacted]	1	ICMP			6		41564	49652
25	24	1510010067	42	34:e6:d7:45:71:8c	2c:41:38:4d:41:d1	[redacted]	[redacted]	17	UDP			7		9930	49652
26	25	1510010067	70	2c:41:38:4d:41:d1	34:e6:d7:45:71:8c	[redacted]	[redacted]	1	ICMP			7		9930	49652
27	26	1510010067	60	f8:b1:56:74:9c:52	01:80:c2:00:00:00				STP						
28	27	1510010068	42	34:e6:d7:45:71:8c	2c:41:38:4d:41:d1	[redacted]	[redacted]	17	UDP			8		31535	49652
29	28	1510010068	70	2c:41:38:4d:41:d1	34:e6:d7:45:71:8c	[redacted]	[redacted]	1	ICMP			8		31535	49652

NetworkCapture-20171106-161419-

Figure 14. PCAP Data Extracted to CSV Format

With the Tularosa study PCAP data transformed into the CSV format supported by Neo4j, the next step was importing the data into Neo4j and enriching it with additional information. Figure 15 depicts the data schema used to model subject activity captured during the Tularosa study. The "NetFlow" node contains data imported from the PCAP in CSV format. Following the import of the PCAP data in CSV format, a time tree was created to connect the timestamp of each packet in Unix epoch time using Coordinated Universal Time (UTC) with the relative time of the experiment conducted in the Mountain Time Zone of North America. This time tree is represented by the "Second," "Minute," "Hour," "Day," "Month" and "Year" nodes. Each "Second" node was then updated to also include respective Unix epoch time for each second. Data from two additional spreadsheets was also imported, the first containing information on experiment start, lunch and end times, represented by "day1" and "day2" nodes and the second containing information on the computer systems present on the cyber range network for Day 1 and Day 2 experiment trials, represented by the "system" node.

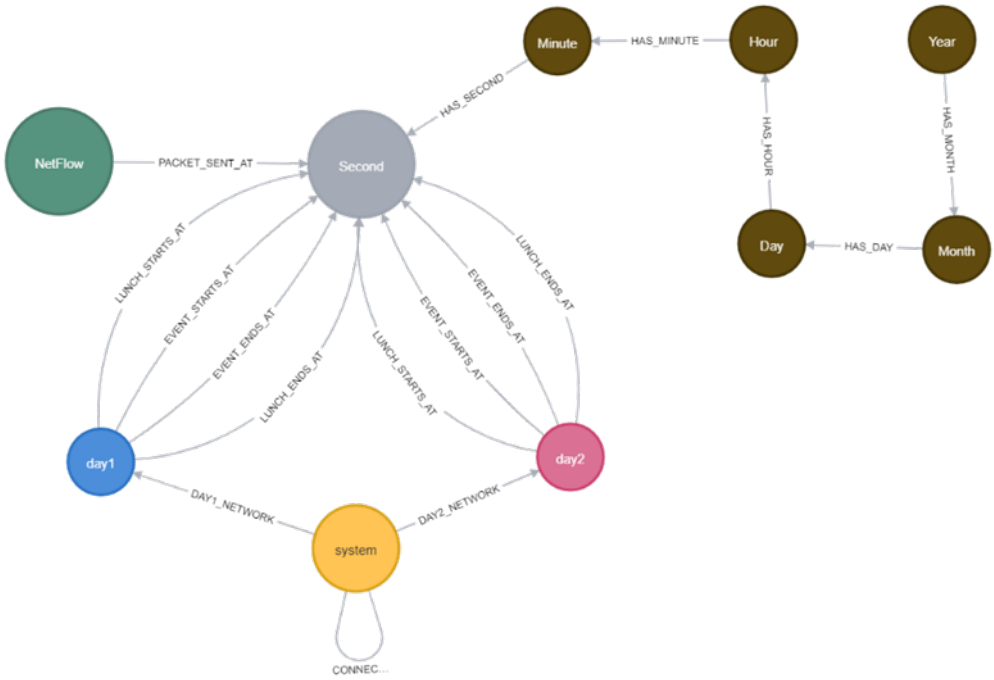


Figure 15. Data Model Schema for Tularosa Network Traffic Data

The "NetFlow" nodes were connected to the "Second" node by matching the Unix epoch times for each node. The "day1" and "day2" nodes were connected to the "Second" node by matching the year, month, day, hour, minute and second for the experiment start and end and lunch start and end with the matching second in the time tree. With this relationship created, the absolute time relative to the start of each experiment trial and excluding the lunch period can calculated and added as a property to the "Second" node for future queries of the data. A visual depiction of these relationships and properties is depicted in Figure 16.

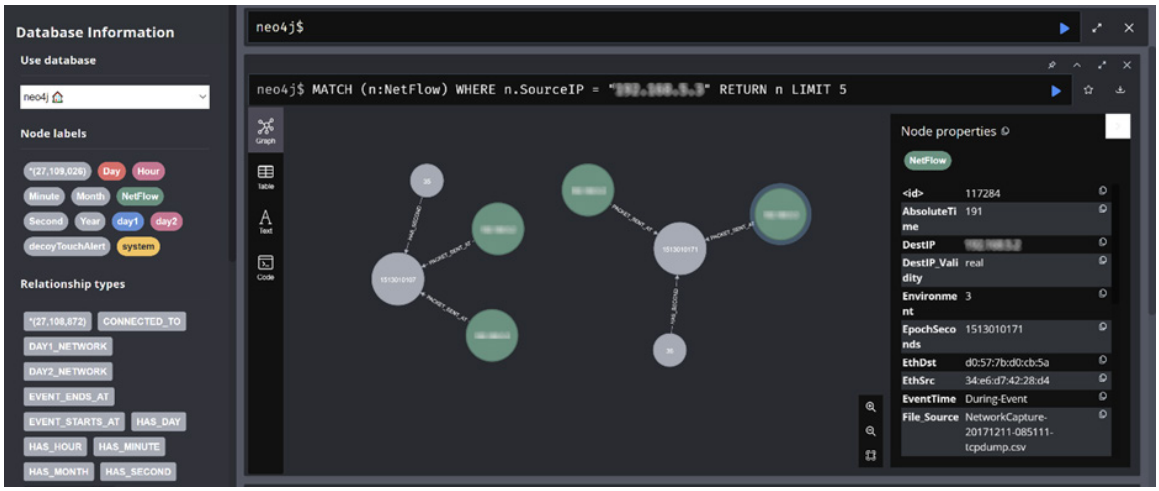


Figure 16. Time Relationships and Properties as Represented in Neo4j

Pseudo-Code 2 provides pseudo-code for the import and extraction of temporal variables from data collected during the Tularosa study. The full program written in a combination of Python and the Neo4j scripting language Cypher can be found in Appendix A

Pseudo-Code 2: Network Graph-based Temporal Variable Extraction Methodology

1: *Create network graph database and initialize Python environment with required modules*
2: *Create indexes within network graph database for high use data elements*
3: *Provide Subject ID number for data being imported*
4: *Create "Day1" and "Day2" node containing date and time information for each experiment trial for Day 1 and Day 2 trials*
5: *Create Unique Identifier (UUID) for experiment trial for tagging all related data*
6: *Create "System" nodes containing system information for the cyber range environment present for experiment trial*
7: *Create relationships between experiment "Day" and "System" nodes*
8: *Create time tree with nodes for year, month, day, hours, minutes and seconds elapsed during experiment trial based on "Day" node*
9: *Create relationships between "Year", "Day", "Hour", "Minute" and "Second" nodes*
10: *Create relationships between "Day1", "Day2" and corresponding "Second" nodes for the experiment trial start, experiment trial end, lunch time start and lunch time end.*
11: *Import network traffic data from CSV spreadsheets creating "NetFlow" nodes containing properties for each of the packet information fields*
12: *Create relationships between "NetFlow" and corresponding "Second" nodes based on matching UNIX epoch seconds*
13: *Set event property for all "Second" nodes based on when they occurred before, during or after the specified experiment trial times and if they occurred during the specified subject lunch period*
14: *Set absolute times for all "Second" nodes based on when they occurred from the initial experiment trial start time. Assign negative time values to "Second" nodes occurring outside of experiment trial (before start time, during lunch time, after end time)*
15: *Match "NetFlow" and "System" nodes based on Destination IP address and set "NetFlow" validity property to facilitate subsequent data queries*
16: *Set absolute times for "NetFlow" nodes based on connected "Second" nodes to facilitate subsequent data queries*
17: *Execute queries to return absolute time for the first and last packet sent to each IP addresses in the 192.168.5.0/24 network from any IP address and the number of packets sent to each of these addresses. To obtain the temporal variables of latency, duration and frequency for each IP address, the absolute time of the first packet provides the latency.*

2. PCAP Transformation Validation Check

On account of the relatively complex transformation and ingestion of the PCAP data into a network graph database, a validation check using the raw PCAP data was developed. *Pseudo-Code 3* block provides the pseudo-code used for this validation check for the extraction of temporal variables using the network graph based PCAP data transformation methodology previously described. This validation check uses Wireshark to parse original PCAP files to extract the temporal variables. However, processing limitations within Wireshark limit this validation check to subjects generating total PCAP

data capture of approximately 500 MB or less per experiment trial. The full Linux script used found in Appendix A.

Pseudo-Code 3: Wireshark-based Temporal Variable Extraction Validation

1: Merge all PCAP files captured for Subject experiment trial into single file
2: Split merged PCAP file into two distinct PCAP files, using experiment start and end time and start and end of lunch period during experiment. The first file will capture all traffic from the start of the experiment trial to the start of the lunch period. The second file will capture all traffic from the end of the lunch period to the end of the experiment trial
3: Merge two files from previous step into single file. This file captures all subject activity during designated experiment time frame.
4: Merge all PCAP files captured for Subject experiment trial into single PCAP file
5: Analyze combined PCAP file using native tools within Wireshark to identify time of first target system interaction (latency), duration of target system interaction (duration) and number of packets sent to target system (frequency)
6: Compare latency, duration and frequency values for subject system interactions with the values obtained using the network graph-based methodology

C. SUBJECT ACTIVITY EXPORT

1. Temporal Variables

Using a combination of Python and Cypher in a Jupyter Notebook, queries of subject activity were executed to export temporal variables of latency, duration, and frequency for each subject for each system on the network. The Cypher syntax used to extract this information is displayed in the Cypher Syntax 1 block and the spreadsheet output is depicted in Figure 17. Day 2 activity is extracted using the same syntax, only replacing (d:day1) with (d:day2) to match the appropriate network traffic captured during the second day of experimentation.

Cypher Syntax 1: Day 1 Subject Activity Extraction

```
WITH [<List of Destination IP Addresses>] AS systems UNWIND systems AS system MATCH (n:NetFlow),(d:day1) WHERE n.SubjectID = <Subject ID> AND d.Node_Key = n.Node_Key AND n.DestIP = system AND n.SourceIP IN [<List of Source IP Addresses>] AND n.EventTime = 'During-Event' RETURN d.SubjectID, d.Day, d.Environment, d.Condition, apoc.coll.sort(COLLECT(n.AbsoluteTime))[0] AS latency, SIZE(COLLECT(n.AbsoluteTime)) AS frequency, apoc.coll.sort(COLLECT(n.AbsoluteTime))[(SIZE(COLLECT(n.AbsoluteTime))-1)] AS last, n.DestIP, n.DestIP_Validity
```

	A	B	C	D	E	F	G	H	I	J	K	L
1	SubjectID	Day	Environme	Condition	latency	duration	frequency	last	SourceIP	DestIP	n.SourceIP_Validity	n.DestIP_Validity
1153	S116	1	3	PI	13546	74	14	13619	[redacted]	[redacted]	real	decoy
1154	S116	1	3	PI	6810	12159	29	18968	[redacted]	[redacted]	real	decoy
1155	S116	1	3	PI	10291	8678	31	18968	[redacted]	[redacted]	real	real
1156	S116	1	3	PI	9360	4227	38	13586	[redacted]	[redacted]	real	decoy
1157	S116	1	3	PI	13478	109	14	13586	[redacted]	[redacted]	real	decoy
1158	S116	1	3	PI	10737	3662	72	14398	[redacted]	[redacted]	real	real
1159	S116	1	3	PI	11200	2117	24	13316	[redacted]	[redacted]	real	decoy
1160	S116	1	3	PI	681	18946	3124	19626	[redacted]	[redacted]	real	real
1161	S116	1	3	PI	11357	137	18	11493	[redacted]	[redacted]	real	decoy
1162	S116	1	3	PI	10342	2069	117	12410	[redacted]	[redacted]	real	real
1163	S116	1	3	PI	18968	3	7	18970	[redacted]	[redacted]	real	decoy
1164	S116	1	3	PI	10755	8214	31	18968	[redacted]	[redacted]	real	real
1165	S116	1	3	PI	18968	3	7	18970	[redacted]	[redacted]	real	decoy
1166	S116	1	3	PI	10936	5742	228	16677	[redacted]	[redacted]	real	real
1167	S116	1	3	PI	10307	3959	74	14265	[redacted]	[redacted]	real	real
1168	S116	1	3	PI	18968	3	7	18970	[redacted]	[redacted]	real	decoy
1169	S116	1	3	PI	16474	2495	14	18968	[redacted]	[redacted]	real	real
1170	S116	1	3	PI	13561	6	8	13566	[redacted]	[redacted]	real	decoy
1171	S116	1	3	PI	18968	1	7	18968	[redacted]	[redacted]	real	decoy
1172	S116	1	3	PI	10721	8248	43	18968	[redacted]	[redacted]	real	real
1173	S116	1	3	PI	18968	1	7	18968	[redacted]	[redacted]	real	decoy
1174	S116	1	3	PI	13409	178	14	13586	[redacted]	[redacted]	real	decoy
1175	S116	1	3	PI	9869	9101	16	18969	[redacted]	[redacted]	real	decoy
1176	S116	1	3	PI	16669	9	227	16677	[redacted]	[redacted]	real	real
1177	S116	1	3	PI	9310	4179	28	13488	[redacted]	[redacted]	real	decoy
1178	S116	1	3	PI	18968	3	7	18970	[redacted]	[redacted]	real	decoy
1179	S116	1	3	PI	13451	136	14	13586	[redacted]	[redacted]	real	decoy
1180	S116	1	3	PI	13506	111	14	13616	[redacted]	[redacted]	real	decoy

Figure 17. Extracted Subject Activity by System

2. Subject Activity Per Minute

In addition to extracting the temporal variables, subject network activity by number of packets sent for each 60 second intervals for the entire duration of each experiment trial were extracted using a combination of Python and Cypher in a Jupyter Notebook. These queries of subject activity in were used to export number of packets sent during each time interval along with source IPs used by each subject and destination IP addresses. The

Cypher query is displayed in Cypher Syntax 2 with the spreadsheet output of sent packets depicted in Figure 18. Day 2 subject activity can be extracted by the s.Day value with ’2’ to capture “Second” values during the second day of experimentation.

Cypher Syntax 2: Day 1 Subject Activity per Minute Extraction

```
MATCH (n:NetFlow)-[p:PACKET_SENT_AT]->(s:Second) WHERE
n.DestIP_Validity = 'real' AND s.Day = '1' AND ' +  str(int(time))+ '<=
s.AbsoluteTime <= ' + str(int(time) +59) + ' AND n.SourceIP = <Source IP>
RETURN COUNT (p) AS packets, toFloat(COUNT(p))/60 AS pps,
size(COLLECT(DISTINCT(n.SourceIP))) AS source_length,
size(COLLECT(DISTINCT(n.DestIP))) AS dest_length,
```

	A	B	C	D	E	F	G	
1	Start	End	Packets	PPS - Real	SourceNum	DestNum	SourceIP	DestIP
47	2700	2759	0	0	0	0	[]	[]
48	2760	2819	1329	22.15	1	1	[redacted]	
49	2820	2879	204	3.4	1	1	[redacted]	
50	2880	2939	75	1.25	1	1	[redacted]	
51	2940	2999	75	1.25	1	1	[redacted]	
52	3000	3059	75	1.25	1	1	[redacted]	
53	3060	3119	4283	71.38333	1	2	[redacted]	
54	3120	3179	430	7.166667	1	2	[redacted]	
55	3180	3239	435	7.25	1	2	[redacted]	
56	3240	3299	435	7.25	1	2	[redacted]	
57	3300	3359	404	6.733333	1	2	[redacted]	
58	3360	3419	474	7.9	1	3	[redacted]	
59	3420	3479	4576	76.26667	1	3	[redacted]	
60	3480	3539	828	13.8	1	3	[redacted]	
61	3540	3599	795	13.25	1	3	[redacted]	
62	3600	3659	810	13.5	1	4	[redacted]	
63	3660	3719	5128	85.46667	1	4	[redacted]	
64	3720	3779	1356	22.6	1	4	[redacted]	
65	3780	3839	1487	24.78333	1	4	[redacted]	
66	3840	3899	1219	20.31667	1	4	[redacted]	
67	3900	3959	2387	39.78333	1	5	[redacted]	
68	3960	4019	4515	75.25	1	4	[redacted]	
69	4020	4079	1448	24.13333	1	4	[redacted]	
70	4080	4139	1326	22.1	1	4	[redacted]	
71	4140	4199	1440	24	1	4	[redacted]	
72	4200	4259	1434	23.9	1	4	[redacted]	
73	4260	4319	1326	22.1	1	4	[redacted]	
74	4320	4379	1440	24	1	4	[redacted]	

S06-Day1_real_packets

Figure 18. Subject Network Activity Per Minute

3. Enrichment of Extracted Subject Activity

From the subject activity data previously extracted, formulas were used within Microsoft Excel to enrich the data, calculating the natural log of packets sent, along with reporting system interaction for each 60 second interval for calculation of system

discovered percentages. Separate spreadsheets were generated for real and decoy systems as depicted in Figure 19 and Figure 20. For PI and PU groups, formulas were used to calculate real to decoy packet ratios for each time interval.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Start	End	S06 - Real I	PPS - Real	SourceNum	DestNum	SourceIP	DestIP	Real - In Packets	% of Real Systems Discovered			
2	0	59	0	0	0	0	[]	[]	0	0%	n	n	n
3	60	119	0	0	0	0	[]	[]	0	0%	n	n	n
4	120	179	0	0	0	0	[]	[]	0	0%	n	n	n
5	180	239	0	0	0	0	[]	[]	0	0%	n	n	n
6	240	299	0	0	0	0	[]	[]	0	0%	n	n	n
7	300	359	0	0	0	0	[]	[]	0	0%	n	n	n
8	360	419	0	0	0	0	[]	[]	0	0%	n	n	n
9	420	479	110	1.833333	1	1			4.700480366	2%	n	n	n
10	480	539	61361	1022.683	1	52			11.07452973	100%	Y	Y	Y
11	540	599	16712	278.5333	1	30			9.723882303	100%	Y	Y	Y
12	600	659	1515	25.25	1	24			7.323170718	100%	n	n	n
13	660	719	2940	49	1	52			7.98616486	100%	Y	Y	Y
14	720	779	91484	1524.733	1	45			11.42391937	100%	Y	Y	Y
15	780	839	131141	2185.683	1	31			11.78402836	100%	n	Y	Y
16	840	899	443	7.383333	1	22			6.09356977	100%	n	Y	Y
17	900	959	617	10.28333	1	26			6.424869024	100%	n	Y	Y
18	960	1019	120	2	1	9			4.787491743	100%	n	n	n
19	1020	1079	0	0	0	0			0	100%	n	n	n
20	1080	1139	0	0	0	0			0	100%	n	n	n
21	1140	1199	0	0	0	0			0	100%	n	n	n
22	1200	1259	4	0.066667	1	1			1.386294361	100%	n	n	n
23	1260	1319	6	0.1	1	1			1.791759469	100%	n	n	n
24	1320	1379	6	0.1	1	1			1.791759469	100%	n	n	n
25	1380	1439	0	0	0	0			0	100%	n	n	n
26	1440	1499	0	0	0	0			0	100%	n	n	n
27	1500	1559	0	0	0	0			0	100%	n	n	n
28	1560	1619	0	0	0	0			0	100%	n	n	n
29	1620	1679	342	5.7	1	1			5.834810737	100%	n	n	n

Figure 19. Real System Subject Activity Enrichment

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Start	End	Decoy - Packets	PPS - Decoy	SourceNum	DestNum	SourceIP	DestIP	Decoy - In Packets	DecoyInter	RealInteraction	Real Packets	Ratio				
2	0	59	0	0	0	0	[]	[]	0	0%	0%	0	0		n	n	n
3	60	119	0	0	0	0	[]	[]	0	0%	0%	0	0		n	n	n
4	120	179	0	0	0	0	[]	[]	0	0%	0%	0	0		n	n	n
5	180	239	0	0	0	0	[]	[]	0	0%	0%	0	0		n	n	n
6	240	299	0	0	0	0	[]	[]	0	0%	0%	0	0		n	n	n
7	300	359	0	0	0	0	[]	[]	0	0%	0%	0	0		n	n	n
8	360	419	0	0	0	0	[]	[]	0	0%	0%	0	0		n	n	n
9	420	479	0	0	0	0	[]	[]	0	0%	0%	0	0		n	n	n
10	480	539	0	0	0	0	[]	[]	0	0%	0%	0	0		n	n	n
11	540	599	0	0	0	0	[]	[]	0	0%	2%	51	1		n	n	n
12	600	659	0	0	0	0	[]	[]	0	0%	2%	141	1		n	n	n
13	660	719	0	0	0	0	[]	[]	0	0%	2%	117	1		n	n	n
14	720	779	123529	2058.816667	1	50			11.72423123	100%	100%	44653	-0.46899		Y	Y	Y
15	780	839	1714	28.56666667	1	3			7.446585099	100%	100%	5812	0.544512		n	n	n
16	840	899	30	0.5	1	2			3.401197382	100%	100%	393	0.858156		n	Y	n
17	900	959	15	0.25	1	1			2.708050201	100%	100%	390	0.925926		n	n	n
18	960	1019	112	1.866666667	1	4			4.718498871	100%	100%	303	0.460241		Y	n	n
19	1020	1079	114	1.9	1	4			4.736198448	100%	100%	289	0.434243		n	n	n
20	1080	1139	53	0.883333333	1	3			3.970291914	100%	100%	227	0.621429		n	n	n
21	1140	1199	3298	54.96666667	1	50			8.101071503	100%	100%	2598	-0.11872		Y	Y	Y
22	1200	1259	692	11.53333333	1	48			6.539585956	100%	100%	736	0.030812		Y	Y	Y
23	1260	1319	887	14.78333333	1	50			6.787844982	100%	100%	1462	0.244785		Y	Y	Y
24	1320	1379	600	10	1	7			6.396929655	100%	100%	1439	0.411476		n	n	n
25	1380	1439	947	15.78333333	1	11			6.853299093	100%	100%	1297	0.155971		n	n	n
26	1440	1499	938	15.63333333	1	12			6.843749949	100%	100%	1164	0.107517		n	n	n
27	1500	1559	821	13.68333333	1	13			6.710523109	100%	100%	1584	0.317256		Y	Y	Y
28	1560	1619	1136	18.93333333	1	10			7.035268599	100%	100%	1336	0.080906		Y	Y	Y
29	1620	1679	1981	33.01666667	1	14			7.591357047	100%	100%	1477	-0.14575		Y	Y	n

Figure 20. Decoy System Subject Activity Enrichment

D. TECHNICAL ARTIFACTS

During transformation, ingest and extraction of subject activity during the Tularosa study, numerous technical artifacts were created including:

- 3,692 files in CSV format (.csv) were generated during conversion of PCAP data collected for all 139 subjects during the Tularosa study
- 123 files in Neo4j database dump format (.dump) were generated by importing the CSV files containing the PCAP data into Neo4j. These files contain complete self-contained archives of reconstructed Day 1 and Day 2 subject activity including PCAP and other data imported along with all relationships created using the previously described Jupyter Notebooks. Neo4j database dump files were not created for the 16 subjects previously identified as having been impacted by data collection issues or determined to be unqualified subjects.
- Several hundred spreadsheets containing information on subject activity were exported from Neo4j using relationships and properties created through the import and processing of data within Neo4j.
    - 246 spreadsheets (one per subject per day) containing latency, frequency, and duration information for system interaction for each subject were generated using queries of the data within Neo4j to aggregate timing and number of packets, subject interaction with each real and decoy system on the network was returned with values of latency (time of first interaction with system), frequency (number of packets sent to system) and last (time of last interaction with system). Duration was calculated by subtracting the last value from the latency value and adding one second to account for the interaction time occurring for the duration of the last value.
    - 246 spreadsheets (one per subject per day) containing subject activity for interaction with real systems on the cyber range environment were generated using queries of the data within Neo4j. Information extracted for each 60 second time interval of each experiment trial included the start and stop time for each time interval, number of packets sent to real systems, source IP addresses used by the participant for the given time interval and real system destination IP addresses interacted during the given time interval.
    - 92 spreadsheets (59 Day 1 subjects, 33 Day 2 subjects) containing subject activity for interaction with decoy systems on the cyber range environment generated using queries of the data within Neo4j. Information extracted for each 60 second time interval of each experiment trial included the start and stop time for each time interval, number of packets sent to decoy systems, source IP addresses used by the participant for the given time interval and decoy system destination IP addresses interacted during the given time interval.

V. EXPLORATORY DATA ANALYSIS

As highlighted by Tukey, "Exploratory data analysis is detective work," emphasizing the need to use multiple methods to find clues within the data to generate new insights or formulate new hypotheses (Tukey, 1977, p. 1). This sentiment is echoed by the *NIST/SEMATECH e-Handbook of Statistical Methods* which highlights that for EDA "focus is on the data--its structure, outliers, and models suggested by the data" (NIST, 2012, 1.1.2.2). Use of multiple methods to analyze the Tularosa study data seeks to minimize the impact of flaws and biases inherent in any individual analytic technique acknowledged by the post-positivist epistemic perspective applied in this research.

A. ANALYTIC FRAMEWORK

For each of the criteria grouping evaluated a series of statistical techniques were applied to describe and evaluate the temporal variables of latency, frequency, and duration across each of the experimental groups. Within these temporal variables latency is interval data, with equally split numerical values and a zero value that indicates activity occurs at the start of the experiment trial. Frequency and duration are both ratio data also having equally split numerical values, but with zero values indicating these variables do not exist.

First descriptive statistics for each variable were calculated including the minimum and maximum value, median, mean, standard deviation (sample), variance (sample), skewness, kurtosis, and coefficient of variability. The mean is the average of values within a number set. The median is the middle value within a number set. The median is less impacted by outliers, which makes it well suited for analysis of network traffic which can contain values that vary between subjects by several magnitudes of order. Variance is the squared differences from the mean, which are then averaged. Standard deviation captures the average difference of each data point around the mean using the square root of the variance. It provides a useful measure of the dispersion of values within a dataset.

The Shapiro-Wilk-Royston Normality Test was used to test for a normal distribution of each variable using the Shapiro-Wilk normality test using the Royston algorithm (Mun, 2017). Levene's test was used to test if sample groupings share

homogeneity of variance (NIST,2012). Levene's test was used instead of Bartlett's Test for Homogeneity of Variances as the variables analyzed largely have non-normal distributions. (NIST, 2012). Both tests were applied to the data as many statistical tests assume data come from a normal distribution and that samples have equal variance. Violations of these assumptions result in unreliable analysis. In-line with previous analysis of the Tularosa study data, the Kruskal-Wallis test was applied to compare multiple samples to determine if population medians for these groups were statistically equal (Mun, 2017). The Kruskal-Wallis test was used as the data evaluated was usually non-normally distributed with unequal variances. Also, in-line with previous analysis of the Tularosa study data, for variables where population medians were not statistically equal, Dunn's Post Hoc Test with Benjamanini-Hochberg Correction was used to determine which group pairs were statistically different. Dunn's Post Hoc Test uses rank sums to calculate differences between the group pairs. Statistical calculations were cross validated using a minimum of two different applications in their computation. Applications used for these calculations include Real Options Value (ROV) BizStats, JMP® Pro, Microsoft Excel, multiple Python libraries including *dataprep, scipy,* and *statsmodels*, and the R library *rstatix*.

Graphical representation of the Tularosa study data and concepts derived from subsequent analysis was guided by Tufte's work on the visualization of quantitative data which emphasized "Graphical excellence consists of complex ideas communicated with clarity, precision and efficiency" (Tufte, 2001, p. 51). Applications used to generate data visualization graphics include Tableau and multiple Python libraries including *matplotlib* and *seaborn*. Histograms provide the frequency of values for the temporal variables for visual comparison across groups. Box-and-whisker plots provide a visual display of the distribution of data, with the middle of the box indicating the median and the bottom and top of the box indicating the middle 50% of the observations (Fields, 2009). Whiskers emanating from the box indicate the top and bottom 25% of the observations with any outliers located outside of the whiskers (Fields, 2009).

For each subject the number of packets sent in each 60 second interval from the start of the experiment (t = 0 seconds) to the end of the experiment (t = 23,400 seconds) to

real systems (all groups) and to decoy systems (PI and PU groups) were extracted and aggregated to generate timelines of subject activity for each group. Charts were created for average and median packets per minute along with standard deviation and variance for packets per minute for each group. An 8th degree polynomial fit trend line was added for comparison of subject activity across groups. The number of packets sent was also used to generate charts with the probability of subject activity at different packet thresholds (greater than 2, 7, 20, 54, 148 and 403). These packet thresholds are based off natural logarithm values ($\ln 2 \approx 1$, $\ln 7 \approx 2$, etc.). An 8th degree polynomial fit trend line was also added for comparison of subject activity across groups.

Timelines of individual subject activity were also generated to highlight the high degree of variability of subject activity over time and between subjects. These charts are available as a supplemental material accessible through the Dudley Knox Library at the Naval Postgraduate School. For subjects where cyber deception was present (Present-Informed, Present-Uninformed) these charts include packets sent per minute to real and decoy systems (actual packets per minute and natural log of packets per minute), the ratio of packets sent to real and decoy systems over time, timeline of interaction percentages with real and decoy systems, number of real and decoy system interactions per minute over time, running average of real to decoy packet ratio and a running sum of packets sent to real and decoy systems. For subjects where cyber deception was not present (Absent-Informed, Absent-Uninformed), these charts only include packets sent per minute to real systems, number of real and decoy systems interactions per minute, and running sum of packets sent to real systems.

B. EXPERIMENT SUBJECT SELECTION

The network-graph based approach used to import and aggregate subject activity captured during the Tularosa study provided unparalleled insights into individual subjects through second-by-second accounting of network activity generated during the study. While the Tularosa study captured cyber-attack behavior from a total of 139 subjects, only data from 120 subjects on Day 1 and 111 subjects on Day 2 was used for analysis. The number of subjects analyzed in this study differs from previous research due to new

insights available from the granular analysis enabled by the methodology created for the analysis of subject network activity.

For Day 1 data analysis, previously identified data collection issues excluded data from 15 subjects (S01-S05, S33-S42) and identification of an unqualified subject excluded data from one additional subject (S69). Following extraction of the temporal variables from the network traffic, data from three additional subjects was excluded. Two of these subjects (S15, S110) interacted only with a single system on the network, which indicates they were either unable or unwilling to execute the network reconnaissance task outlined in the task briefing provided for the experiment. The third subject excluded (S64) started interaction with systems on the cyber range network nearly 5 minutes prior to their designated start time, as evidenced by packets sent to multiple systems on the network by the subject's system starting 248 seconds prior to the beginning of the experiment trial.

For Day 2 data analyzed there was an overall reduction of 9 subjects. Two subjects (S70 and S80) did not return for Day 2 of the experiment, with data collection issues impacting eight additional subjects (S18, S24, S25, S26, S28, S49, S55 and S88). Two subjects (S15 and S110) were excluded from Day 2 analysis for the same reason they were excluded from Day 1 analysis, interacting with only a single system on the network. Day 2 data from one subject (S64) excluded from Day 1 analysis due to starting activity prior to the start of the experiment was included in the Day 2 analysis as all Day 2 system interactions occurred after the designated start time. After the completion of data analysis for this dissertation, a data collection issue was identified for one subject (S132) included in the Day 2 analysis, which indicated only partial collection of the overall activity for the subject. While inclusion of this subject's data did not impact the knowledge claims made in this dissertation, this issue should be noted for future research on the Tularosa study data.

Analysis of the network traffic generated by subjects during the Tularosa study also revealed the use of multiple source IP addresses by several subjects to send packets to both real and decoy systems. Initially the computer system used by the subjects to perform reconnaissance and exploitation tasks was assigned a .3 address on the cyber range network, but subjects were able to change this address to one or more unassigned IP

addresses on the cyber range. Use of multiple source IP addresses by a cyber-attacker would serve to obfuscate their activity, potentially decreasing the probability of detection by a cyber defender. For Day 1, ten subjects (S13, S27, S60, S80, S91, S93, S95, S96, S133 and S135) were observed using multiple source IP addresses to send packets to target systems on the cyber range network. For Day 2, five subjects (S57, S58, S91, S96 and S116) were observed using multiple source IP addresses to send packets to target systems on the cyber range network. Only two subjects (S91 and S96) used multiple source IP addresses on both Day 1 and Day 2. For Day 1 experimentation, one subject in the AI condition, four subjects in the AU condition and five subjects in the PI condition used multiple source IP addresses. However, in Day 2 experimentation all five subjects using multiple source IP addresses were in the PIAU condition.

C. DAY 1 EXPERIMENT DATA ANALYSIS

This section provides the statistical and graphical analysis for all Tularosa study data collected during Day 1 experiment trials.

1. Real System Interactions For All Groups (AI, AU, PI, PU)

This sub-section provides the statistical and graphical analysis for real system interactions for all experiment groups during Day 1 experiment trials.

a. Descriptive and Comparative Statistics for All Real System Interactions

This sub-section provides the descriptive and comparative statistics for real system interactions for all experiment groups during Day 1 experiment trials.

(1) Latency

Table 1 provides the descriptive statistics for the latency variable for real system interactions for all experiment groups during Day 1 experiment trials. Figure 21 depicts a histogram of latency values for these groups. Figure 22 depicts a box-and-whisker plot of latency values for these groups. Table 2 provides the results of the Shapiro-Wilk-Royston test for normality for these groups. Table 3 provides the results of Levene's test for equal variance for these groups. Table 4 provides the results of the Kruskal-Wallis test for these

groups. Finally, Table 5 provides the results of the Dunn's Post Hoc test with Benjamini-Hochberg correction for these groups.

Table 1. Day 1 Real System Latency Descriptive Statistics

Value	**AI**	**AU**	**PI**	**PU**
Subjects (N)	*27*	**34**	30	29
Count of Real Systems	50	50	50	50
Count of Observed Real System Interactions (n)	1306	**1629**	*1336*	1393
Count of Possible Real System Interactions	*1350*	**1700**	1500	1450
Percentage of Real Systems Interactions	**96.74%**	95.82%	*89.07%*	96.07%
Minimum Value	*1*	*1*	**151**	8
Maximum Value	*17,484*	19,569	**20,155**	18,888
Median	1127	*992*	**1369**	1068
Mean	*2208*	2580	**3059**	2935
Standard Deviation (Sample)	*3007*	3465	4187	**4225**
Variance (Sample)	9.04×10^{06}	1.20×10^{07}	1.75×10^{07}	$\mathbf{1.79\times10^{07}}$
Skewness	2.35	**2.51**	2.41	*2.23*
Kurtosis	4.82	**6.76**	5.44	*3.79*
Coefficient of Variability	1.36	*1.34*	1.37	**1.44**

Highest values in **Bold**, Lowest values in *Italics*

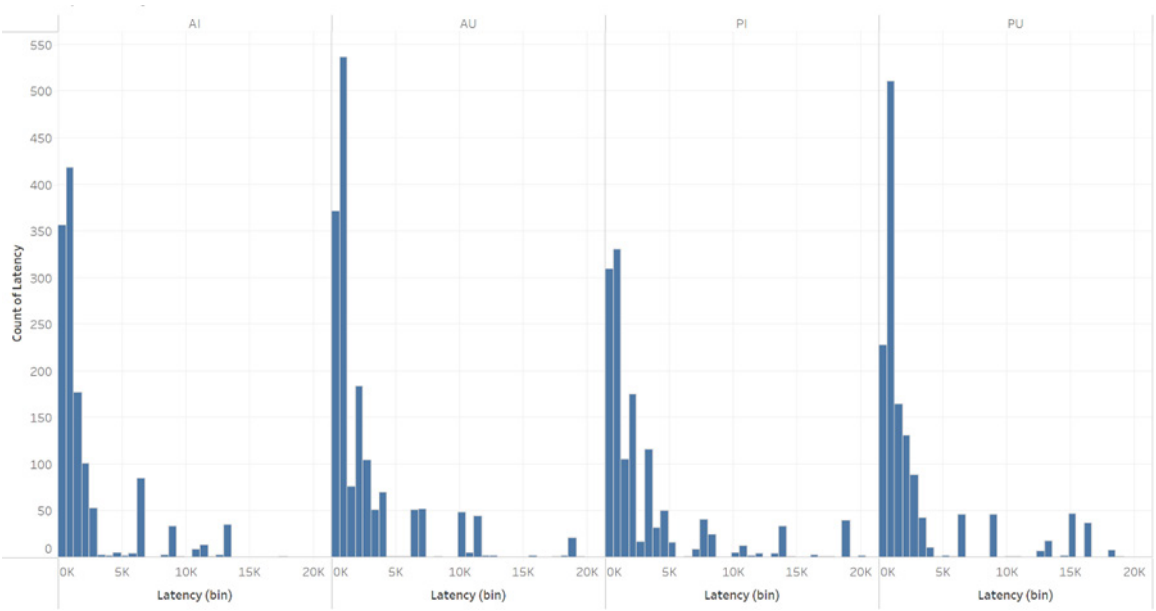


Figure 21. Day 1 Real System Latency Histogram

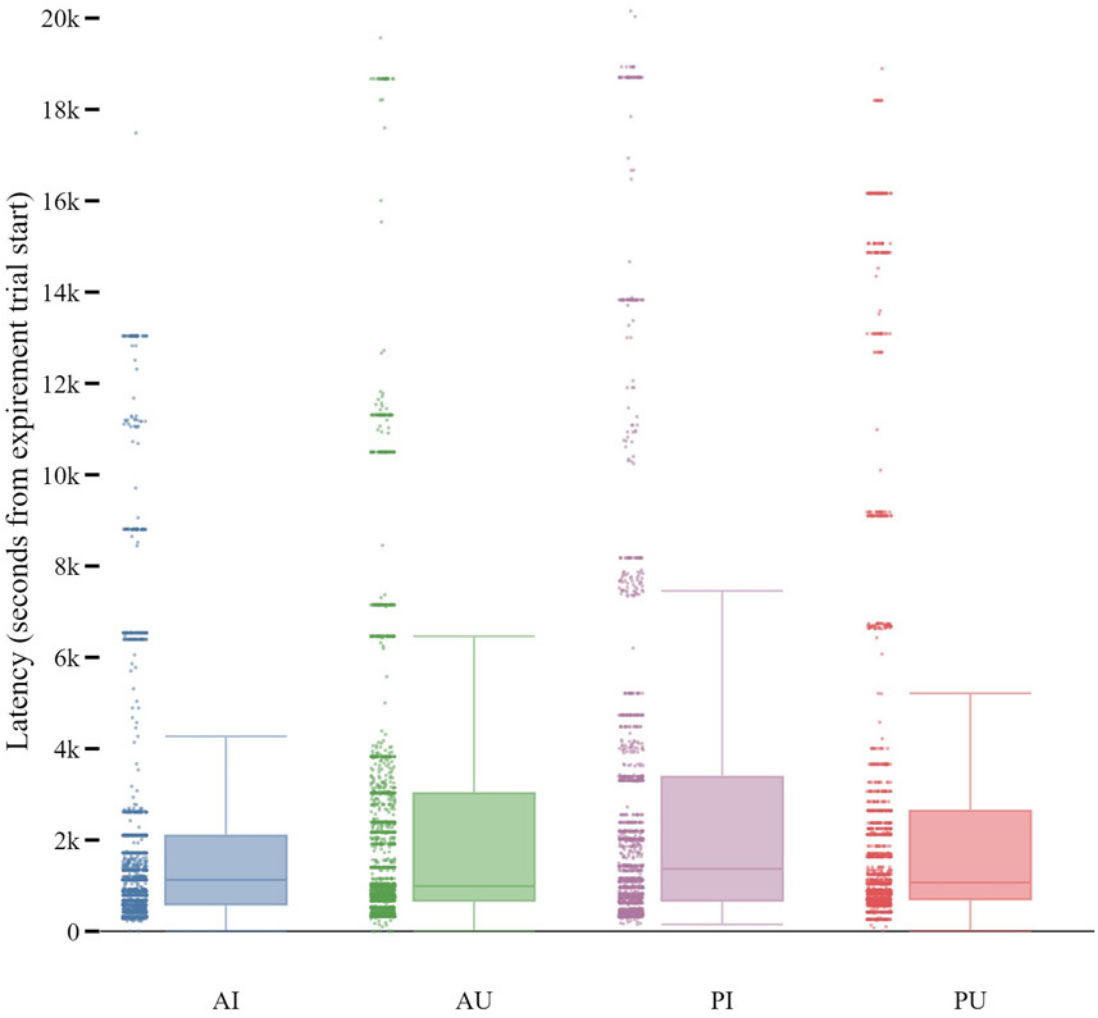


Figure 22. Day 1 Real Systems Latency Box-and-Whisker Plot

Table 2. Day 1 Real Systems Latency Shapiro-Wilk-Royston Test for Normality

	AI	AU	PI	PU
W	0.617	0.649	0.645	0.604
p-value	0.00	0.00	0.00	0.00

For all conditions the null hypothesis that the latency data is normal is rejected with a $p < 0.05$.

Table 3. Day 1 Real Systems Latency Levene's Test for Equal Variance

	Centered on Median	Centered on Mean
W	12.4	28.7
p-value	0.00	0.00

For both Median and Mean $p < 0.05$. Null Hypothesis that the data has equal variance is rejected.

Table 4. Day 1 Real Systems Latency Kruskal-Wallis Test

	Real System Interactions
Chi-squared	43.3
p-value	0.000

The population medians are not statistically equal at 1%, 5%, or 10% significance.

Table 5. Day 1 Real Systems Latency Dunn's Post Hoc Test with Benjamini-Hochberg Correction

	AI	AU	PI	PU
AI	1.00			
AU	**0.0022**	1.00		
PI	**0.0000**	**0.0017**	1.00	
PU	**0.0000**	0.0641	0.1701	1.00

Pairwise comparisons with Dunn's test indicate latency values for the AI group are significantly different from the AU group (p=0.0022), PI group (p=0.0000) and PU group (p=0.0000) and for the AU group and PI group (p=0.0017)

(2) Frequency

Table 6 provides the descriptive statistics for the frequency variable for real system interactions for all experiment groups during Day 1 experiment trials. Figure 23 depicts a histogram of frequency values for these groups. Figure 24 depicts a box-and-whisker plot of frequency values for these groups. Table 7 provides the results of the Shapiro-Wilk-Royston test for normality for these groups. Table 8 provides the results of Levene's test

for equal variance for these groups. Table 9 provides the results of the Kruskal-Wallis test for these groups. Finally, Table 10 provides the results of the Dunn's Post Hoc test with Benjamini-Hochberg correction for these groups.

Table 6. Day 1 Real System Frequency Descriptive Statistics

Value	**AI**	**AU**	**PI**	**PU**
Subjects	*27*	**34**	30	29
Observations	1306	1629	1336	1393
Minimum Value	8	4	**13**	*1*
Maximum Value	5,376,200	9,620,700	**35,699,000**	*4,794,500*
Median	6646	*4188*	**16,086**	4252
Mean	75,092	*64,448*	**91,204**	73,946
Standard Deviation (Sample)	287,160	336,654	**979,876**	*224,294*
Variance (Sample)	8.25×10^{10}	1.13×10^{11}	$\mathbf{9.60\times10^{11}}$	5.03×10^{10}
Skewness	13.4	23.7	**36.0**	*12.17*
Kurtosis	212.3	628.7	**1309**	*206.9*
Coefficient of Variability	3.82	5.22	**10.74**	*3.03*

Highest values in **Bold**, Lowest values in *Italics*

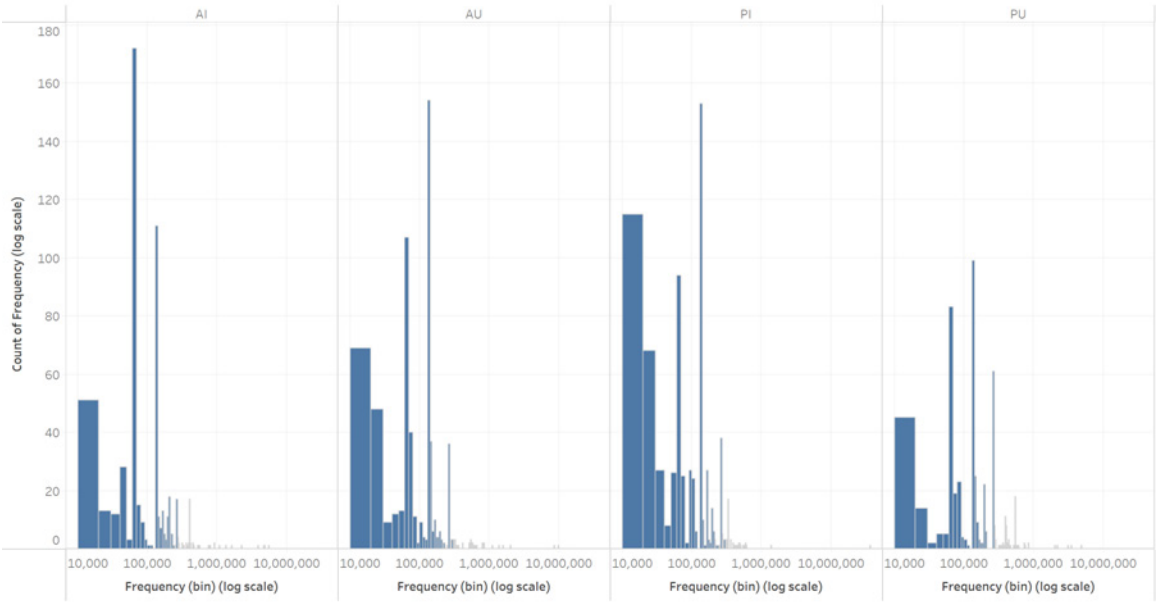


Figure 23. Day 1 Real System Frequency Histogram.

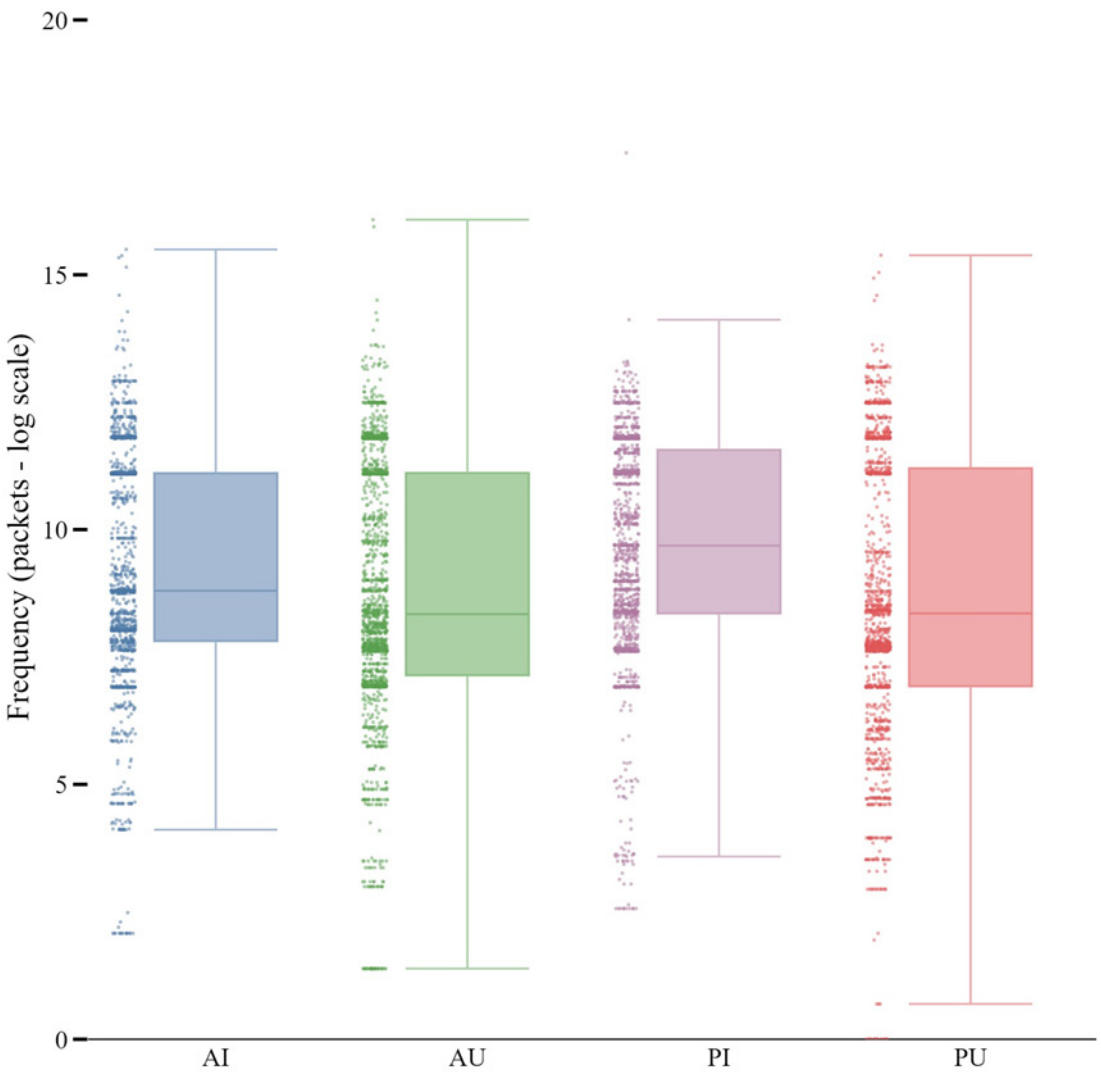


Figure 24. Day 1 Real System Frequency Box-and-Whisker Plot

Table 7. Day 1 Real System Frequency Shapiro-Wilk-Royston Test for Normality

	AI	AU	PI	PU
W	0.202	0.119	0.031	0.296
p-value	0.00	0.00	0.00	0.00

For all conditions $p<0.05$ therefore the null hypothesis that the frequency data is normal is rejected.

Table 8. Day 1 Real System Frequency Levene's Test for Equal Variance

	Centered on Median	Centered on Mean
W	0.473	0.528
p-value	0.701	0.663

For both Median and Mean $p < 0.05$. Null Hypothesis that the data has equal variance is accepted. This is the only temporal variable that shows equal variance across all groups.

Table 9. Day 1 Real System Frequency Kruskal-Wallis Test

	Real System Interactions
Chi-squared	150.22
p-value	0.000

The population medians are not statistically equal at 1%, 5%, or 10% significance.

Table 10. Day 1 Real System Frequency Dunn's Post Hoc Test with Benjamini-Hochberg Correction

	AI	AU	PI	PU
AI	1.00			
AU	**0.0000**	1.00		
PI	**0.0000**	**0.0000**	1.00	
PU	**0.0000**	0.4903	**0.000**	1.00

Pairwise comparisons with Dunn's test indicate frequency values for the AI group are significantly different from the AU group ($p = 0.0000$), PI group ($p=0.0000$) and PU group ($p=0.0000$), for the AU group and PI group ($p=0.0000$) and PU group and PI group ($p=0.0000$).

(3) Duration

Table 11 provides the descriptive statistics for the duration variable for real system interactions for all experiment groups during Day 1 experiment trials. Figure 25 depicts a histogram of duration values for these groups. Figure 26 depicts a box-and-whisker plot of duration values for these groups. Table 12 provides the results of the Shapiro-Wilk-

Royston test for normality for these groups. Table 13 provides the results of Levene's test for equal variance for these groups. Table 14 provides the results of the Kruskal-Wallis test for these groups. Finally, Table 15 provides the results of the Dunn's Post Hoc test with Benjamini-Hochberg correction for these groups.

Table 11. Day 1 Real System Duration Descriptive Statistics

Value	**AI**	**AU**	**PI**	**PU**
Subjects	*27*	**34**	30	29
Observations	1306	1629	1336	1393
Minimum Value	**2**	*1*	*1*	*1*
Maximum Value	23,395	**23,400**	23,366	*23,200*
Median	**15,951**	15,358	*14,075*	15,078
Mean	13,469	**14,002**	13,098	*12,521*
Standard Deviation (Sample)	7,247	*7,049*	7,469	**8,229**
Variance (Sample)	5.25×10^{7}	4.97×10^{7}	5.58×10^{7}	$\mathbf{6.77\times10^{7}}$
Skewness	*-0.620*	-0.497	-0.401	**-0.283**
Kurtosis	-1.03	-1.05	-1.21	*-1.51*
Coefficient of Variability	0.538	*0.503*	0.570	**0.652**

Highest values in **Bold**, Lowest values in *Italics*

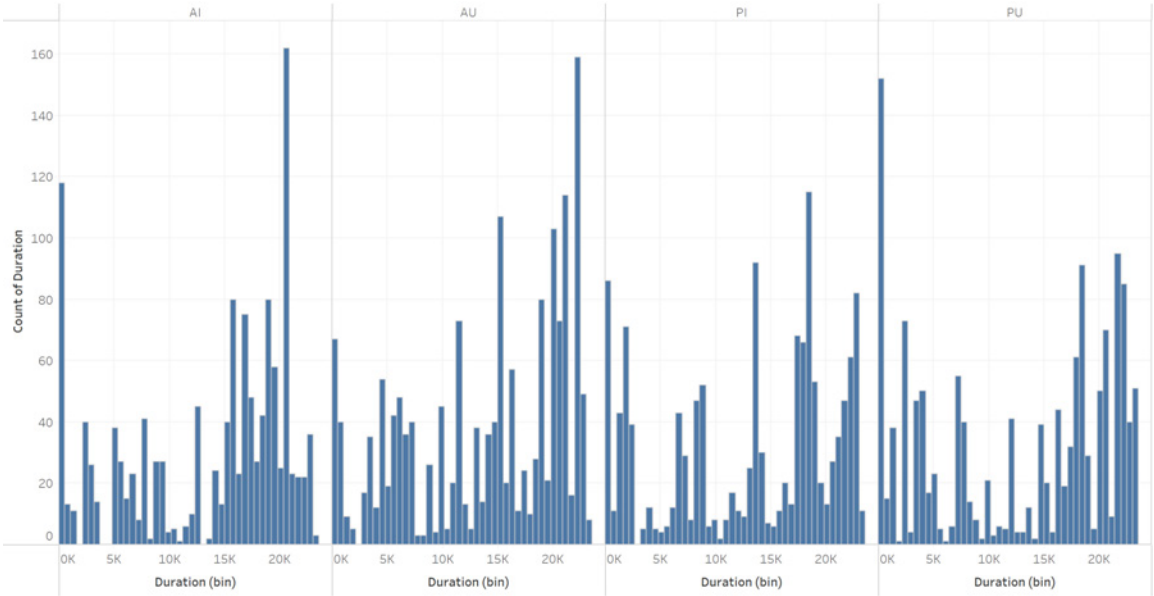


Figure 25. Day 1 Real System Duration Histogram.

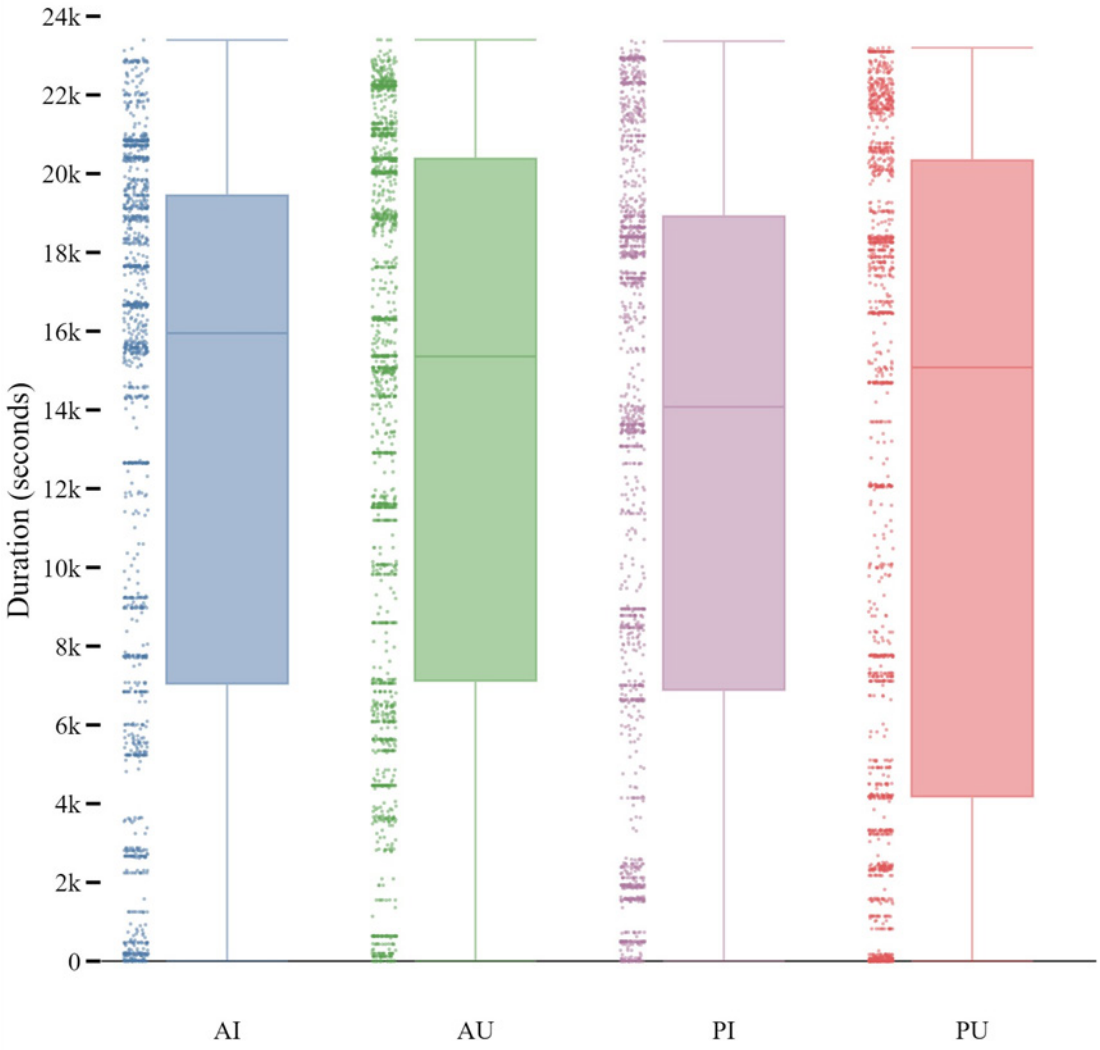


Figure 26. Day 1 Real System Duration Box-and-Whisker Plot.

Table 12. Day 1 Real System Duration Shapiro-Wilk-Royston Test for Normality

	AI	AU	PI	PU
W	0.882	0.909	0.907	0.876
p-value	0.00	0.00	0.00	0.00

For all conditions $p<0.05$ therefore the null hypothesis that the duration data is normal is rejected.

Table 13. Day 1 Real System Duration Levene's Test for Equal Variance

	Centered on Median	Centered on Mean
W	34.67	47.70
p-value	0.00	0.00

For both Median and Mean $p < 0.05$. Null Hypothesis that the data has equal variance is rejected.

Table 14. Day 1 Real System Duration Kruskal-Wallis Test

	Real System Interactions
Chi-squared	17.91
p-value	0.005

The population medians are not statistically equal at 1%, 5%, or 10% significance.

Table 15. Day 1 Real System Duration Dunn's Post Hoc Test with Benjamini-Hochberg Correction

	AI	AU	PI	PU
AI	1.00			
AU	**0.0255**	1.00		
PI	0.5181	**0.0043**	1.00	
PU	0.2926	**0.0006**	0.5181	1.00

Pairwise comparisons with Dunn's test indicate duration values for the AI group are significantly different from the AU group ($p = 0.0255$), for the AU group and PI group ($p=0.0043$) and for the PU group and AU group ($p=0.0006$)

b. Subject Activity Analysis for All Real System Interactions

This sub-section provides the graphical analysis for real system interactions for all experiment groups during Day 1 experiment trials. Figure 27 depicts the natural log of the average packets per minute sent by each group from the start to the end of the Day 1 experiment trials. Figure 28 depicts the natural log of the median packets per minute sent by each group from the start to the end of the Day 1 experiment trials. Figure 29 depicts

the standard deviation of the natural log of the average packets per minute sent by each group from the start to the end of the Day 1 experiment trials. Figure 30 depicts the variance of the natural log of the average packets per minute sent by each group from the start to the end of the Day 1 experiment trials.

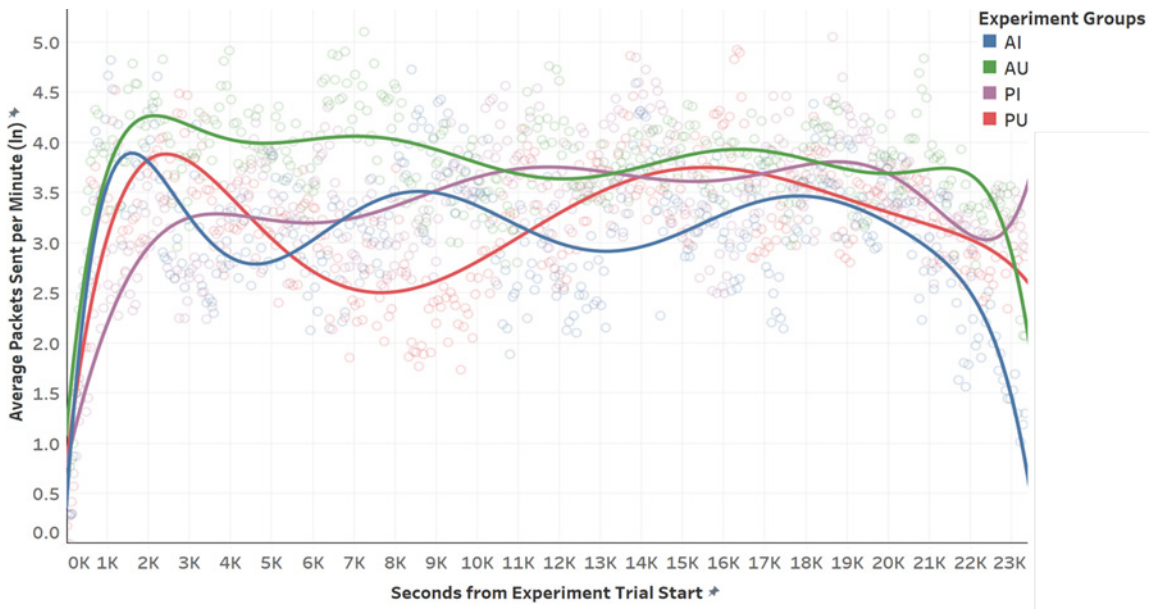


Figure 27. Day 1 Real System Average Packets Per Minute.

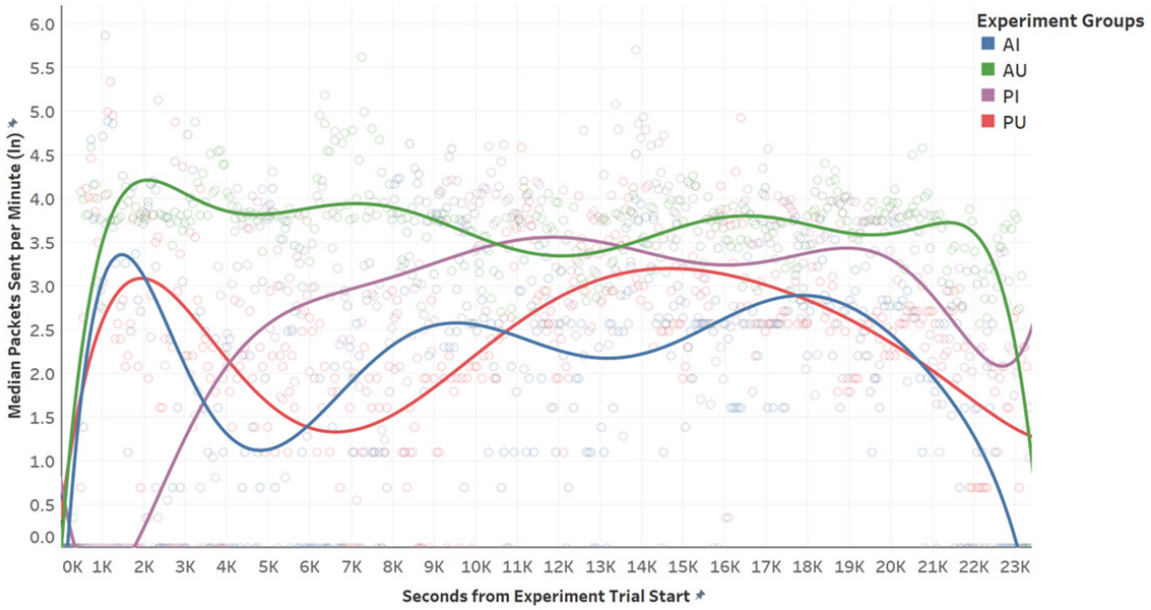


Figure 28. Day 1 Real System Median Packets Per Minute.

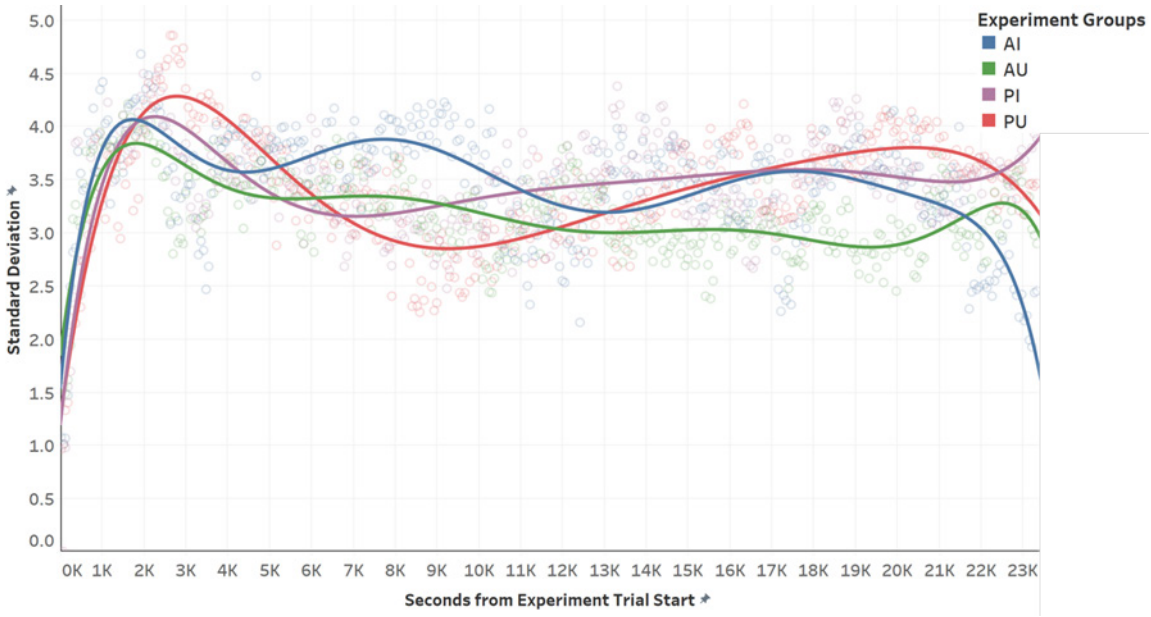


Figure 29. Day 1 Real System Standard Deviation Per Minute

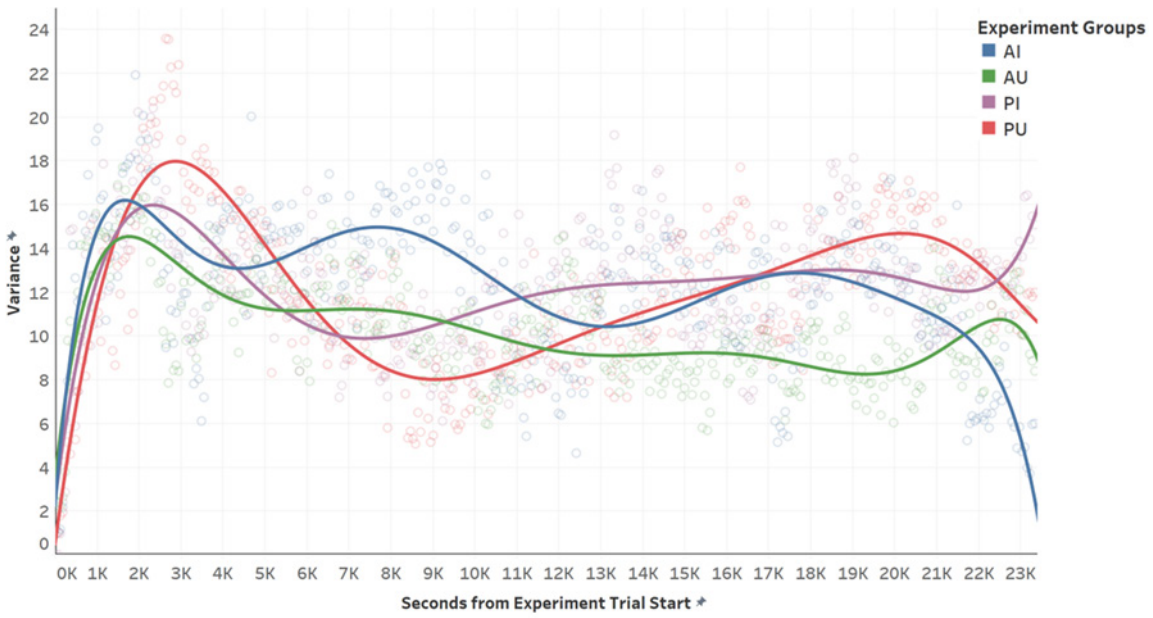


Figure 30. Day 1 Real System Variance Per Minute.

Figure 31 depicts subjects sending more than 2 packets for each one-minute interval by each group from the start to the end of the Day 1 experiment trials. Figure 32 depicts subjects sending more than 7 packets for each one-minute interval by each group from the start to the end of the Day 1 experiment trials. Figure 33 depicts subjects sending more than 33 packets for each one-minute interval by each group from the start to the end of the Day 1 experiment trials. Figure 34 depicts subjects sending more than 54 packets for each one-minute interval by each group from the start to the end of the Day 1 experiment trials. Figure 35 depicts subjects sending more than 148 packets for each one-minute interval by each group from the start to the end of the Day 1 experiment trials. Figure 36 depicts subjects sending more than 403 packets for each one-minute interval by each group from the start to the end of the Day 1 experiment trials.

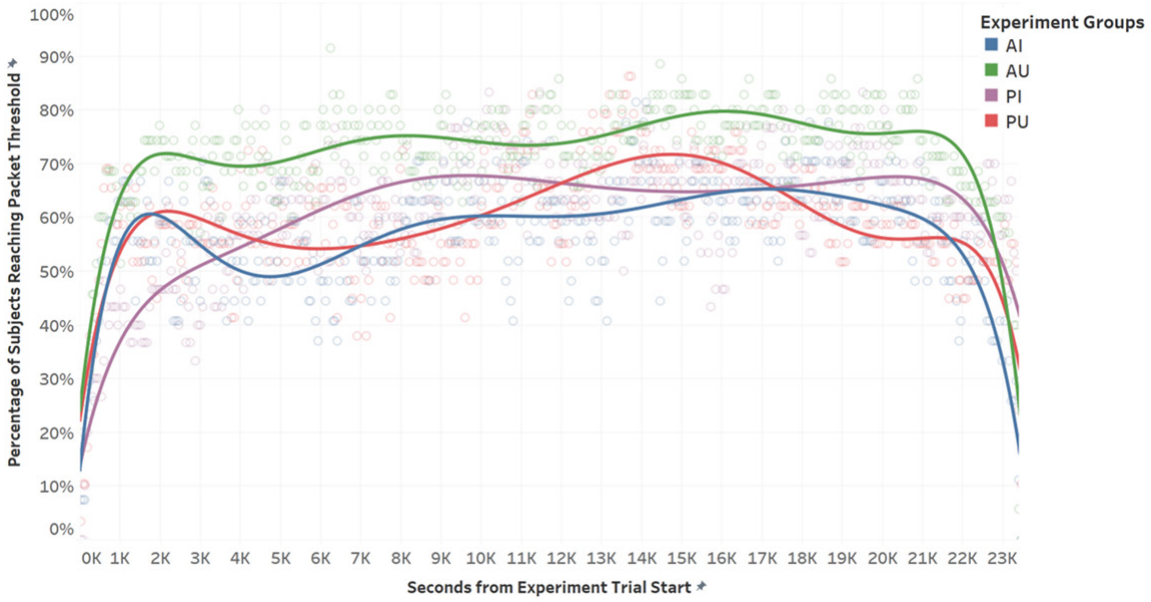


Figure 31. Day 1 Real System Subject Activity Threshold – 2 Packets Per Minute

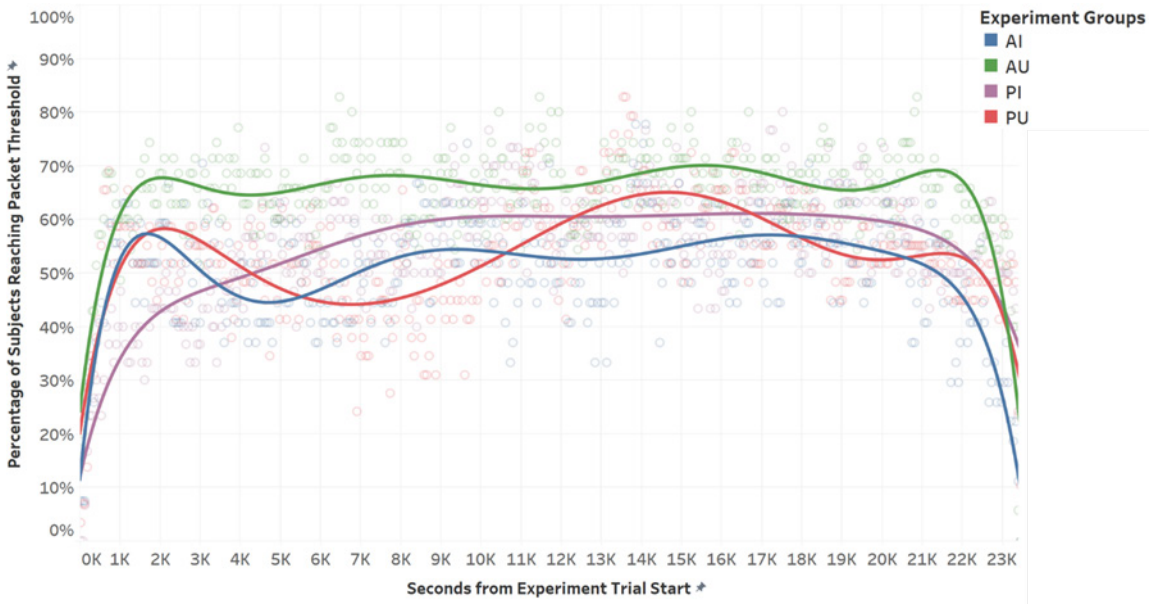


Figure 32. Day 1 Real System Subject Activity Threshold – 7 Packets Per Minute

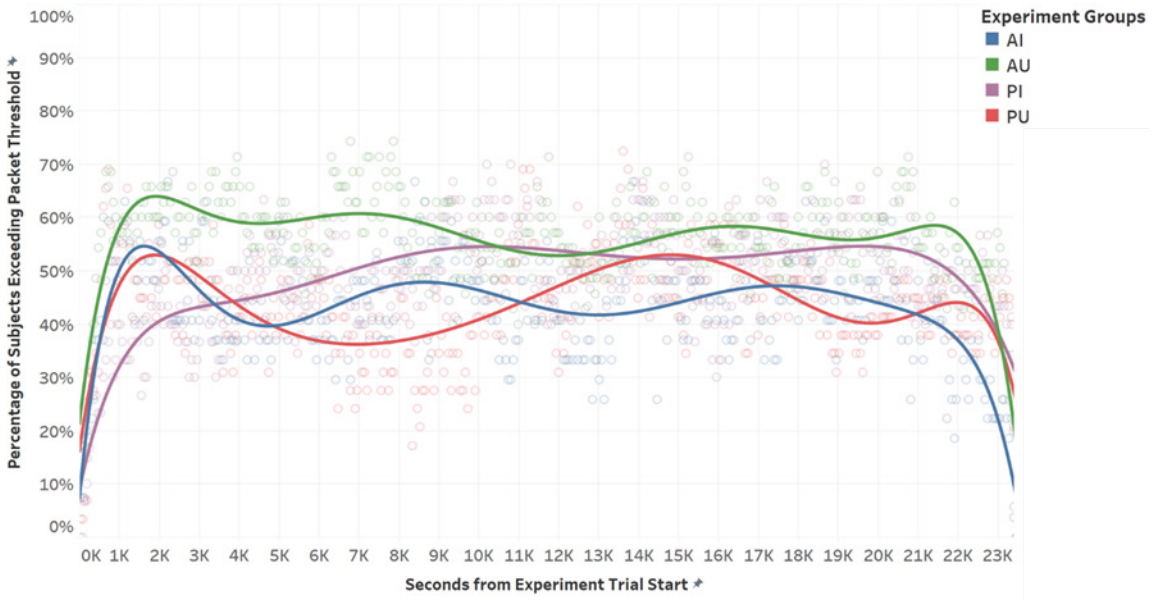


Figure 33. Day 1 Real System Subject Activity Threshold – 20 Packets Per Minute

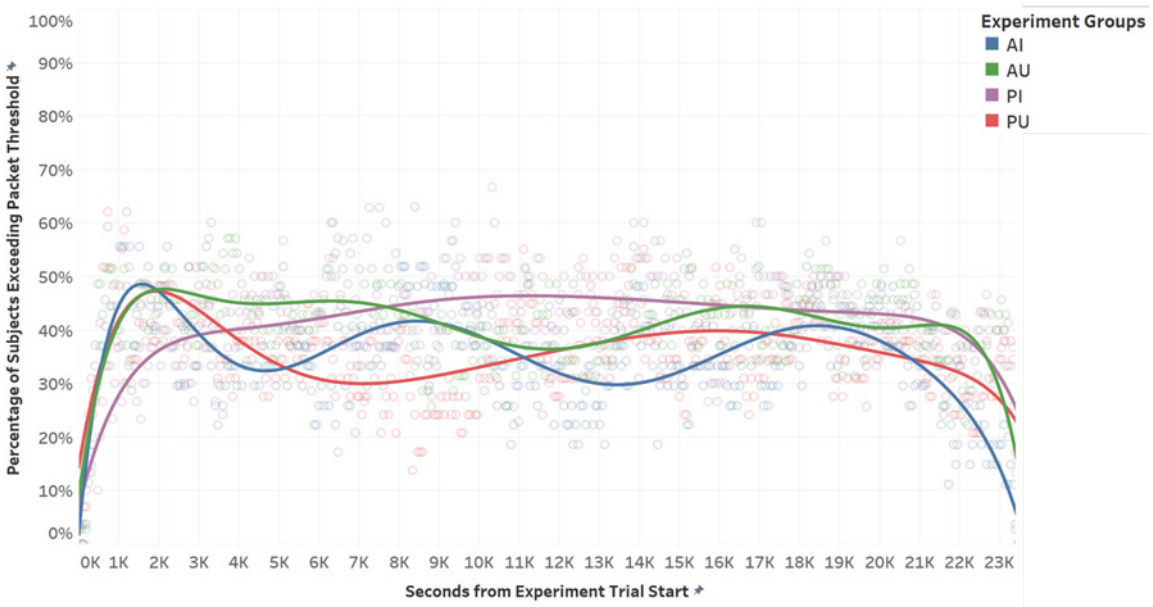


Figure 34. Day 1 Real System Subject Activity Threshold – 54 Packets Per Minute

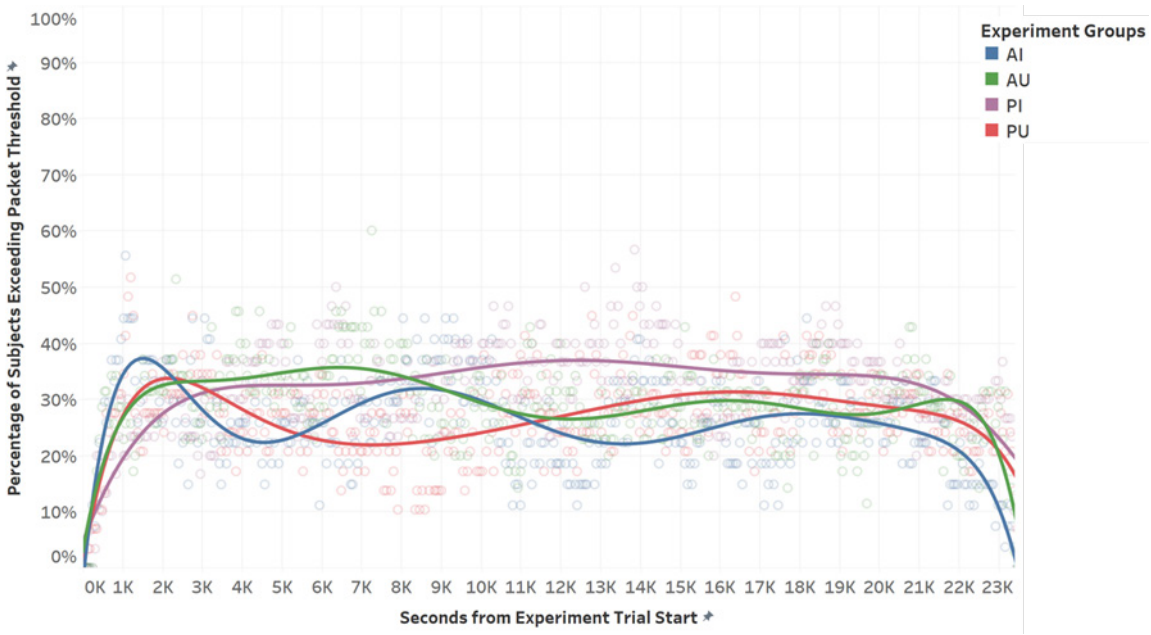


Figure 35. Day 1 Real System Subject Activity Threshold – 148 Packets Per Minute

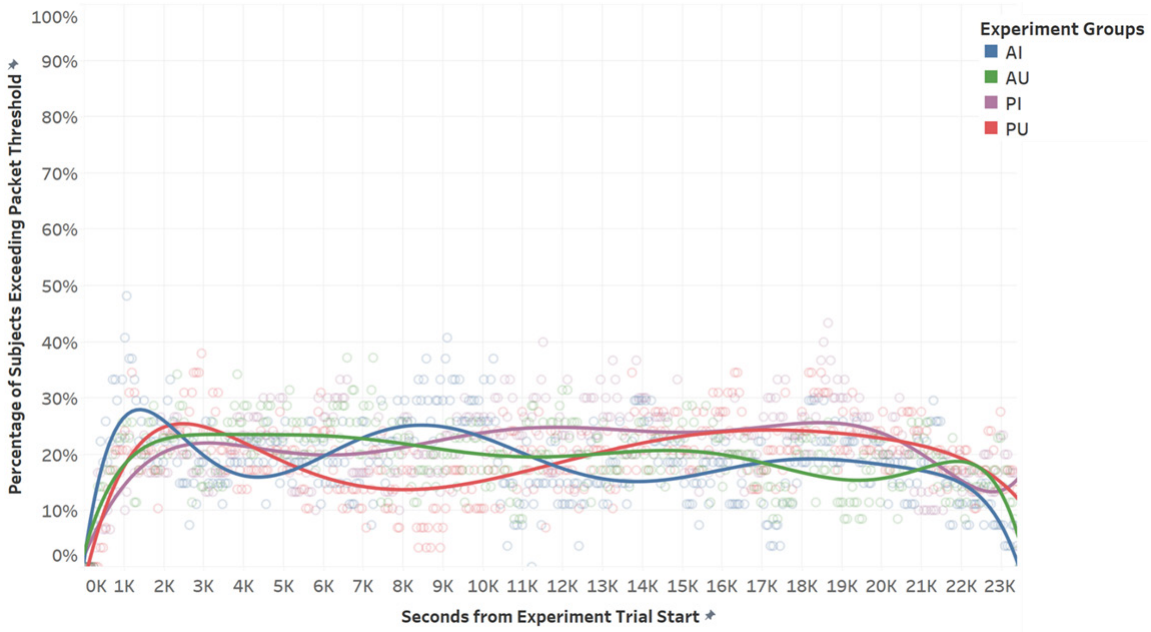


Figure 36. Day 1 Real System Subject Activity Threshold – 403 Packets Per Minute.

Timelines of aggregate subject interaction with real systems on the cyber range network based on the percentage of systems interacted with over time were generated. These charts provide insight into the impact of varying conditions of cyber deception and psychological deception over time on attacker behavior at the group level. Figure 37 depicts the real system interaction percentages over the duration of the Day 1 experiment trials for all groups.

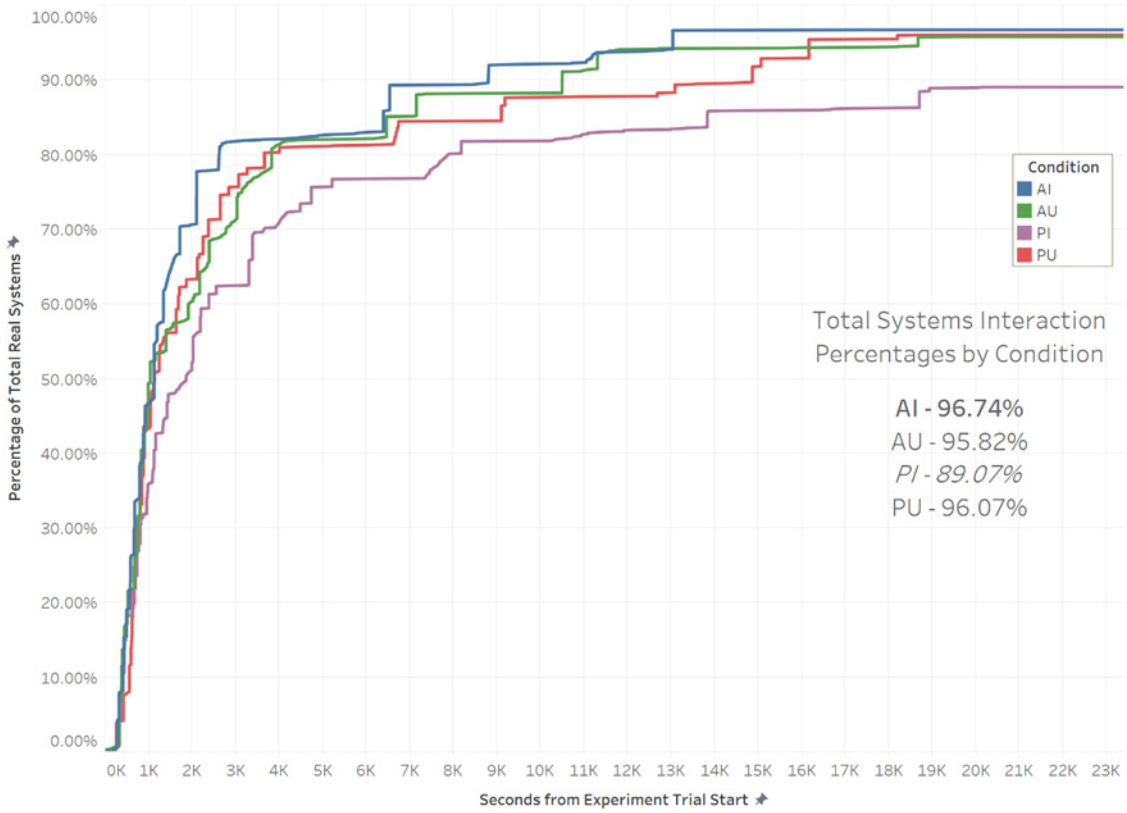


Figure 37. Day 1 Real System Interaction Timeline by Group

c. Transformed Data Distributions

With the latency, duration and frequency variables identified as having non-normal distributions with unequal variances, transformations were applied to the data. Transformations are frequently used for addressing distribution issues, unequal variance,

and outliers in data (Field, 2009). Log, Cube Root, Box-Cox were selected for use as they had the most significant impact on transforming the largely positively skewed data.

(1) Log Transformation

Log transformations apply the inverse function of exponentiation to an input (Osborne, 2013). Each of the values for the variables of latency, frequency and duration for real system interactions were transformed using the function $Log_2(x)$. The resulting values for each condition were then tested using the Shapiro-Wilk-Royston test for normality. For all variables the log transformation results in the data remaining non-normal. Table 16 provides the results of the Shapiro-Wilk-Royston test for normality applied to the log transformation of the latency values for all Day 1 experiment groups. Table 17 provides the results of the Shapiro-Wilk-Royston test for normality applied to the log transformation of the frequency values for all Day 1 experiment groups. Table 18 provides the results of the Shapiro-Wilk-Royston test for normality applied to the log transformation of the duration values for all Day 1 experiment groups.

Table 16. Day 1 Real System Log Transformed Latency Shapiro-Wilk-Royston Test for Normality

	AI	AU	PI	PU
W	0.924	0.919	0.960	0.923
p-value	0.00	0.00	0.00	0.00

For all conditions $p < 0.05$ therefore the null hypothesis that the log transformation of the latency data resulted in a normal distribution is rejected.

Table 17. Day 1 Real System Log Transformed Frequency Shapiro-Wilk-Royston Test for Normality

	AI	AU	PI	PU
W	0.961	0.957	0.944	0.960
p-value	0.00	0.00	0.00	0.00

For all conditions $p<0.05$ therefore the null hypothesis that the log transformation of the frequency data resulted in a normal distribution is rejected.

Table 18. Day 1 Real System Log Transformed Duration Shapiro-Wilk-Royston Test for Normality

	AI	AU	PI	PU
W	0.600	0.592	0.615	0.624
p-value	0.00	0.00	0.00	0.00

For all conditions $p<0.05$ therefore the null hypothesis that the log transformation of the duration data resulted in a normal distribution is rejected.

(2) Cube Root Transformation

Cube Root transformations apply exponentiation of 1/3 to an input. Each of the values for the variables of latency, frequency and duration were transformed using the function $(x)^{1/3}$. These values for each condition were then tested using the Shapiro-Wilk-Royston Test for Normality. For all variables the cube root transformation results in the data remaining non-normal. Table 19 provides the results of the Shapiro-Wilk-Royston test for normality applied to the cube root transformation of the latency values for all Day 1 experiment groups. Table 20 provides the results of the Shapiro-Wilk-Royston test for normality applied to the cube root transformation of the frequency values for all Day 1 experiment groups. Table 21 provides the results of the Shapiro-Wilk-Royston test for normality applied to the cube root transformation of the duration values for all Day 1 experiment groups.

Table 19. Day 1 Real System Cube Root Transformed Latency Shapiro-Wilk-Royston Test for Normality

	AI	AU	PI	PU
W	0.840	0.883	0.884	0.821
p-value	0.00	0.00	0.00	0.00

For all conditions $p<0.05$ therefore the null hypothesis that the cube root transformation of the latency data resulted in a normal distribution is rejected.

Table 20. Day 1 Real System Cube Root Transformed Frequency Shapiro-Wilk-Royston Test for Normality

	AI	AU	PI	PU
W	0.845	0.826	0.840	0.834
p-value	0.00	0.00	0.00	0.00

For all conditions $p<0.05$ therefore the null hypothesis that the cube root transformation of the frequency data resulted in a normal distribution is rejected.

Table 21. Day 1 Real System Cube Root Transformed Duration Shapiro-Wilk-Royston Test for Normality

	AI	AU	PI	PU
W	0.779	0.817	0.834	0.822
p-value	0.00	0.00	0.00	0.00

For all conditions $p<0.05$ therefore the null hypothesis that the cube transformation of the duration data resulted in a normal distribution is rejected.

(3) Box-Cox Transformation

Finally, Box-Cox transformation was applied using the Box-Cox module available in the Python *scipy* library. This module accepts inputs for the lambda transformation coefficient, but also supports automatic calculation of the lambda value that maximizes the log-likelihood function (Virtanen et al., 2020). Each of the values for the variables of latency, frequency and duration were transformed using Box-Cox with calculated lambda values. These values for each condition were then tested using the Shapiro-Wilk-Royston Test for Normality. For all variables the Box-Cox transformation results in the data remaining non-normal. Table 22 provides the results of the Shapiro-Wilk-Royston test for normality applied to the Box-Cox transformation of the latency values for all Day 1 experiment groups. Table 23 provides the results of the Shapiro-Wilk-Royston test for normality applied to the Box-Cox transformation of the frequency values for all Day 1 experiment groups. Table 24 provides the results of the Shapiro-Wilk-Royston test for normality applied to the Box-Cox transformation of the duration values for all Day 1 experiment groups.

Table 22. Day 1 Real System Box-Cox Transformed Latency Shapiro-Wilk-Royston Test for Normality

A	AI	AU	PI	PU
W	0.921	0.929	0.974	0.941
p-value	0.00	0.00	0.00	0.00
Lambda	-0.055	0.043	-0.205	-0.187

For all conditions $p<0.05$ therefore the null hypothesis that the Box Cox transformation of the latency data resulted in a normal distribution is rejected.

Table 23. Day 1 Real System Box-Cox Transformed Frequency Shapiro-Wilk-Royston Test for Normality

	AI	AU	PI	PU
W	0.966	0.961	0.965	0.963
p-value	0.00	0.00	0.00	0.00
Lambda	0.050	0.045	0.107	0.044

For all conditions $p<0.05$ therefore the null hypothesis that the Box Cox transformation of the frequency data resulted in a normal distribution is rejected.

Table 24. Day 1 Real System Box-Cox Transformed Duration Shapiro-Wilk-Royston Test for Normality

	AI	AU	PI	PU
W	0.864	0.904	0.892	0.860
p-value	0.00	0.00	0.00	0.00
Lambda	0.786	0.878	0.727	0.526

For all conditions $p<0.05$ therefore the null hypothesis that the Box Cox transformation of the duration data resulted in a normal distribution is rejected.

d. Winsorization of Temporal Variables

Winsorization provides another technique for addressing outliers within a data distribution (Reifman & Keyton, 2022). Winsorization trims data outliers by replacing extreme high and low values with values that are on the top end or low end of the distribution defined by the specified value percentage (Reifman & Keyton, 2022). The key

benefit of Winsorization is that outliers are normalized with minimal impact on the overall distribution of data (Reifman & Keyton, 2022). Winsorization was performed using Microsoft Excel and validated using the *winsorize* module available in the Python *scipy* library. The temporal variables of latency, frequency and duration were Winsorized at 90% and 80% levels. While median values are unaffected by Winsorization, they are included in the tables as a point of reference. This sub-section provides the descriptive and comparative statistics for real system interactions with 90% Winsorization applied to real systems interaction data from all experiment groups during Day 1 experiment trials.

(1) Latency

Table 25 provides the descriptive statistics for the latency variable for real system interactions for all experiment groups during Day 1 experiment trials. Figure 38 depicts a box-and-whisker plot of latency values for these groups.

Table 25. Day 1 Real System Latency 90% Winsorization Descriptive Statistics

Value	**AI**	**AU**	**PI**	**PU**
Subjects	*27*	**34**	30	29
Observations	1306	1629	1336	1393
Minimum Value	321	**327**	361	*421*
Maximum Value	8,812	*10,499*	**13,829**	14,866
Median	1127	*992*	**1369**	1068
Mean	*2,039*	2,422	**2,894**	2,883
Standard Deviation (Sample)	*2,461*	2,892	3,609	**4,036**
Variance (Sample)	6.06×10^{6}	8.36×10^{6}	1.30×10^{7}	$\mathbf{1.63\times10^{7}}$
Skewness	1.88	*1.83*	**2.01**	2.14
Kurtosis	2.16	*2.28*	**3.11**	3.31
Coefficient of Variability	1.21	*1.19*	1.25	**1.40**

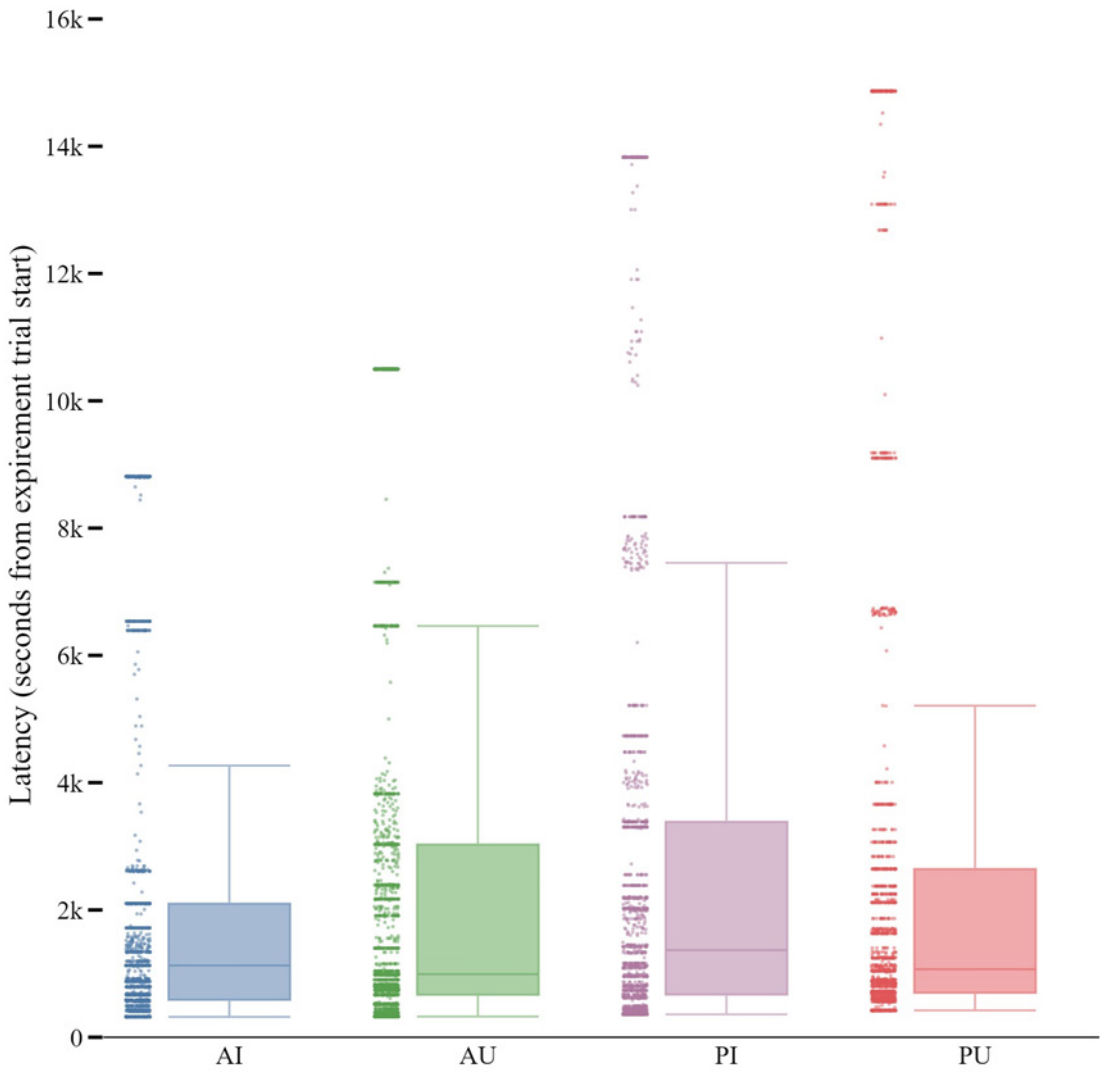


Figure 38. Day 1 Real System Latency 90% Winsorization Box-and-Whisker

(2) Frequency

Table 26 provides the descriptive statistics for the frequency variable for real system interactions for all experiment groups during Day 1 experiment trials. Figure 39 depicts a box-and-whisker plot of frequency values for these groups.

Table 26. Day 1 Real System Frequency 90% Winsorization Descriptive Statistics

Value	**AI**	**AU**	**PI**	**PU**
Subjects	*27*	**34**	30	29
Observations	1306	1629	1336	1393
Minimum Value	**135**	110	699	*52*
Maximum Value	264,657	*259,109*	265,998	**283,996**
Median	6,646	*4188*	**16,086**	4,252
Mean	51,117	*45,076*	**59,656**	55,109
Standard Deviation (Sample)	74,505	*70,852*	77,681	**88,969**
Variance (Sample)	5.55×10^{9}	5.02×10^{9}	6.03×10^{9}	$\mathbf{7.92\times10^{9}}$
Skewness	**1.60**	1.71	*1.40*	1.58
Kurtosis	**1.61**	2.05	*0.96*	1.15
Coefficient of Variability	1.46	1.57	*1.30*	**1.61**

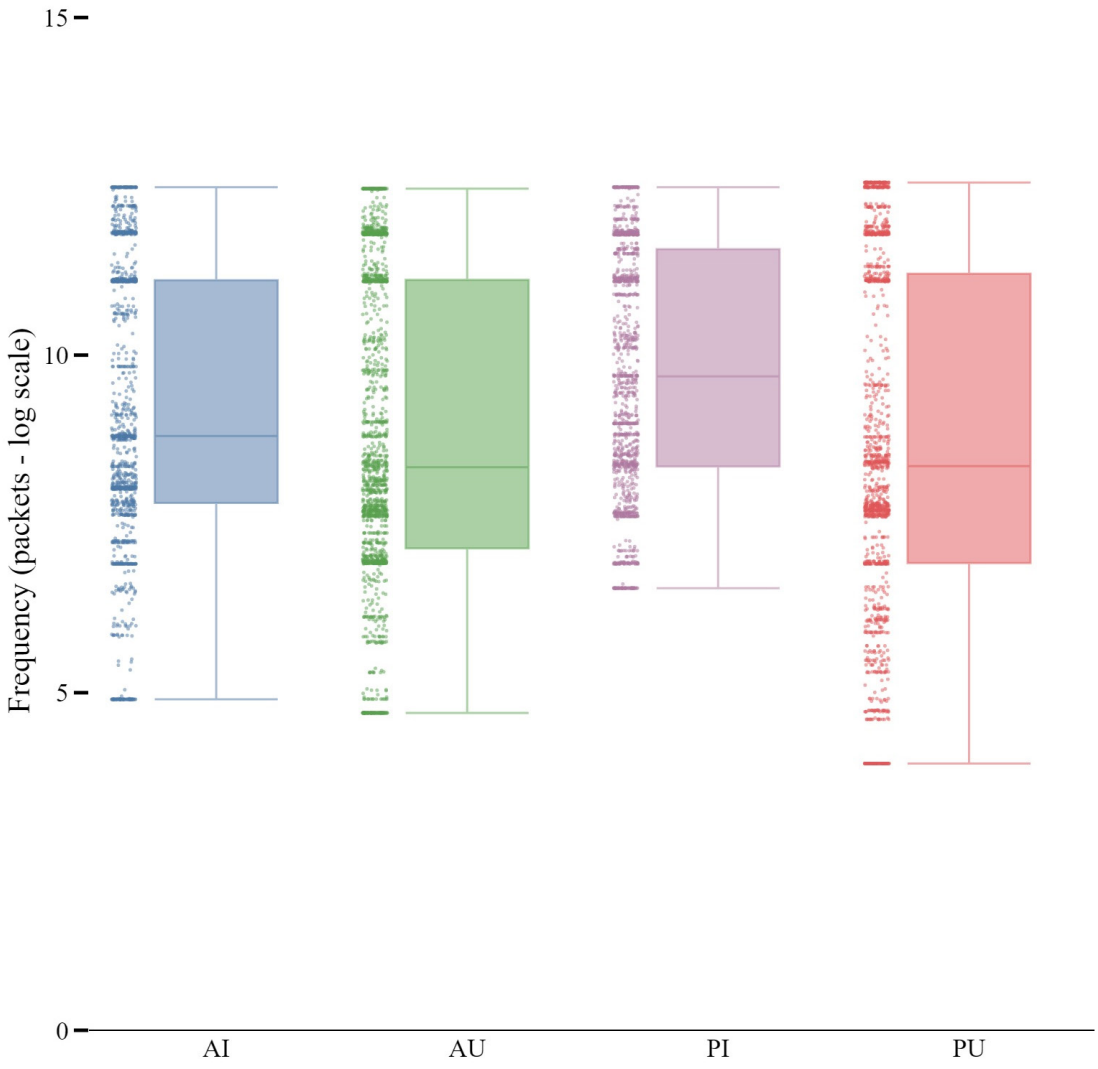


Figure 39. Day 1 Real System Frequency 90% Winsorization Box-and-Whisker Plot.

(3) Duration

Table 27 provides the descriptive statistics for the duration variable for real system interactions for all experiment groups during Day 1 experiment trials. Figure 40 depicts a box-and-whisker plot of duration values for these groups.

Table 27. Day 1 Real System Duration 90% Winsorization Descriptive Statistics

Value	**AI**	**AU**	**PI**	**PU**
Subjects	*27*	**34**	30	29
Observations	1306	1629	1336	1393
Minimum Value	187	**637**	504	*4*
Maximum Value	*21,835*	22,346	22,879	**22,722**
Median	**15,951**	15,358	*14,075*	15,078
Mean	13,442	**14,005**	13,109	*12,605*
Standard Deviation (Sample)	7,194	*6.992*	7,436	**8,210**
Variance (Sample)	5.18×10^{7}	4.89×10^{7}	5.53×10^{7}	$\mathbf{6.74\times10^{7}}$
Skewness	*-.0.636*	-0.490	-0.395	**-0.288**
Kurtosis	***-1.04***	-1.09	-1.23	*-1.51*
Coefficient of Variability	0.535	*0.500*	0.567	**0.651**

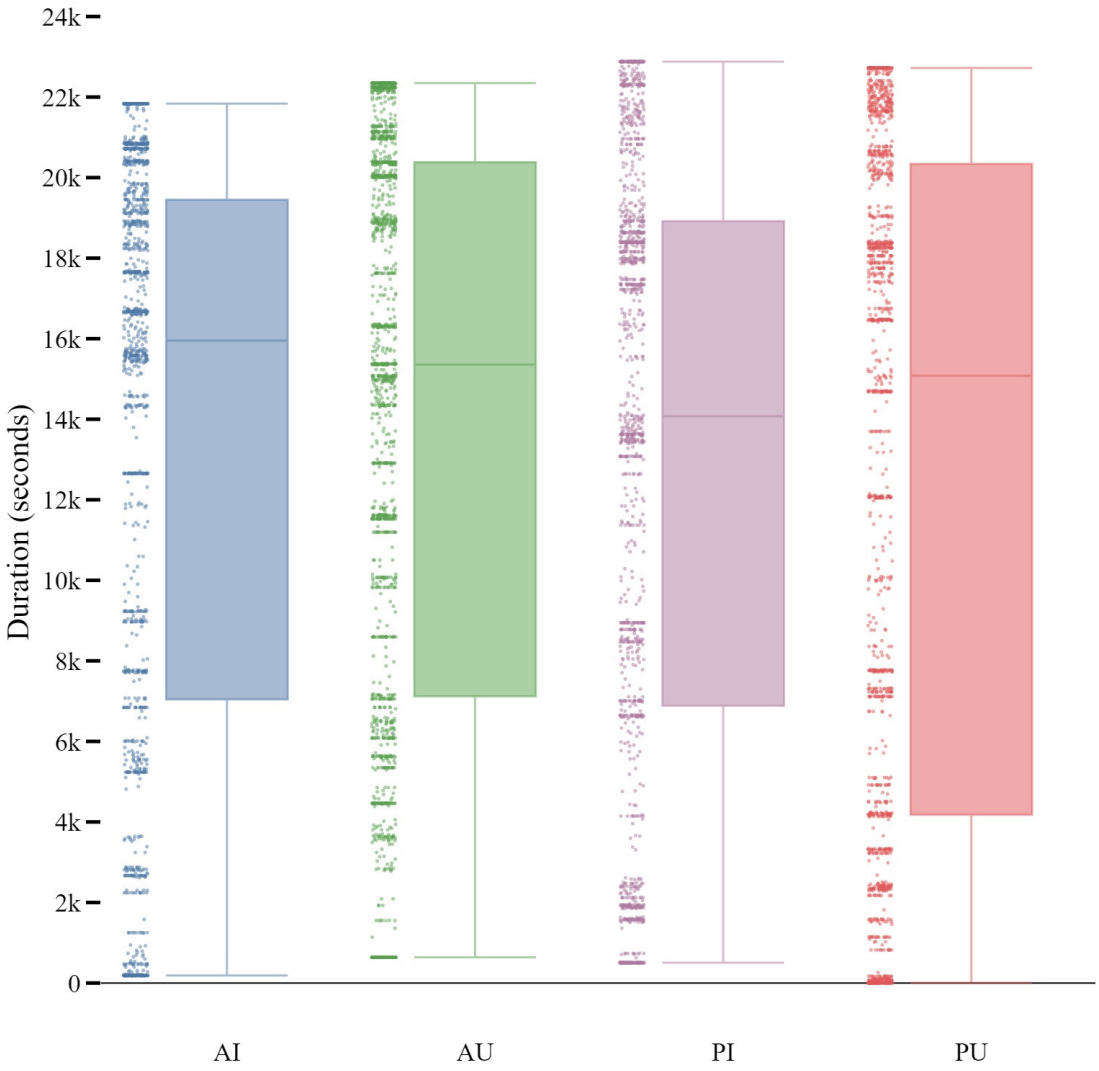


Figure 40. Day 1 Real System Duration 90% Winsorization Box-and-Whisker Plot

e. 80% Winsorization

This sub-section provides the descriptive and comparative statistics for real system interactions with 80% Winsorization applied to real systems interaction data from all experiment groups during Day 1 experiment trials.

(1) Latency

Table 28 provides the descriptive statistics for the latency variable for real system interactions for all experiment groups during Day 1 experiment trials. Figure 41 depicts a box-and-whisker plot of latency values for these groups.

Table 28. Day 1 Real System Latency 80% Winsorization Descriptive Statistics

Value	**AI**	**AU**	**PI**	**PU**
Subjects	*27*	**34**	30	29
Observations	1306	1629	1336	1393
Minimum Value	408	*382*	404	**559**
Maximum Value	*6537*	7149	8,046	**9101**
Median	1127	*992*	**1369**	1068
Mean	*1873*	2160	**2474**	2429
Standard Deviation (Sample)	*2,013*	2,207	2,486	**2,773**
Variance (Sample)	4.05×10^{6}	4.87×10^{6}	6.18×10^{6}	$\mathbf{7.69\times10^{6}}$
Skewness	1.65	1.37	*1.32*	**1.68**
Kurtosis	1.16	0.533	*0.420*	**1.32**
Coefficient of Variability	1.08	1.02	*1.00*	**1.14**

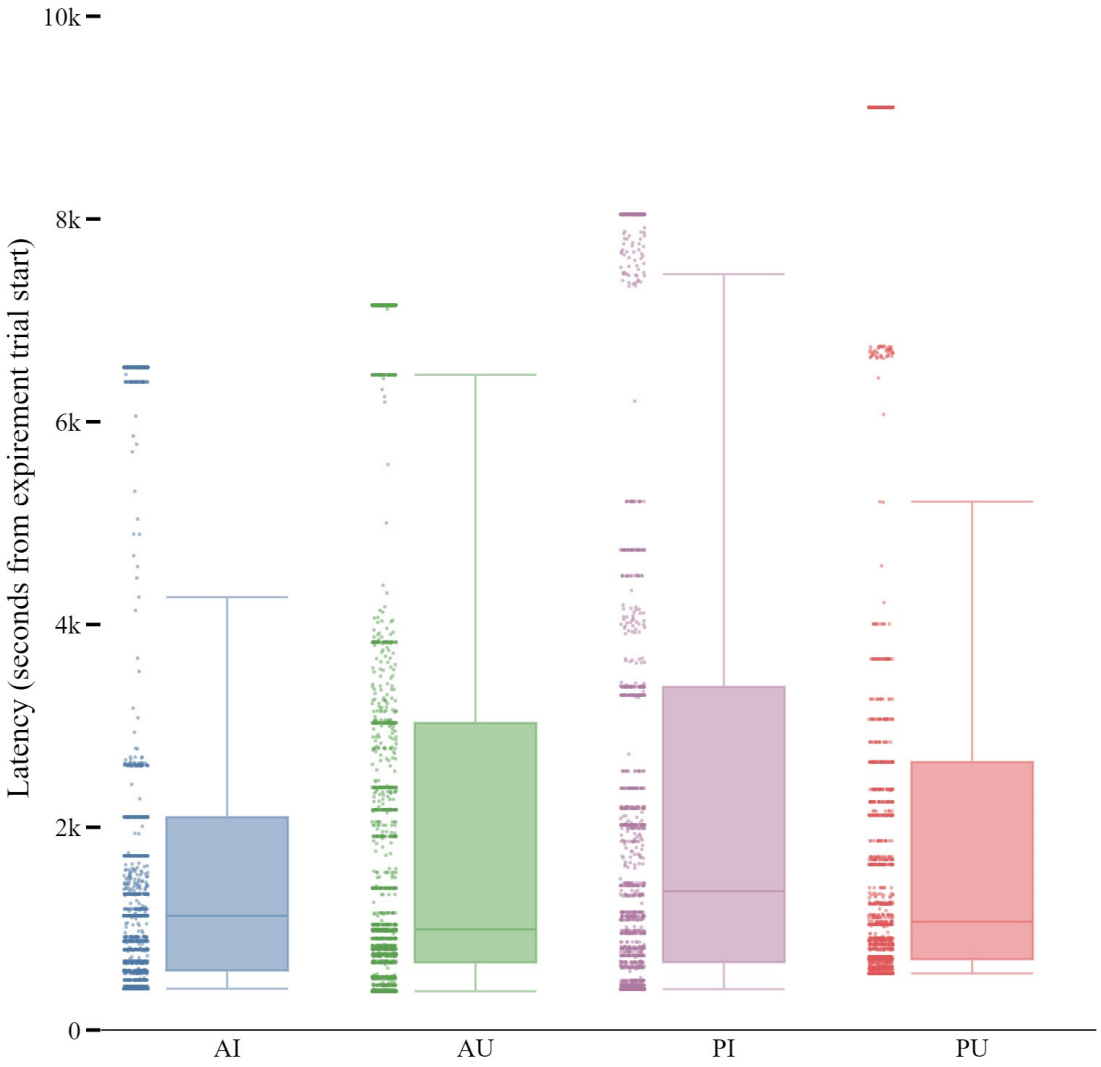


Figure 41. Day 1 Real System Latency 80% Winsorization Box-and-Whisker Plot

(2) Frequency

Table 29 provides the descriptive statistics for the frequency variable for real system interactions for all experiment groups during Day 1 experiment trials. Figure 42 depicts a box-and-whisker plot of frequency values for these groups.

Table 29. Day 1 Real System Frequency 80% Winsorization Descriptive Statistics

Value	**AI**	**AU**	**PI**	**PU**
Subjects	27	34	30	29
Observations	1306	1629	1336	1393
Minimum Value	1000	438	**1215**	*150*
Maximum Value	157,043	*139,036*	165,298	**251,885**
Median	6,646	*4,188*	**16,085**	4,252
Mean	43,746	*37,868*	52,001	**52,747**
Standard Deviation (Sample)	56,663	*53,061*	60,218	**83,139**
Variance (Sample)	3.21×10^{9}	2.82×10^{9}	3.63×10^{9}	$\mathbf{6.91\times10^{9}}$
Skewness	1.04	1.10	*0.840*	**1.47**
Kurtosis	-0.502	-0.516	*-0.927*	**0.738**
Coefficient of Variability	1.30	1.40	*1.16*	**1.58**

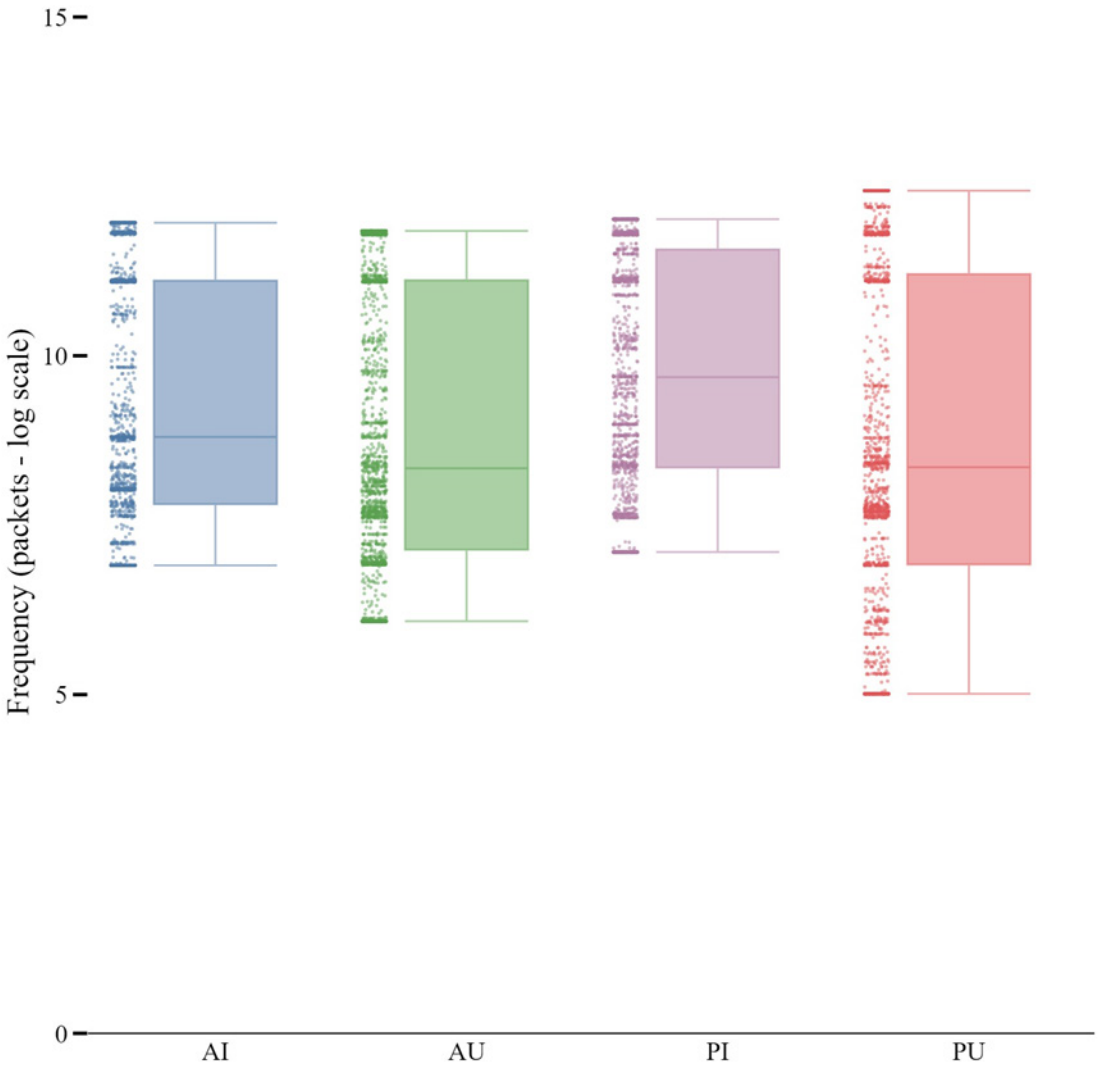


Figure 42. Day 1 Real System Frequency 80% Winsorization Box-and-Whisker Plot

(3) Duration

Table 30 provides the descriptive statistics for the duration variable for real system interactions for all experiment groups during Day 1 experiment trials. Figure 43 depicts a box-and-whisker plot of duration values for these groups.

Table 30. Day 1 Real System Duration 80% Winsorization Descriptive Statistics

Value	**AI**	**AU**	**PI**	**PU**
Subjects	*27*	**34**	30	29
Observations	1306	1629	1336	1393
Minimum Value	1092	**3639**	1585	*166*
Maximum Value	*20,850*	22,221	**22,300**	22,144
Median	**15,951**	15,358	*14,075*	15,078
Mean	13,459	**14,220**	13,150	*12,577*
Standard Deviation (Sample)	6,979	*6,591*	7,261	**8,140**
Variance (Sample)	4.87×10^{7}	4.34×10^{7}	5.27×10^{7}	$\mathbf{6.63\times10^{7}}$
Skewness	*-0.623*	-0.362	-0.378	**-0.296**
Kurtosis	**-1.12**	-1.33	-1.28	*-1.52*
Coefficient of Variability	0.519	*0.463*	0.552	**0.647**

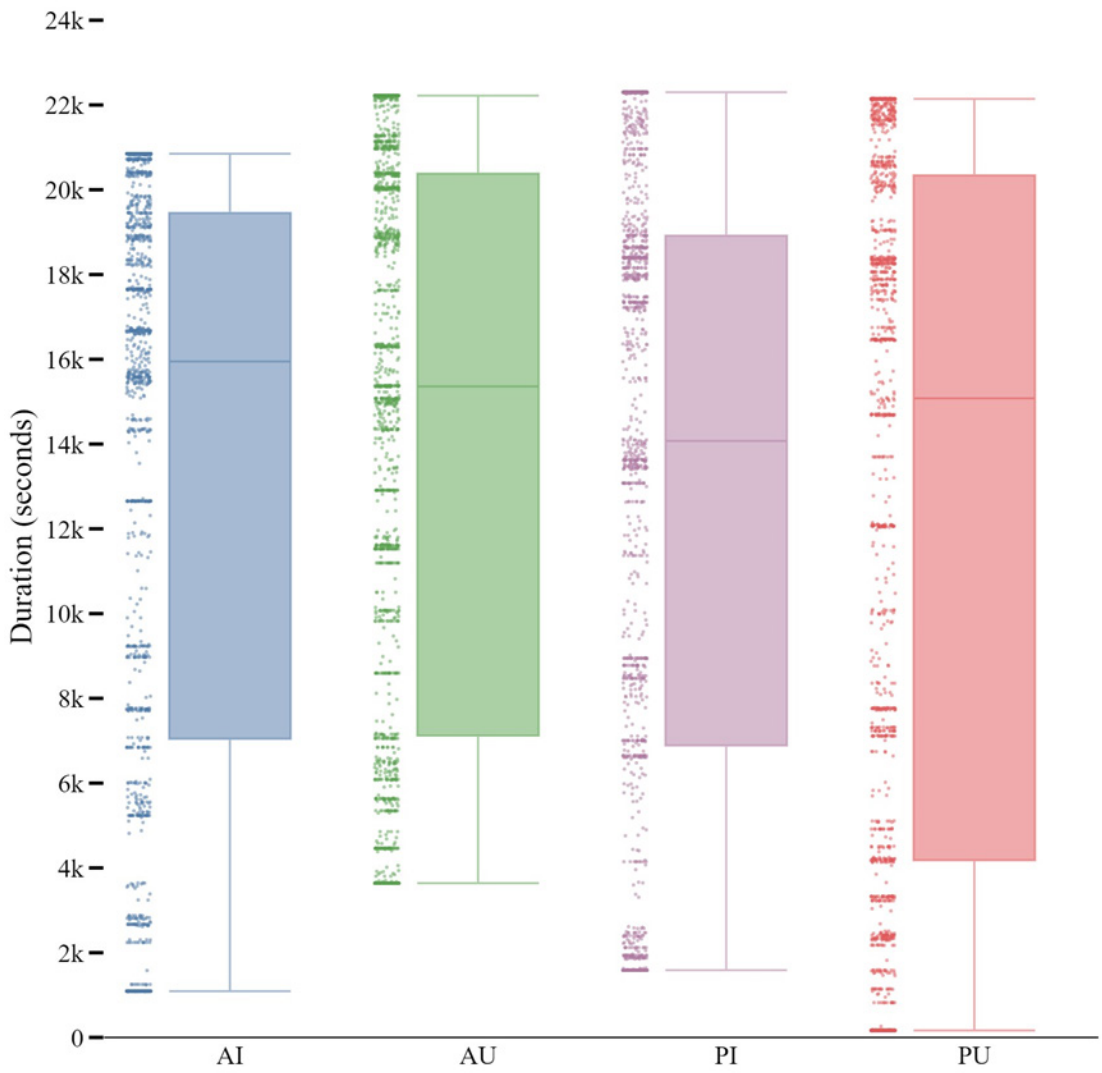


Figure 43. Day 1 Real System Duration 80% Winsorization Box-and-Whisker Plot

2. Decoy System Interactions for Cyber Deception Groups (PI, PU)

This sub-section provides the statistical and graphical analysis for decoy system interactions for Present-Informed and Present-Uninformed groups during Day 1 experiment trials.

a. Descriptive Statistics for All Decoy System Interactions

This sub-section provides the descriptive and comparative statistics for decoy systems interactions for Present-Informed and Present-Uninformed groups during Day 1 experiment trials.

(1) Latency

Table 31 provides the descriptive statistics for the latency variable for decoy system interactions for Present-Informed and Present-Uninformed groups during Day 1 experiment trials. Table 32 provides the results of the Shapiro-Wilk-Royston test for normality for these groups. Table 33 provides the results of Levene's test for equal variance for these groups. Table 34 provides the results of the Kruskal-Wallis test for these groups.

Table 31. Day 1 Decoy System Latency Descriptive Statistics

Value	**PI**	**PU**
Subjects	**30**	*29*
Observations	**1,398**	*1,395*
Minimum Value	**317**	*260*
Maximum Value	**23,058**	*18,896*
Median	**1,425**	*1,111*
Mean	**3,289**	*3,000*
Standard Deviation (Sample)	**4,423**	*4,276*
Variance (Sample)	**1.96×10^{7}**	*1.83×10^{7}*
Skewness	**2.243**	*2.195*
Kurtosis	**4.476**	*3.659*
Coefficient of Variability	*1.345*	**1.425**

Highest values in **Bold**, Lowest values in *Italics*

Table 32. Day 1 Decoy System Latency Shapiro-Wilk-Royston Test for Normality

	PI	PU
W	0.660	0.609
p-value	0.00	0.00

For all conditions $p<0.05$ therefore the null hypothesis that the latency data is normal is rejected.

Table 33. Day 1 Decoy System Latency Levene's Test for Equal Variance

	Centered on Median	Centered on Mean
W	2.93	0.779
p-value	0.087	0.378

For both Median and Mean $p > 0.05$. Null Hypothesis that the data has equal variance is accepted.

Table 34. Day 1 Decoy System Latency Kruskal-Wallis Test

	Real System Interactions
Chi-squared	6.64
p-value	0.010

The population medians are not statistically equal at 1%, 5%, or 10% significance.

(2) Frequency

Table 35 provides the descriptive statistics for the frequency variable for decoy system interactions for Present-Informed and Present-Uninformed groups during Day 1 experiment trials. Table 36 provides the results of the Shapiro-Wilk-Royston test for normality for these groups. Table 37 provides the results of Levene's test for equal variance for these groups. Table 38 provides the results of the Kruskal-Wallis test for these groups.

Table 35. Day 1 Decoy System Frequency Descriptive Statistics

Value	**PI**	**PU**
Subjects	**30**	*29*
Observations	**1,398**	*1,395*
Minimum Value	**7**	*1*
Maximum Value	*1.63×10^{6}*	**8.07×10^{6}**
Median	**9,044**	*2,450*
Mean	*49,513*	**56,399**
Standard Deviation (Sample)	*105,419*	**252,471**
Variance (Sample)	*1.11×10^{10}*	**6.37×10^{10}**
Skewness	*8.264*	**24.215**
Kurtosis	*95.75*	**737.35**
Coefficient of Variability	*2.129*	**4.477**

Highest values in **Bold**, Lowest values in *Italics*

Table 36. Day 1 Decoy System Frequency Shapiro-Wilk-Royston Test for Normality

	PI	PU
W	0.409	0.158
p-value	0.00	0.00

For all conditions $p<0.05$ therefore the null hypothesis that the frequency data is normal is rejected.

Table 37. Day 1 Decoy System Frequency Levene's Test for Equal Variance

	Centered on Median	Centered on Mean
W	1.61	10.747
p-value	0.205	0.001

For both Levene's Test for Equal Variance Centered on Mean $p < 0.05$. Null Hypothesis that the data has equal variance is rejected.

Table 38. Day 1 Decoy System Frequency Kruskal-Wallis Test

	Real System Interactions
Chi-squared	122.21
p-value	0.000

The population medians are not statistically equal at 1%, 5%, or 10% significance.

(3) Duration

Table 39 provides the descriptive statistics for the duration variable for decoy system interactions for Present-Informed and Present-Uninformed groups during Day 1 experiment trials. Table 40 provides the results of the Shapiro-Wilk-Royston test for normality for these groups. Table 41 provides the results of Levene's test for equal variance for these groups. Table 42 provides the results of the Kruskal-Wallis test for these groups.

Table 39. Day 1 Decoy System Duration Descriptive Statistics

Value	**PI**	**PU**
Subjects	**30**	*29*
Observations	**1,398**	*1,395*
Minimum Value	*1*	*1*
Maximum Value	*22,961*	**23,101**
Median	*13,623*	**13,701**
Mean	**12,661**	*12,106*
Standard Deviation (Sample)	*7,064*	**8,032**
Variance (Sample)	4.99×10^{7}	$\mathbf{6.45\times10^{7}}$
Skewness	*-0.366*	**-0.190**
Kurtosis	**-1.122**	*-1.513*
Coefficient of Variability	*0.558*	**0.663**

Highest values in **Bold**, Lowest values in *Italics*

Table 40. Day 1 Decoy System Duration Shapiro-Wilk-Royston Test for Normality

	PI	PU
W	0.925	0.891
p-value	0.00	0.00

For all conditions $p<0.05$ therefore the null hypothesis that the duration data is normal is rejected.

Table 41. Day 1 Decoy System Duration Levene's Test for Equal Variance

	Centered on Median	Centered on Mean
W	75.32	88.04
p-value	0.00	0.00

For both Median and Mean $p < 0.05$. Null Hypothesis that the data has equal variance is rejected.

Table 42. Day 1 Decoy System Duration Kruskal-Wallis Test

	Real System Interactions
Chi-squared	1.343
p-value	0.246

The population medians are statistically equal at 10% significance.

b. Subject Activity Analysis for All Decoy System Interactions

Figure 44 depicts the decoy system interaction percentages over the duration of the Day 1 experiment trials for Present-Uninformed and Present-Informed groups.

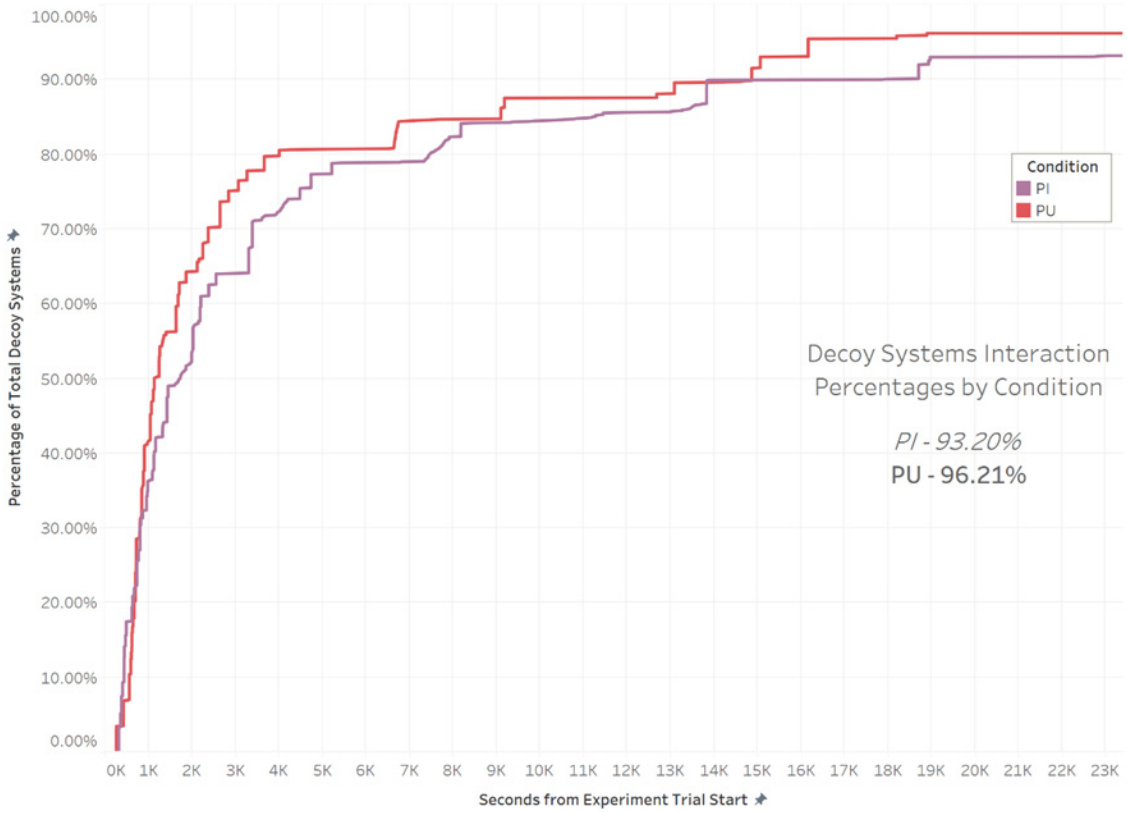


Figure 44. Elapsed Day 1 Decoy System Interaction Percentages By Groups

3. Real and Decoy System Interactions for Cyber Deception Groups (PI, PU)

This sub-section provides the statistical and graphical analysis for real and decoy system interactions for Present-Informed and Present-Uninformed groups during Day 1 experiment trials.

a. Descriptive and Comparative Statistics for All Real and Decoy System Interactions

This sub-section provides the descriptive and comparative statistics for real and decoy systems interactions for Present-Informed and Present-Uninformed groups during Day 1 experiment trials.

(1) Latency

Table 43 provides the descriptive statistics for the latency variable for real and decoy system interactions for Present-Informed and Present-Uninformed groups during Day 1 experiment trials. Table 44 provides the results of the Shapiro-Wilk-Royston test for normality for these groups. Table 45 provides the results of Levene's test for equal variance for these groups. Table 46 provides the results of the Kruskal-Wallis test for these groups.

Table 43. Day 1 Real and Decoy System Latency Descriptive Statistics

Value	**PI**	**PU**
Subjects	**30**	*29*
Observations	*2,734*	**2,788**
Minimum Value	**151**	*8*
Maximum Value	**23,058**	*18,896*
Median	**1,425**	*1,110*
Mean	**3,177**	*2,968*
Standard Deviation (Sample)	**4,310**	*4,250*
Variance (Sample)	$\mathbf{1.86\times10^{7}}$	1.81×10^{7}
Skewness	**2.32**	*2.21*
Kurtosis	**4.91**	*3.72*
Coefficient of Variability	*1.36*	**1.43**

Highest values in **Bold**, Lowest values in *Italics*

Table 44. Day 1 Real and Decoy System Latency Shapiro-Wilk-Royston Test for Normality

	PI	PU
W	0.653	0.607
p-value	0.00	0.00

For all conditions $p<0.05$ therefore the null hypothesis that the latency data is normal is rejected.

Table 45. Day 1 Real and Decoy System Latency Levene's Test for Equal Variance

	Centered on Median	Centered on Mean
W	3.229	0.118
p-value	0.072	0.732

For both Median and Mean $p > 0.05$. Null Hypothesis that the data has equal variance is accepted.

Table 46. Day 1 Real and Decoy System Latency Kruskal-Wallis Test

	Real System Interactions
Chi-squared	9.40
p-value	0.002

The population medians are not statistically equal at 1%, 5%, or 10% significance.

(2) Frequency

Table 47 provides the descriptive statistics for the frequency variable for real and decoy system interactions for Present-Informed and Present-Uninformed groups during Day 1 experiment trials. Table 48 provides the results of the Shapiro-Wilk-Royston test for normality for these groups. Table 49 provides the results of Levene's test for equal variance for these groups. Table 50 provides the results of the Kruskal-Wallis test for these groups.

Table 47. Day 1 Real and Decoy System Frequency Descriptive Statistics

Value	**PI**	**PU**
Subjects	**30**	*29*
Observations	*2,734*	**2,788**
Minimum Value	*1*	*1*
Maximum Value	**3.57×10^{7}**	*8.07×10^{6}*
Median	**12,381**	*2,593*
Mean	**69,043**	*64,644*
Standard Deviation (Sample)	**681,684**	*236,629*
Variance (Sample)	**4.65×10^{11}**	*5.60×10^{10}*
Skewness	**51.14**	*19.45*
Kurtosis	**2,673**	*547.2*
Coefficient of Variability	**9.87**	*3.66*

Highest values in **Bold**, Lowest values in *Italics*

Table 48. Day 1 Real and Decoy System Frequency Shapiro-Wilk-Royston Test for Normality

	PI	PU
W	0.032	0.220
p-value	0.00	0.00

For all conditions $p<0.05$ therefore the null hypothesis that the frequency data is normal is rejected.

Table 49. Day 1 Real and Decoy System Frequency Levene's Test for Equal Variance

	Centered on Median	Centered on Mean
W	0.019	0.331
p-value	0.890	0.565

For both Median and Mean $p < 0.05$. Null Hypothesis that the data has equal variance is rejected.

Table 50. Day 1 Real and Decoy System Frequency Kruskal-Wallis Test

	Real System Interactions
Chi-squared	223.4
p-value	0.000

The population medians are not statistically equal at 1%, 5%, or 10% significance.

(3) Duration

Table 51 provides the descriptive statistics for the duration variable for real and decoy system interactions for Present-Informed and Present-Uninformed groups during Day 1 experiment trials. Table 52 provides the results of the Shapiro-Wilk-Royston test for normality for these groups. Table 53 provides the results of Levene's test for equal variance for these groups. Table 54 provides the results of the Kruskal-Wallis test for these groups.

Table 51. Day 1 Real and Decoy System Duration Descriptive Statistics

Value	**PI**	**PU**
Subjects	30	*29*
Observations	2,734	2,788
Minimum Value	1	1
Maximum Value	23,366	23,201
Median	13,716	**14,695**
Mean	12,875	12,363
Standard Deviation (Sample)	7,266	8,134
Variance (Sample)	5.28×10^7	6.62×10^7
Skewness	-0.379	-0.235
Kurtosis	-1.17	-1.51
Coefficient of Variability	0.564	0.658

Highest values in **Bold**, Lowest values in *Italics*

Table 52. Day 1 Real and Decoy System Duration Shapiro-Wilk-Royston Test for Normality

	PI	PU
W	0.917	0.884
p-value	0.00	0.00

For all conditions $p<0.05$ therefore the null hypothesis that the duration data is normal is rejected.

Table 53. Day 1 Real and Decoy System Duration Levene's Test for Equal Variance

	Centered on Median	Centered on Mean
W	101.83	147.518
p-value	0.000	0.000

For both Median and Mean $p < 0.05$. Null Hypothesis that the data has equal variance is rejected.

Table 54. Day 1 Real and Decoy System Duration Kruskal-Wallis Test

	Real System Interactions
Chi-squared	1.82
p-value	0.177

The population medians are statistically equal at 10% significance.

b. Subject Activity Analysis for Real and Decoy System Interactions

Figure 45 depicts the combined real and decoy system interaction percentages over the duration of the Day 1 experiment trials for Present-Uninformed and Present-Informed groups. Figure 46 depicts the breakout between real and decoy system interaction percentages over the duration of the Day 1 experiment trials for Present-Uninformed and Present-Informed groups.

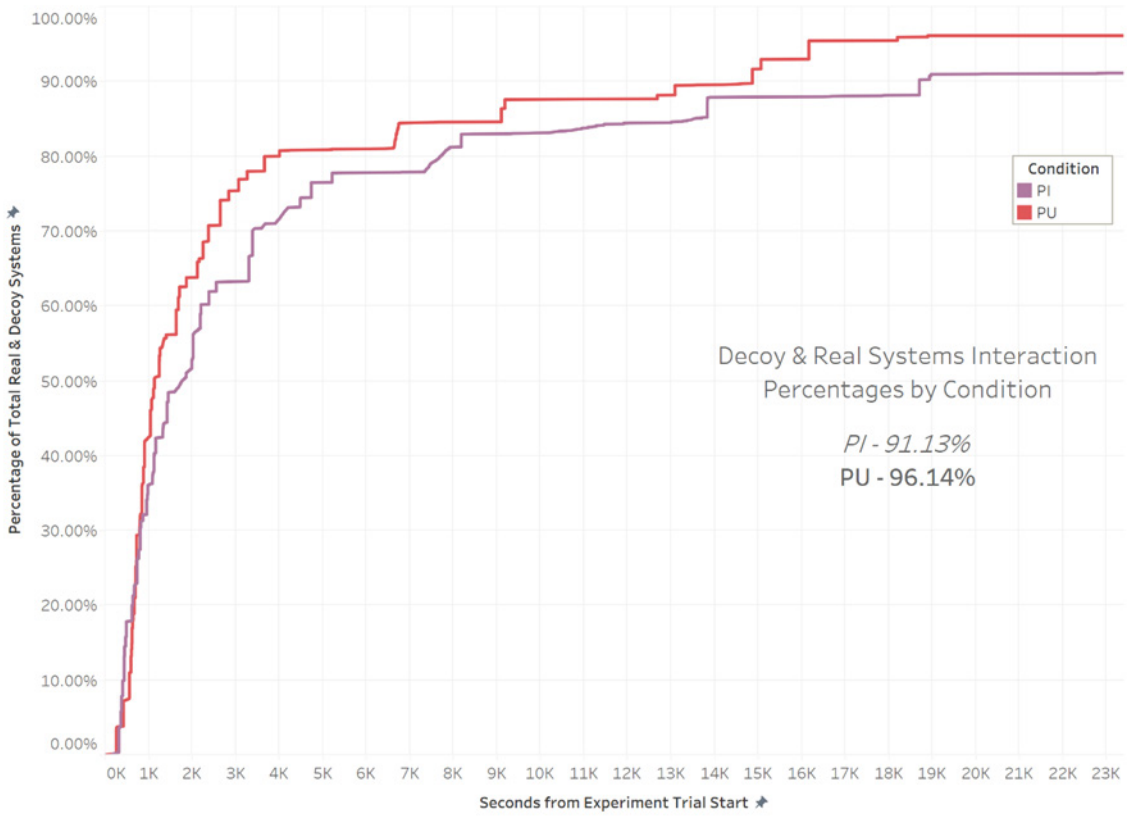


Figure 45. Elapsed Day 1 Real and Decoy System Interaction Percentages

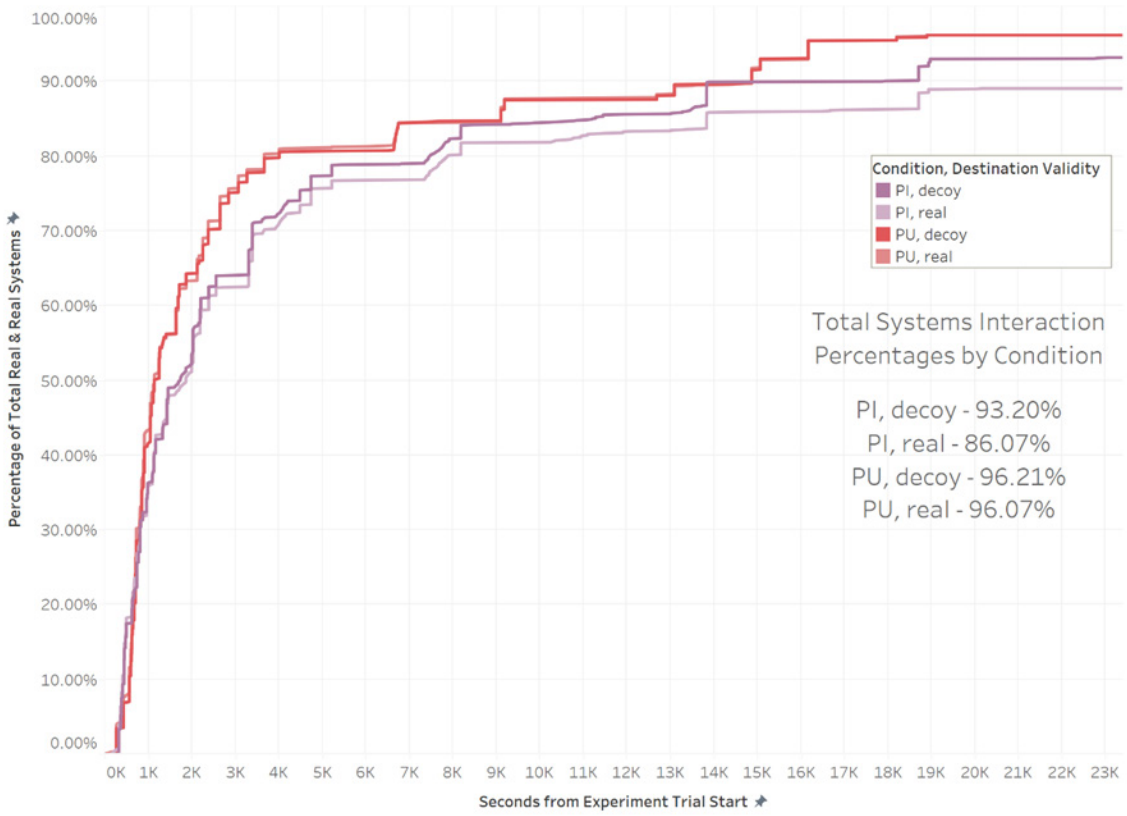


Figure 46. Elapsed Day 1 Split Real and Decoy System Interaction Percentages

D. DAY 2 EXPERIMENT DATA ANALYSIS

This section provides the statistical and graphical analysis for all Tularosa study data collected during Day 2 experiment trials.

1. Real System Interactions Across All Groups (AIAU, AUPU, PIAU, PUAU)

This sub-section provides the statistical and graphical analysis for real system interactions for all experiment groups during Day 2 experiment trials.

a. Descriptive and Comparative Statistics for All Real System Interactions

This sub-section provides the descriptive and comparative statistics for real system interactions for all experiment groups during Day 2 experiment trials.

(1) Latency

Table 55 provides the descriptive statistics for the latency variable for real system interactions for all experiment groups during Day 2 experiment trials. Figure 47 depicts a histogram of latency values for these groups. Figure 48 depicts a box-and-whisker plot of latency values for these groups. Table 56 provides the results of the Shapiro-Wilk-Royston test for normality for these groups. Table 57 provides the results of Levene's test for equal variance for these groups. Table 58 provides the results of the Kruskal-Wallis test for these groups. Finally, Table 59 provides the results of the Dunn's Post Hoc test with Benjamini-Hochberg correction for these groups.

Table 55. Day 2 Real System Latency Descriptive Statistics

Value	**AIAU**	**AUPU**	**PIAU**	**PUAU**
Subjects	*26*	**33**	*26*	*26*
Count of Real Systems	*50*	50	50	50
Count of Observed Real System Interactions (n)	*1,212*	**1,645**	1,283	1,283
Count of Possible Real System Interactions	*1,300*	*1,650*	*1,300*	*1,300*
Percentage of Real Systems Interactions	*93.2%*	**99.7%**	98.7%	98.7%

Value	AIAU	AUPU	PIAU	PUAU
Minimum Value	*0*	*0*	**70**	163
Maximum Value	17,593	13,694	**17,869**	*11,793*
Median	*670*	769	**980**	663
Mean	1,496	1,864	1,464	*1,329*
Standard Deviation (Sample)	2,327	**3,016**	1,713	*2,017*
Variance (Sample)	5.89×10^{6}	$\mathbf{9.09\times10^{6}}$	2.93×10^{6}	4.07×10^{6}
Skewness	3.30	2.96	*2.62*	**3.39**
Kurtosis	11.22	8.16	*7.02*	**10.87**
Coefficient of Variability	1.60	**1.62**	*1.17*	1.52

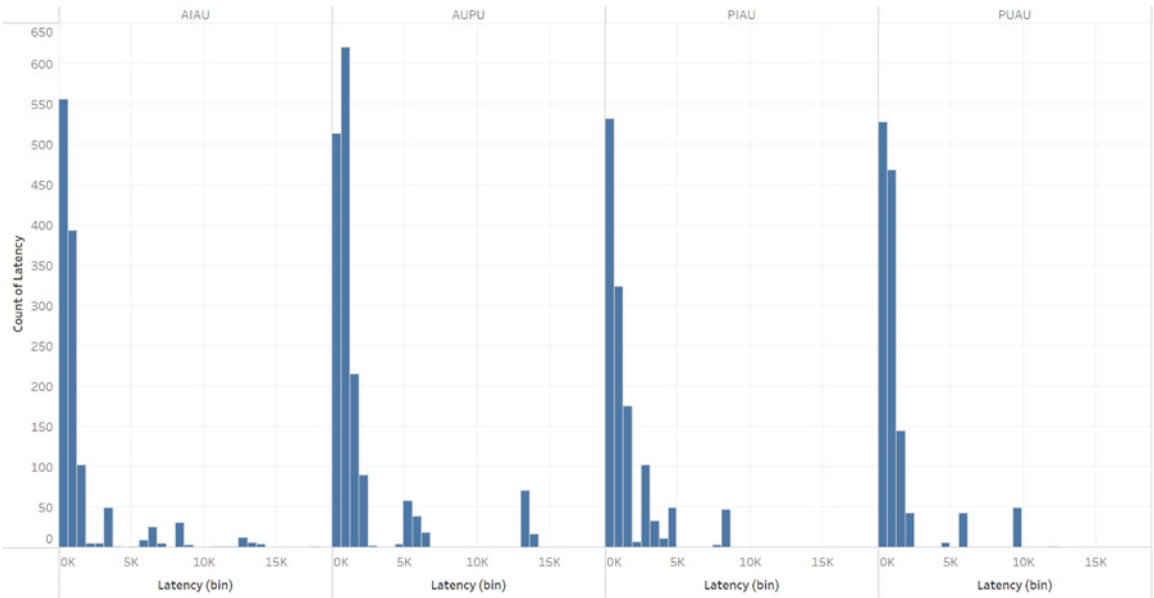


Figure 47. Day 2 Real System Latency Histogram

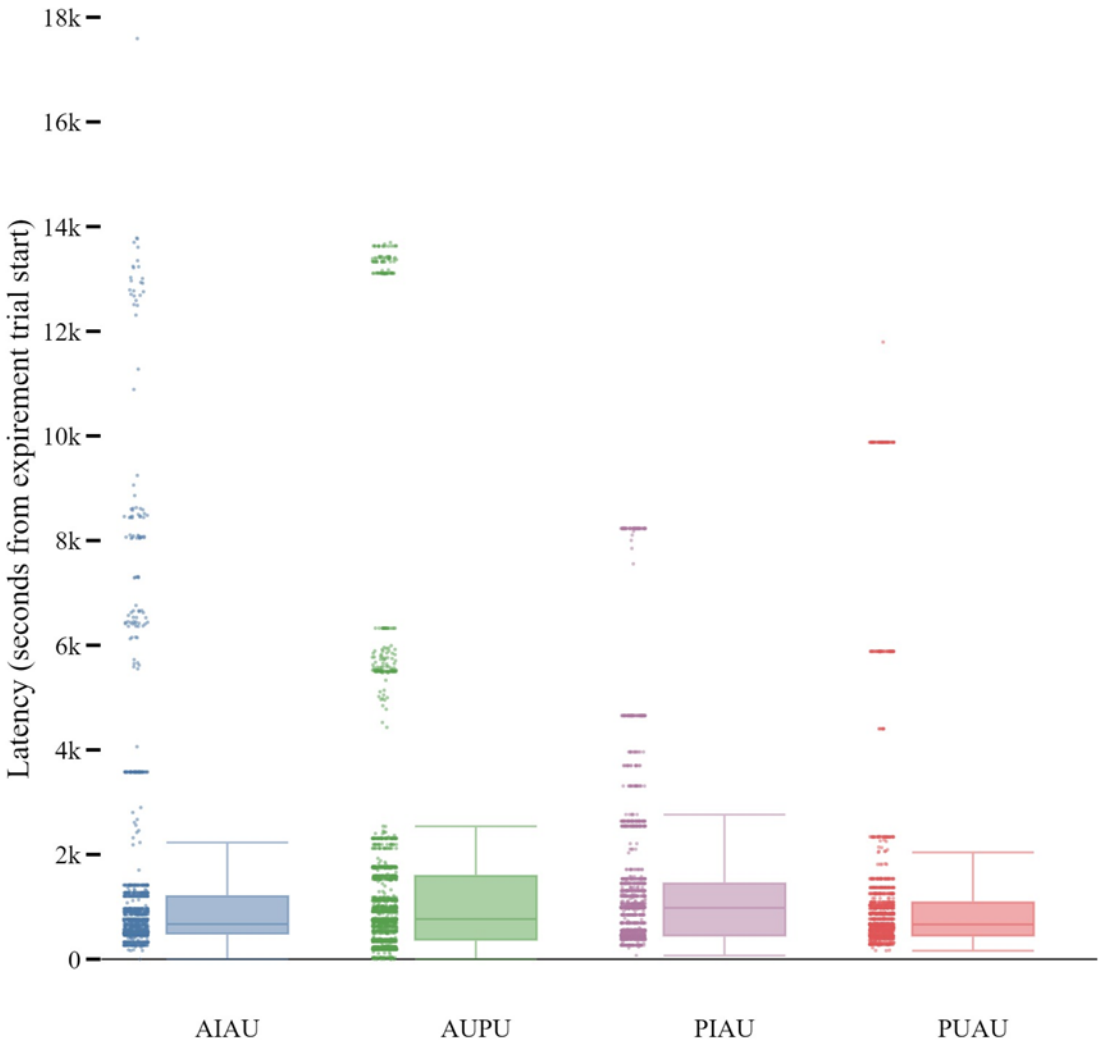


Figure 48. Day 2 Real System Latency Box-and-Whisker Plot

Table 56. Day 2 Real System Latency Shapiro-Wilk-Royston Test for Normality

	AIAU	**AUPU**	**PIAU**	**PUAU**
W	0.488	0.534	0.638	0.464
p-value	0.00	0.00	0.00	0.00

The latency values for each group were assessed for normality using the Shapiro-Wilk-Royston Test. For all groups the null hypothesis that the latency data is normal is rejected with a p-value of 0.00.

Table 57. Day 2 Real System Latency Levene's Test for Equal Variance

	Centered on Median	Centered on Mean
W	17.72	42.38
p-value	0.000	0.000

The latency values for each group were assessed for equal variance using the Levene's test. For both median and mean, the groups have equal variance.

Table 58. Day 2 Real System Latency Kruskal-Wallis Test

	Real System Interactions
Chi-squared	31.42
p-value	0.000

The population medians are not statistically equal at 1%, 5%, or 10% significance.

Table 59. Day 2 Real System Latency Dunn's Post Hoc Test with Benjamini-Hochberg Correction

	AIAU	AUPU	PIAU	PUAU
AIAU	1.00			
AUPU	**0.0002**	1.00		
PIAU	**0.0001**	0.6199	1.00	
PUAU	0.7329	**0.0005**	**0.0002**	1.00

Pairwise comparisons with Dunn's test indicate latency values for the AIAU group are significantly different from the AUPU group (p=0.0002) and PIAU group (p=0.0001), for the AUPU group and PUAU group (p=0.0005) and PIAU group and PUAU group (p=0.0002).

(2) Frequency

Table 60 provides the descriptive statistics for the frequency variable for real system interactions for all experiment groups during Day 2 experiment trials. Figure 49 depicts a histogram of frequency values for these groups. Figure 50 depicts a box-and-whisker plot of frequency values for these groups. Table 61 provides the results of the

Shapiro-Wilk-Royston test for normality for these groups. Table 62 provides the results of Levene's test for equal variance for these groups. Table 63 provides the results of the Kruskal-Wallis test for these groups. Finally, Table 64 provides the results of the Dunn's Post Hoc test with Benjamini-Hochberg correction for these groups.

Table 60. Day 2 Real System Frequency Descriptive Statistics

Value	**AIAU**	**AUPU**	**PIAU**	**PUAU**
Subjects	*26*	**33**	*26*	*26*
Observations	*1,212*	**1,645**	1,283	1,283
Minimum Value	*6*	*1*	*18*	*8*
Maximum Value	*9.10×10^{6}*	9.18×10^{6}	**2.80×10^{7}**	1.52×10^{6}
Median	**65,555**	*4,154*	12,132	12,025
Mean	**104,906**	*53,417*	90,520	76,642
Standard Deviation (Sample)	413,477	325,731	**789,267**	*133,084*
Variance (Sample)	1.71×10^{11}	1.06×10^{11}	**6.23×10^{11}**	*1.77×10^{10}*
Skewness	14.55	20.09	**34.39**	*4.92*
Kurtosis	251.8	471.8	**1,214**	*38.77*
Coefficient of Variability	3.94	6.10	*8.72*	*1.74*

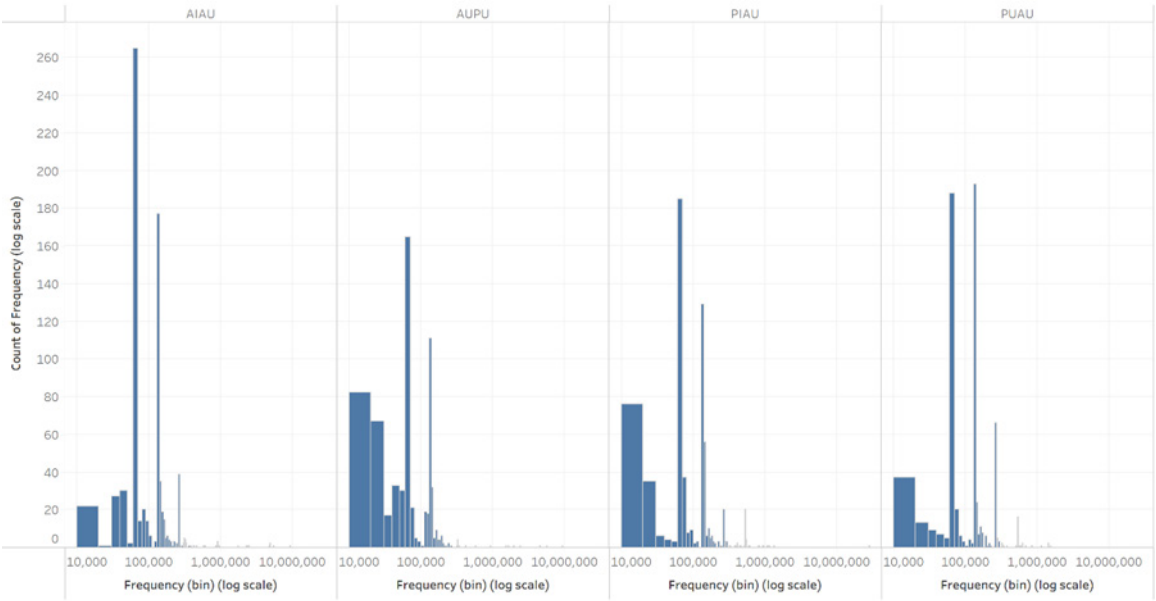


Figure 49. Day 2 Real System Frequency Histogram

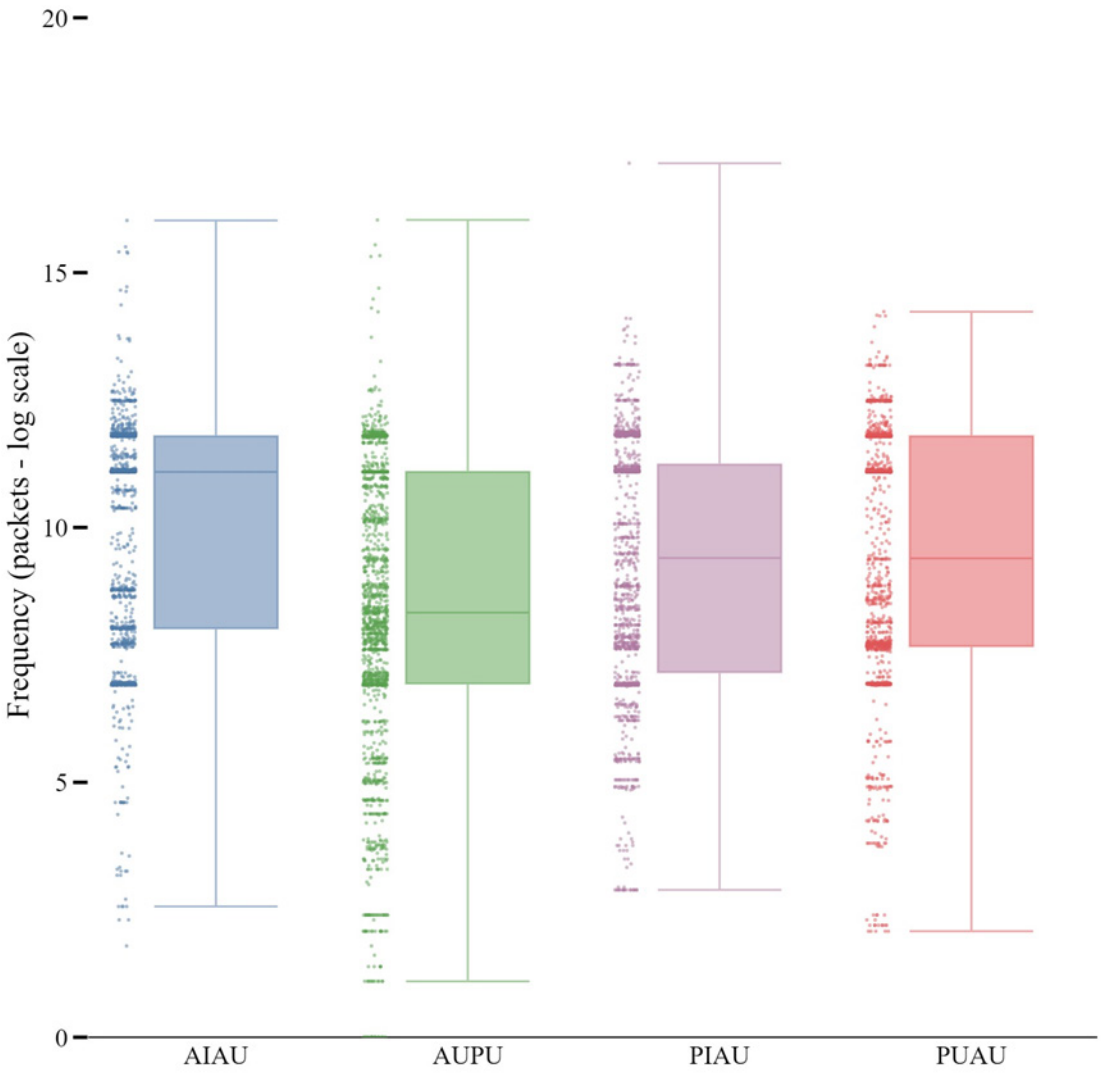


Figure 50. Day 2 Real System Frequency Box-and-Whisker Plot

Table 61. Day 2 Real System Frequency Shapiro-Wilk-Royston Test for Normality

	AIAU	**AUPU**	**PIAU**	**PUAU**
W	0.158	0.099	0.046	0.560
p-value	0.00	0.00	0.00	0.00

The frequency values for each group were assessed for normality using the Shapiro-Wilk-Royston Test. For all groups the null hypothesis that the frequency data is normal is rejected with a p-value of 0.00.

Table 62. Day 2 Real System Frequency Levene's Test for Equal Variance

	Centered on Median	Centered on Mean
W	2.134	2.366
p-value	0.094	0.069

The frequency values for each group were assessed for equal variance using the Levene's test. For both median and mean, the groups do not have equal variance with $p > 0.05$.

Table 63. Day 2 Real System Frequency Kruskal-Wallis Test

	Real System Interactions
Chi-squared	252.43
p-value	0.000

The population medians are not statistically equal at 1%, 5%, or 10% significance.

Table 64. Day 2 Real System Frequency Dunn's Post Hoc Test with Benjamini-Hochberg Correction

	AIAU	AUPU	PIAU	PUAU
AIAU	1.00			
AUPU	**0.0000**	1.00		
PIAU	**0.0000**	**0.0000**	1.00	
PUAU	**0.0000**	**0.0000**	0.1001	1.00

Pairwise comparisons with Dunn's test indicate frequency values for the AIAU group are significantly different from the AUPU group (p=0.0000), PIAU group (p=0.0000) and PUAU group (p = 0.0000), for the AUPU group and both PIAU and PUAU groups (p=0.0000 and p = 0.0000).

(3) Duration

Table 65 provides the descriptive statistics for the duration variable for real system interactions for all experiment groups during Day 2 experiment trials. Figure 51 depicts a histogram of duration values for these groups. Figure 52 depicts a box-and-whisker plot of

duration values for these groups. Table 66 provides the results of the Shapiro-Wilk-Royston test for normality for these groups. Table 67 provides the results of Levene's test for equal variance for these groups. Table 68 provides the results of the Kruskal-Wallis test for these groups. Finally, Table 69 provides the results of the Dunn's Post Hoc test with Benjamini-Hochberg correction for these groups.

Table 65. Day 2 Real System Duration Descriptive Statistics

Value	**AIAU**	**AUPU**	**PIAU**	**PUAU**
Subjects	*26*	**33**	*26*	*26*
Observations	*1,212*	**1,645**	1,283	1,283
Minimum Value	*1*	*1*	*2*	*7*
Maximum Value	23,339	**23,340**	*23,184*	23,190
Median	15,110	**15,826**	*14,368*	14,619
Mean	12,692	**14,026**	*12,314*	12,843
Standard Deviation (Sample)	8,080	*7,874*	**8,689**	8,089
Variance (Sample)	6.53×10^{7}	6.20×10^{7}	$\mathbf{7.55\times10^{7}}$	6.54×10^{7}
Skewness	-0.360	*-0.564*	**-0.245**	-0.394
Kurtosis	-1.387	**-1.104**	*-1.573*	-1.377
Coefficient of Variability	0.637	*0.561*	**0.706**	0.630

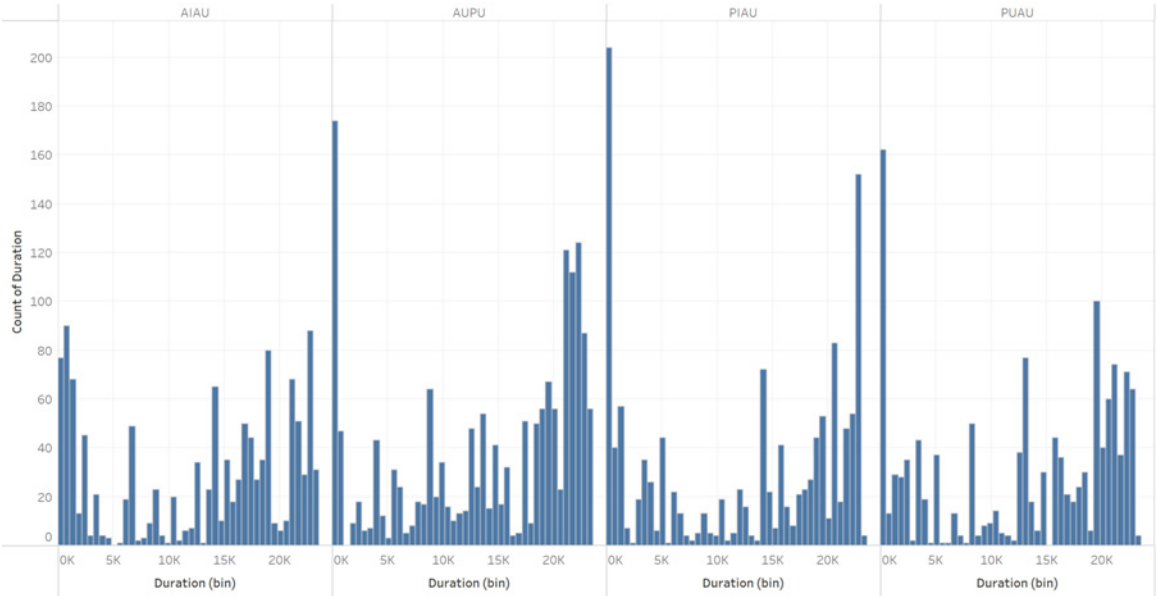


Figure 51. Day 2 Real System Duration Histogram

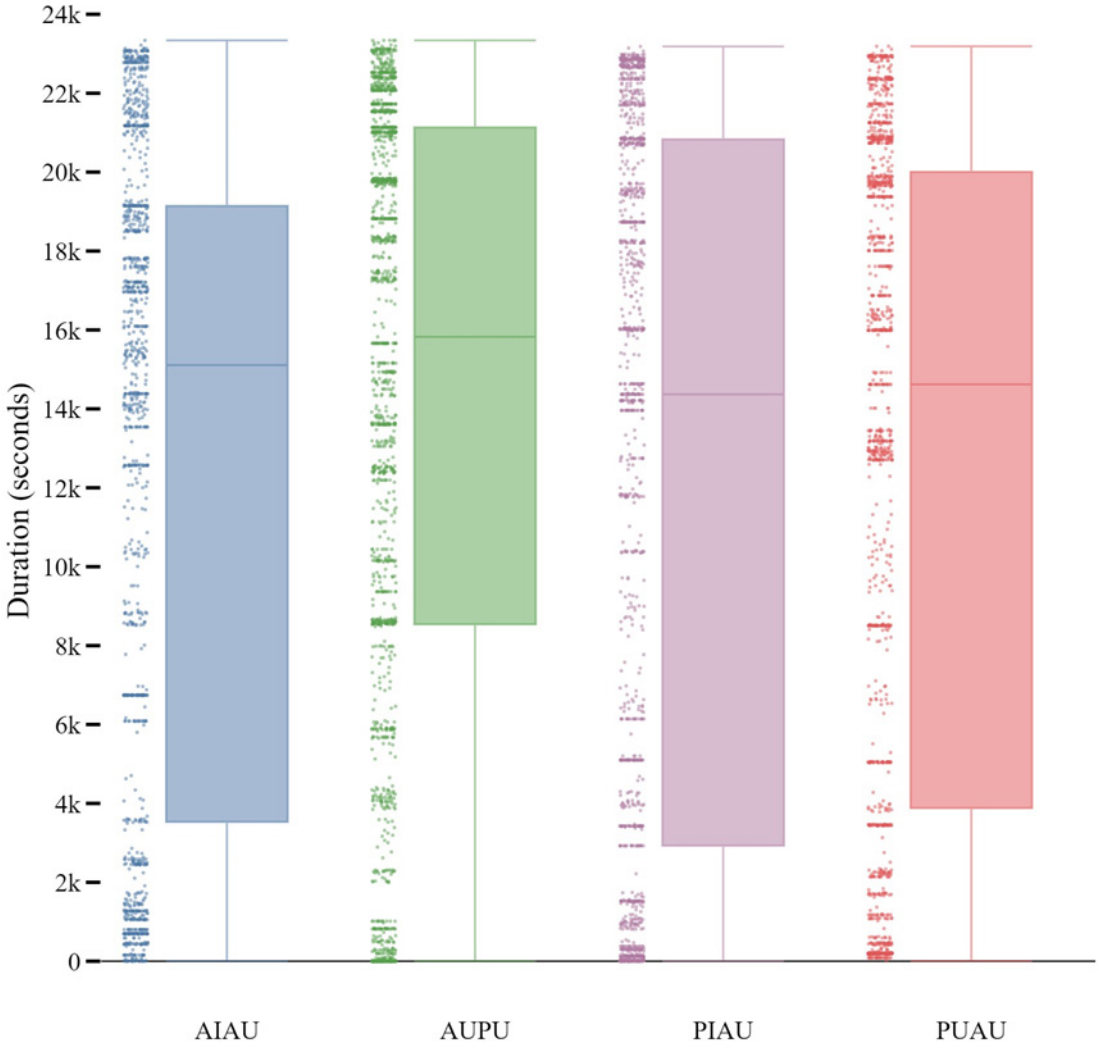


Figure 52. Day 2 Real System Duration Box-and-Whisker Plot

Table 66. Day 2 Real System Duration Shapiro-Wilk-Royston Test for Normality

	AIAU	**AUPU**	**PIAU**	**PUAU**
W	0.880	0.875	0.857	0.875
p-value	0.00	0.00	0.00	0.00

The duration values for each group were assessed for normality using the Shapiro-Wilk-Royston Test. For all groups the null hypothesis that the duration data is normal is rejected with a p-value of 0.00.

Table 67. Day 2 Real System Duration Levene's Test for Equal Variance

	Centered on Median	Centered on Mean
W	2.13	21.32
p-value	0.094	0.000

The duration values for each group were assessed for equal variance using the Levene's test. For both median and mean, the groups do not have equal variance.

Table 68. Day 2 Real System Duration Kruskal-Wallis Test

	Real System Interactions
Chi-squared	29.18
p-value	0.000

The population medians are not statistically equal at 1%, 5%, or 10% significance.

Table 69. Day 2 Real System Duration Dunn's Post Hoc Test with Benjamini-Hochberg Correction

	AIAU	AUPU	PIAU	PUAU
AIAU	1.00			
AUPU	**0.0002**	1.00		
PIAU	0.6023	**0.0001**	1.00	
PUAU	0.9432	**0.0002**	0.6023	1.00

Pairwise comparisons with Dunn's test indicate duration values for the AUPU group are significantly different from the AIAU group (p=0.002), the PIAU group (p=0.0001) and the PUAU group (p=0.0002).

b. Subject Activity Analysis for All Real System Interactions

This sub-section provides the graphical analysis for real system interactions for all experiment groups during Day 2 experiment trials. Figure 53 depicts the natural log of the average packets per minute sent by each group from the start to the end of the Day 2

experiment trials. Figure 54 depicts the natural log of the median packets per minute sent by each group from the start to the end of the Day 2 experiment trials. Figure 55 depicts the standard deviation of the natural log of the average packets per minute sent by each group from the start to the end of the Day 2 experiment trials. Figure 56 depicts the variance of the natural log of the average packets per minute sent by each group from the start to the end of the Day 2 experiment trials.

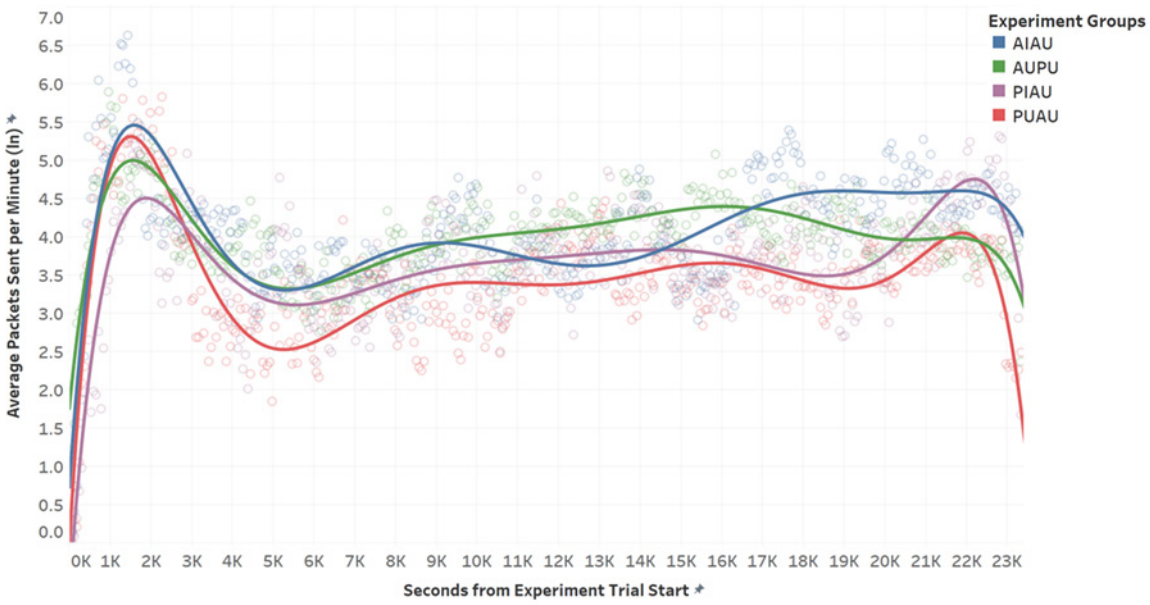


Figure 53. Day 2 Real System Average Packets Per Minute

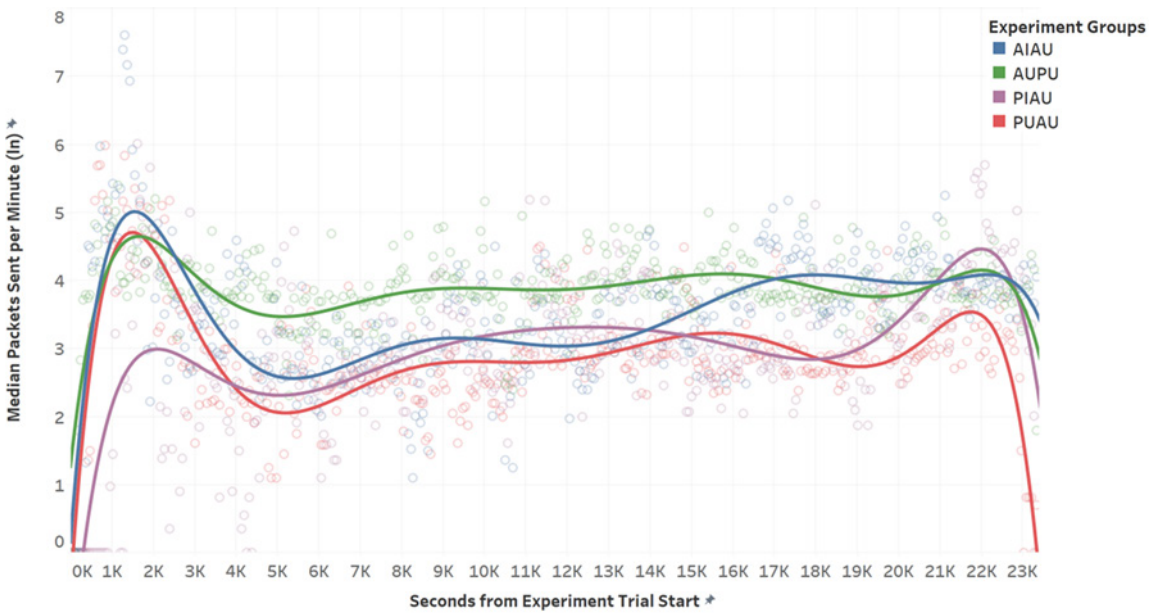


Figure 54. Day 2 Real System Median Packets Per Minute

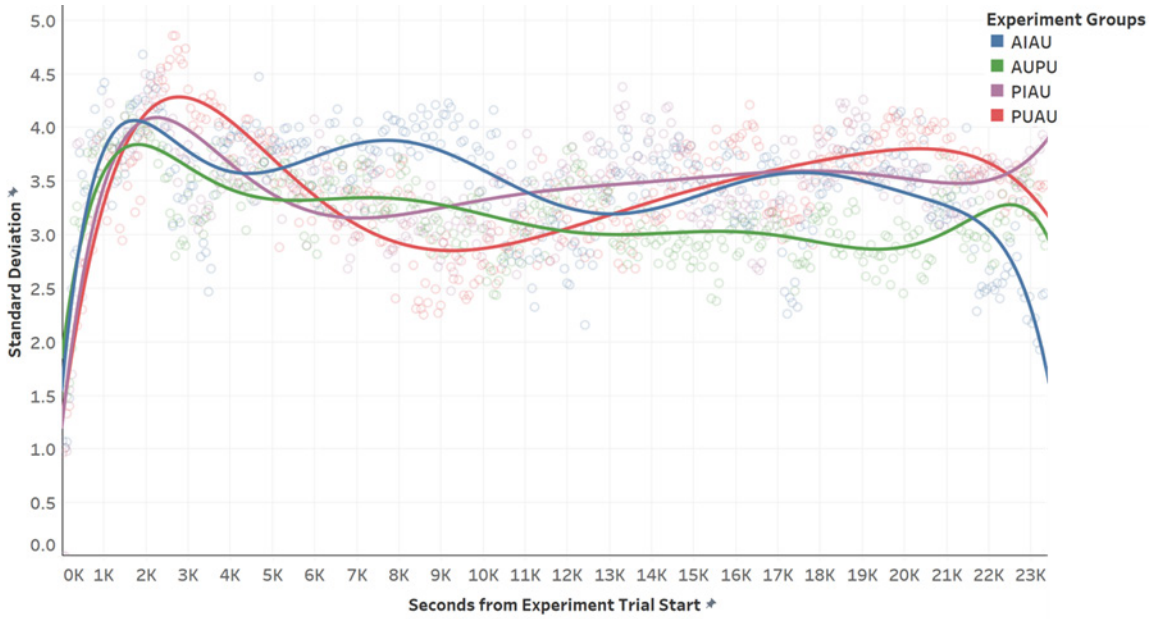


Figure 55. Day 2 Real System Standard Deviation Per Minute

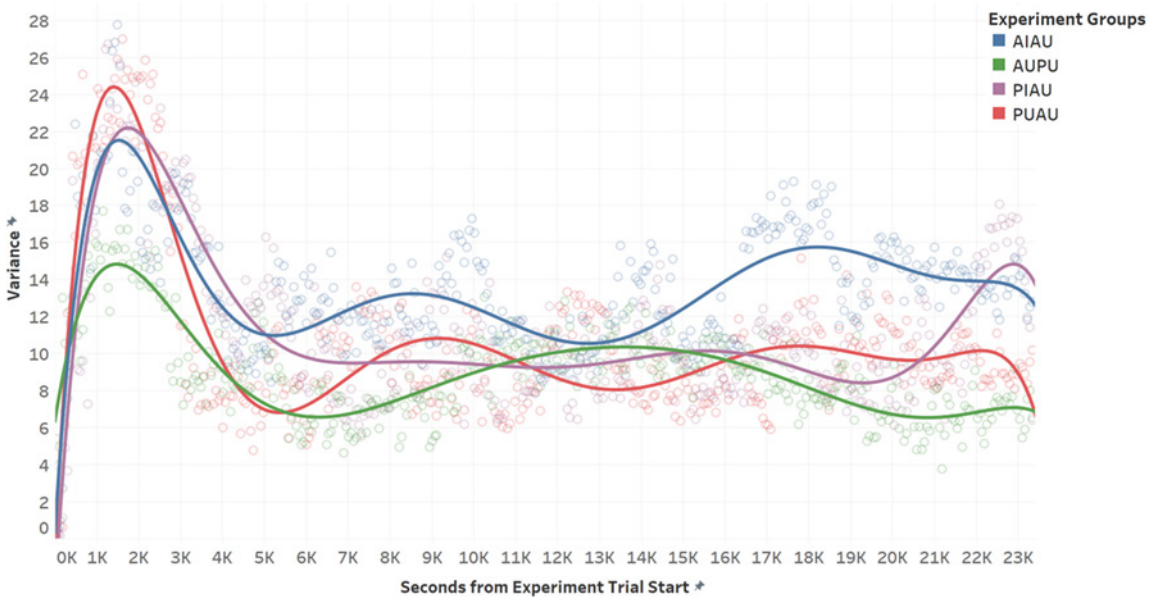


Figure 56. Day 2 Real System Variance Per Minute

Figure 57 depicts subjects sending more than 2 packets for each one-minute interval by each group from the start to the end of the Day 2 experiment trials. Figure 58 depicts subjects sending more than 7 packets for each one-minute interval by each group from the start to the end of the Day 2 experiment trials. Figure 59 depicts subjects sending more than 33 packets for each one-minute interval by each group from the start to the end of the Day 2 experiment trials. Figure 60 depicts subjects sending more than 54 packets for each one-minute interval by each group from the start to the end of the Day 2 experiment trials. Figure 61 depicts subjects sending more than 148 packets for each one-minute interval by each group from the start to the end of the Day 2 experiment trials. Figure 62 depicts subjects sending more than 403 packets for each one-minute interval by each group from the start to the end of the Day 2 experiment trials.

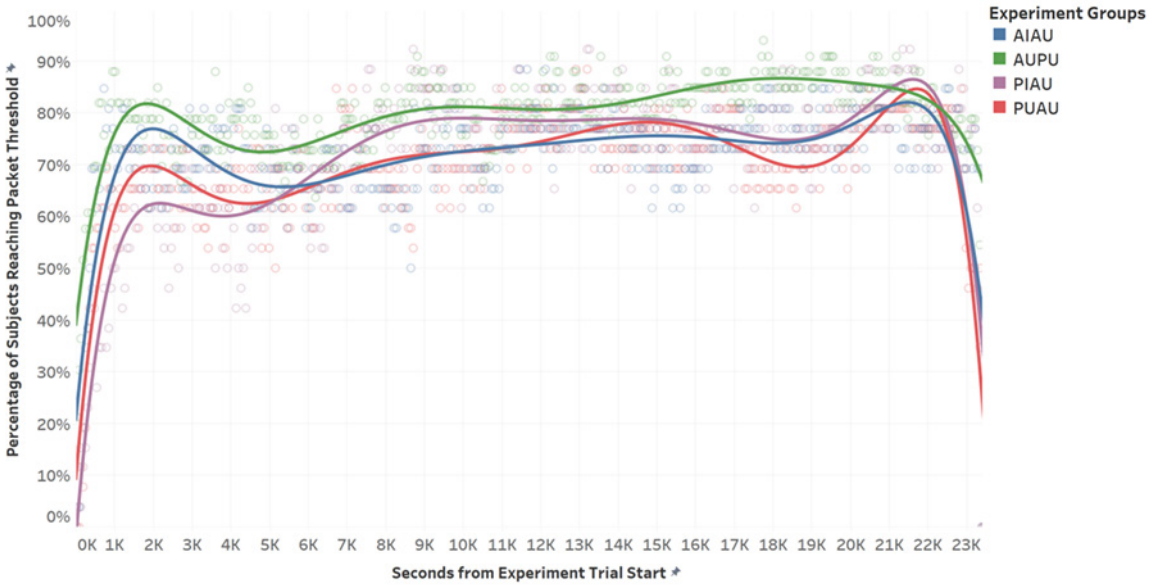


Figure 57. Day 2 Real System Activity Threshold – 2 Packets Per Minute

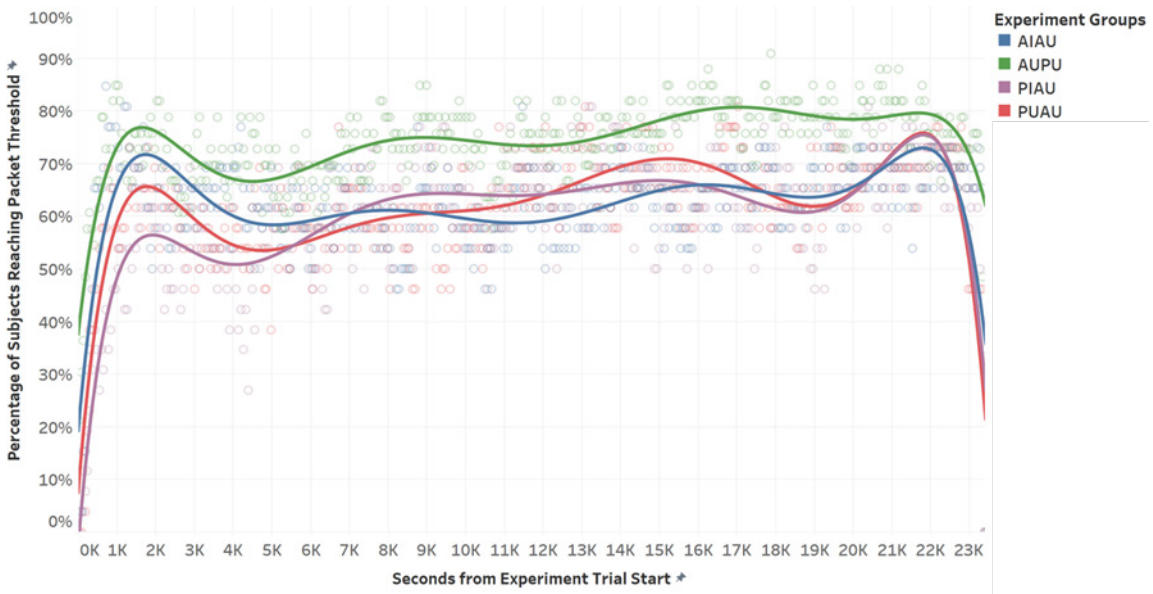


Figure 58. Day 2 Real System Activity Threshold – 7 Packets Per Minute

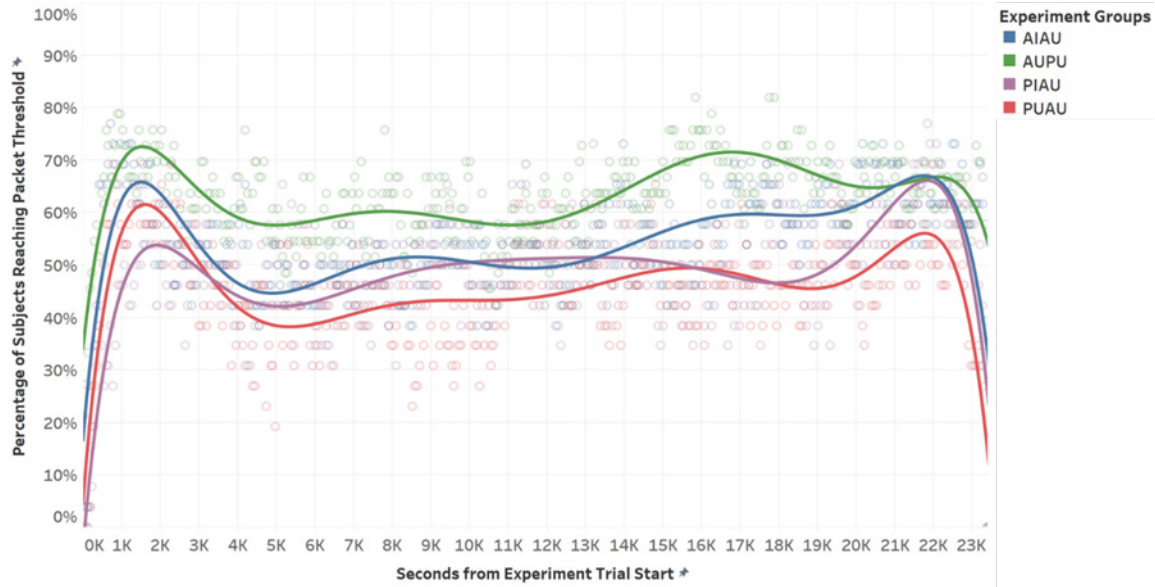


Figure 59. Day 2 Real System Activity Threshold – 20 Packets Per Minute

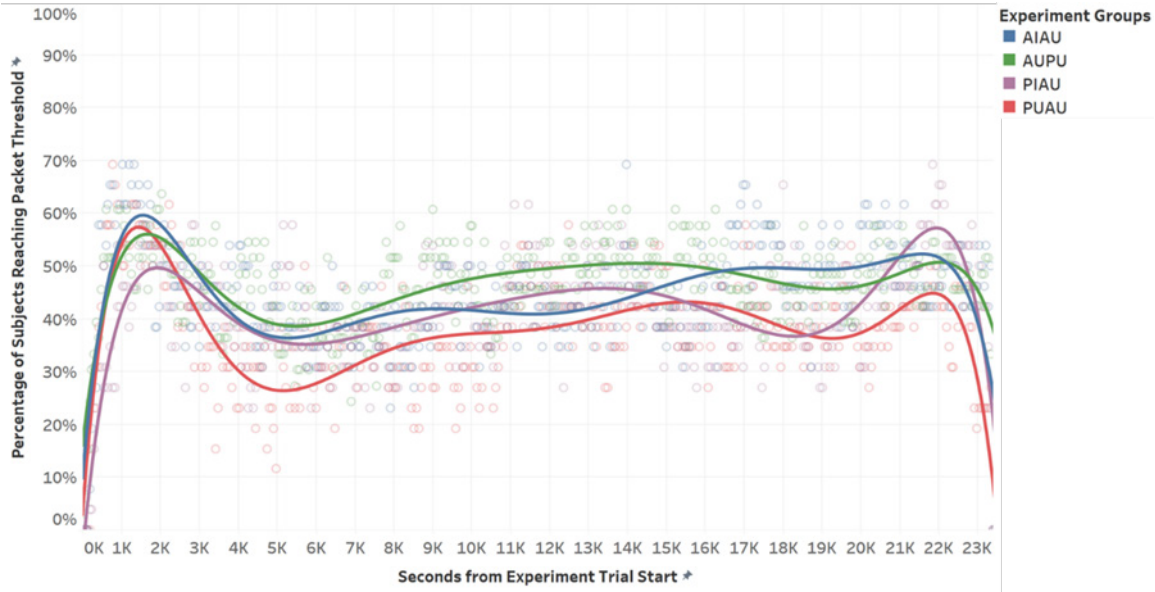


Figure 60. Day 2 Real System Activity Threshold – 54 Packets Per Minute

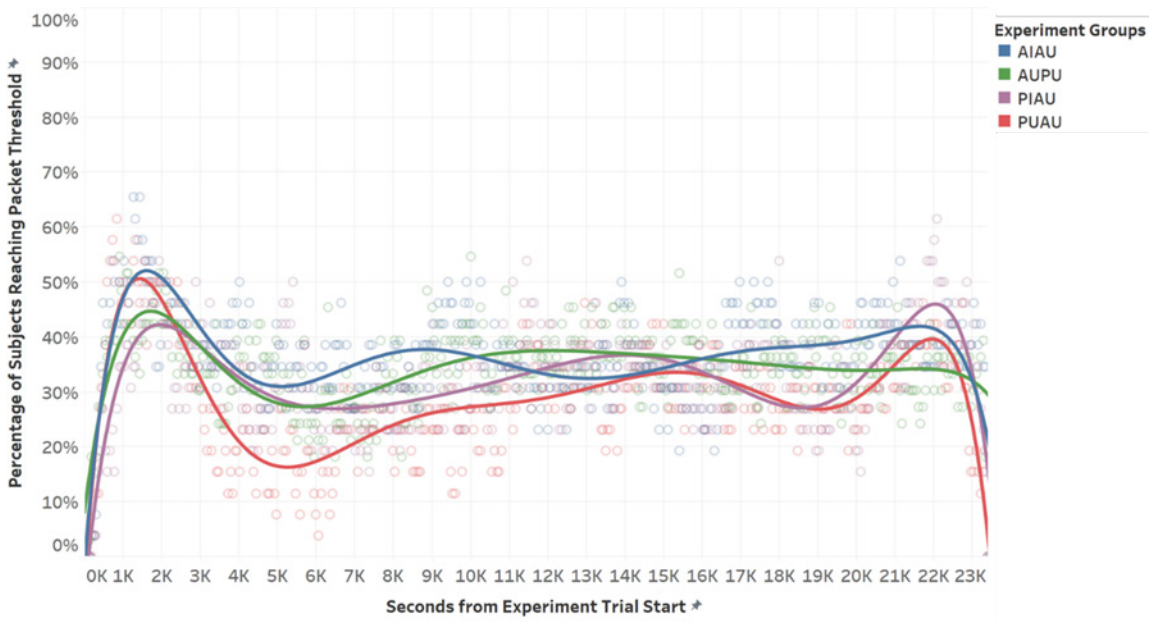


Figure 61. Day 2 Real System Activity Threshold – 148 Packets Per Minute

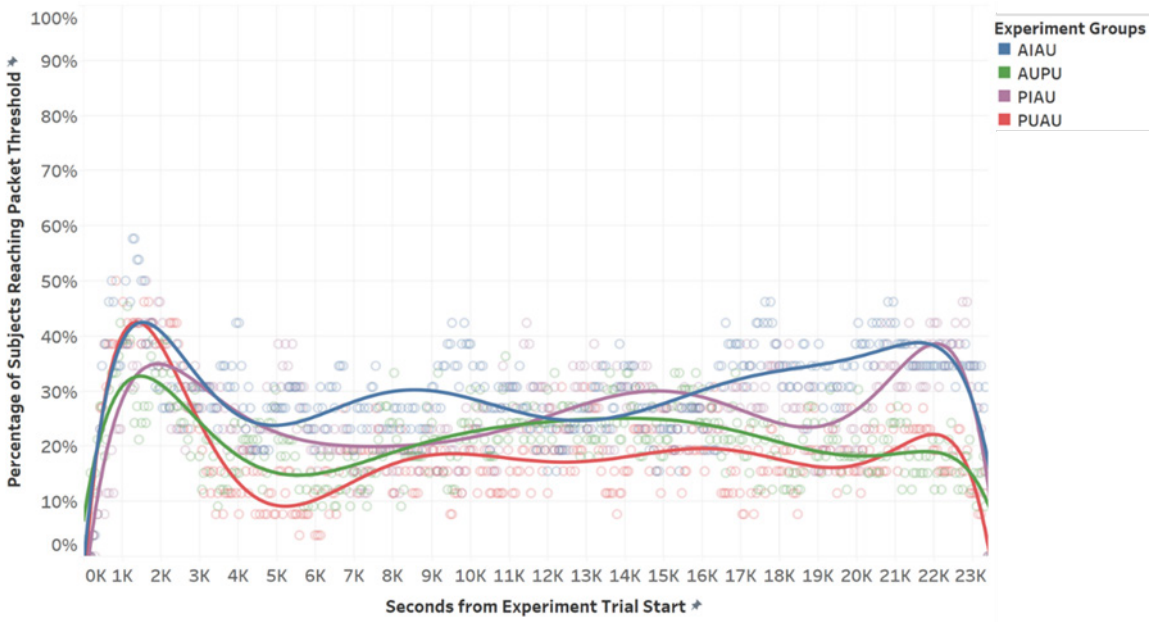


Figure 62. Day 2 Real System Activity Threshold – 403 Packets Per Minute

Figure 63 depicts the combined real system interaction percentages over the duration of the Day 2 experiment trials for all groups.

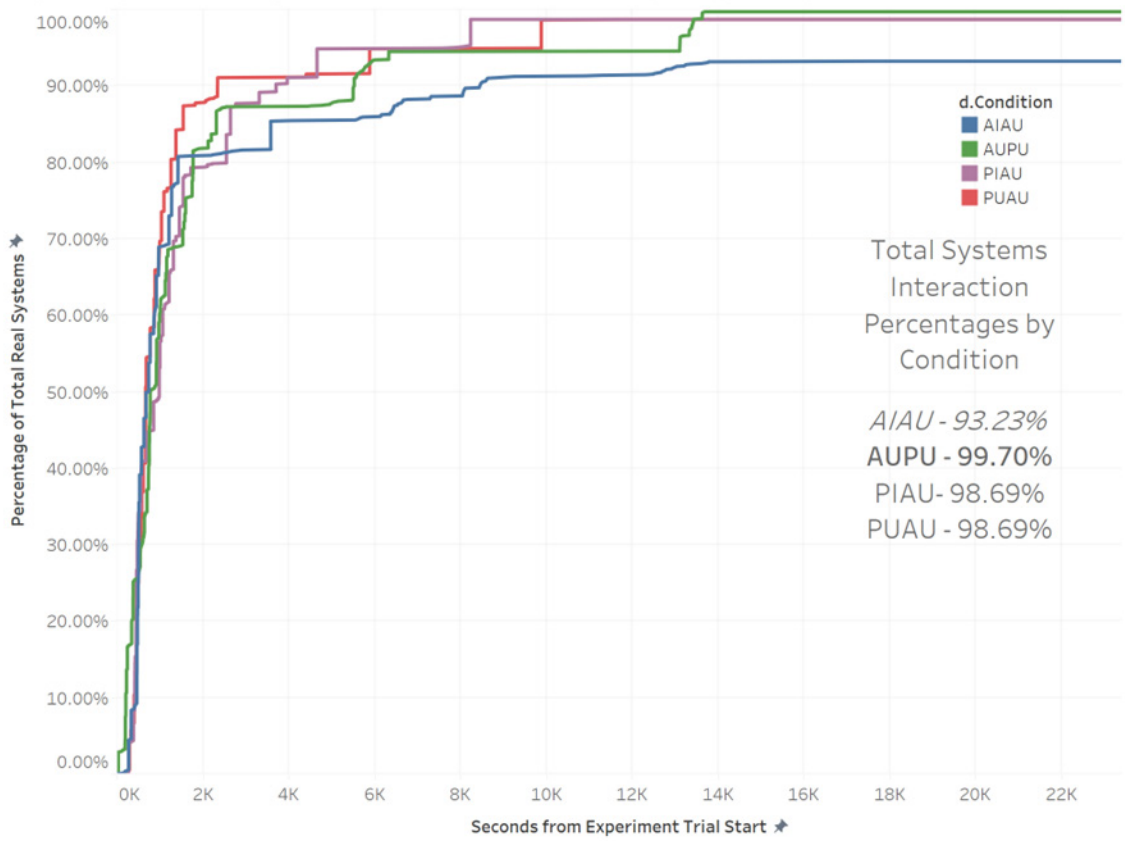


Figure 63. Day 2 Real System Elapsed System Interaction Percentages.

For system interactions depicted in Figure 63, two outliers for the AIAU group should be noted. The AIAU group had one subject (S107) who interacted with only 4 of 50 real systems and another subject (S64) who interacted with only 25 of 50 real systems. In contrast, 40 of 50 real systems was the lowest interaction number for the other three condition groups, each having a subject interaction average of 49 of 50 real systems. Removing these two outliers from the AIAU group results in a Total System Interaction Percentage of 98.58%, in line with the other Day 2 groups.

2. Decoy System Interactions for Cyber Deception Group (AUPU)

This sub-section provides the statistical analysis for decoy system interactions for the Absent-Uninformed/Present-Uninformed group during Day 2 experiment trials. Table 70 provides the descriptive statistics for the latency variable for decoy system interactions for this group. Table 71 provides the descriptive statistics for the frequency variable for

decoy system interactions for this group. Table 72 provides the descriptive statistics for the duration variable for decoy system interactions for this group.

Table 70. Day 2 Decoy System Latency Descriptive Statistics

Value	AUPU
Subjects	33
Observations	1,554
Minimum Value	27
Maximum Value	17,548
Median	1,119
Mean	3,105
Standard Deviation (Sample)	4,309
Variance (Sample)	1.857×10^{7}
Skewness	1.854
Kurtosis	2.065
Coefficient of Variability	1.388

Table 71. Day 2 Decoy System Frequency Descriptive Statistics

Value	AUPU
Subjects	33
Observations	1,554
Minimum Value	1
Maximum Value	2,487,000
Median	2,073
Mean	28,944
Standard Deviation (Sample)	97,365
Variance (Sample)	9.48×10^{9}
Skewness	14.69
Kurtosis	312.23
Coefficient of Variability	3.364

Table 72. Day 2 Decoy System Duration Descriptive Statistics

Value	AUPU
Subjects	33
Observations	1,554
Minimum Value	1
Maximum Value	22,525
Median	13,698

Value	AUPU
Mean	12,770
Standard Deviation (Sample)	7,270
Variance (Sample)	5.28×10^{7}
Skewness	-0.367
Kurtosis	-1.21
Coefficient of Variability	0.569

E. DAY 1 AND DAY 2 EXPERIMENT DATA COMPARISON ANALYSIS

1. Real System Interactions for All Groups (AI, AU, PI, PU and AIAU, AUPU, PIAU, PUAU)

Timelines of aggregate subject interaction with real systems on the cyber range network based on the percentage of systems interacted with over time were generated for Day 1 and Day 2 were combined to provide insight into the impact of varying conditions of cyber deception and psychological deception over time on attacker behavior between groups between Day 1 and Day 2 trials. Figure 64 depicts the real system interaction percentages over the duration of both Day 1 and Day 2 experiment trials for all groups.

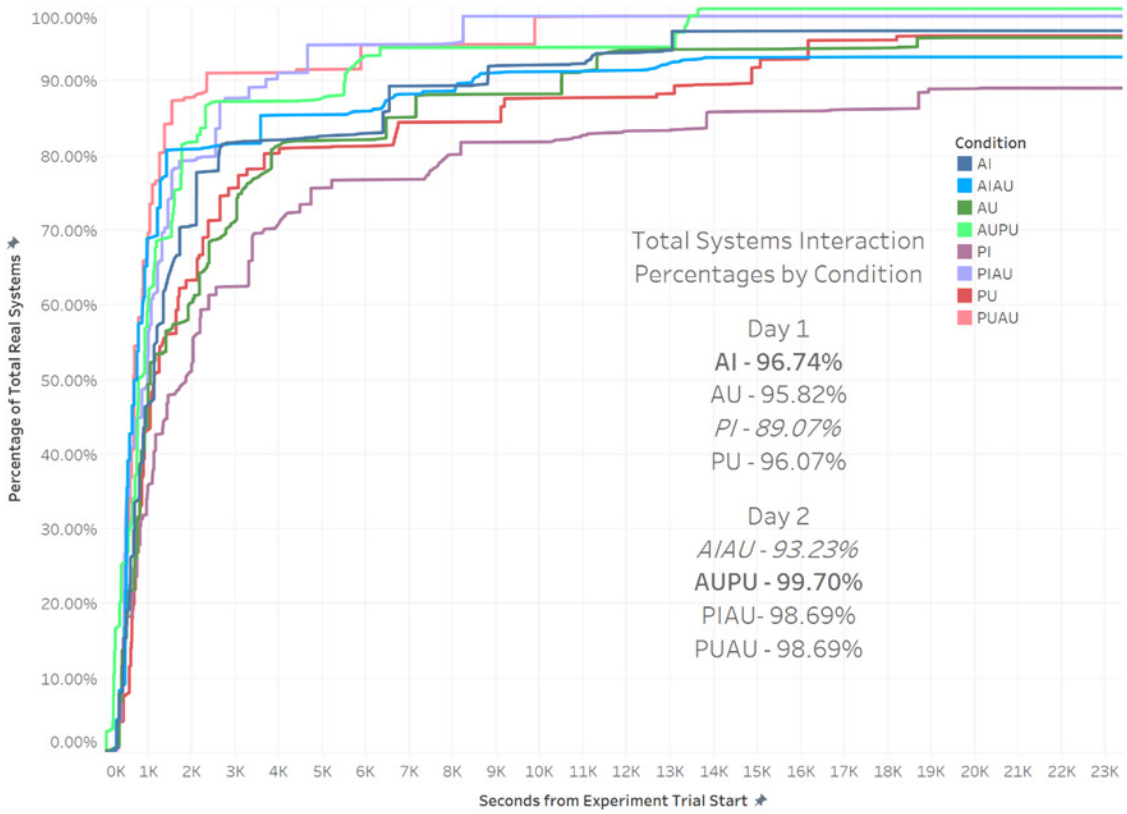


Figure 64. Day 1 and 2 Elapsed Real Systems Interaction Percentages

2. Decoy System Interactions for Cyber Deception Groups (PI, PU and AUPU)

Figure 65 depicts the decoy system interaction percentages over the duration of both Day 1 and Day 2 experiment trials for groups where cyber deception was present.

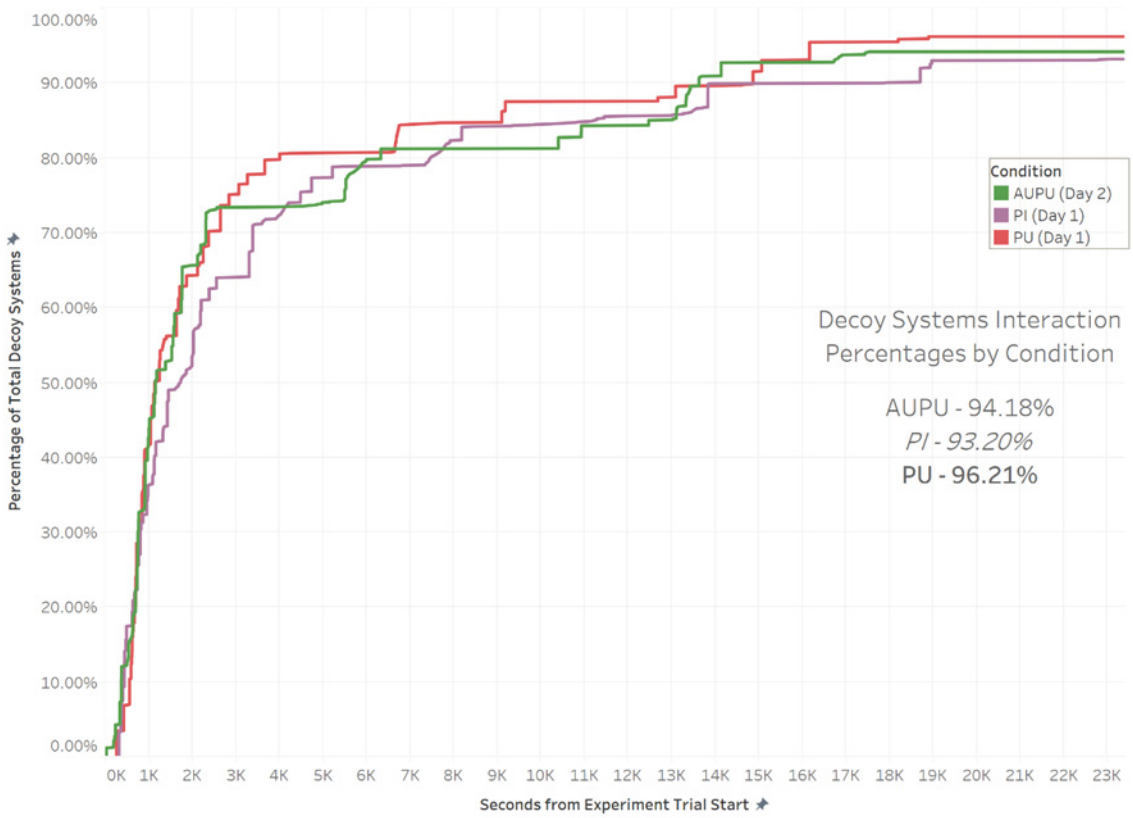


Figure 65. Day 1 and 2 Elapsed Decoy Systems Interaction Percentages

THIS PAGE INTENTIONALLY LEFT BLANK

VI. FACTOR ANALYSIS

Given the nascent state of cyber deception theory, PCA and EFA were used to explore correlations and relationships within the Tularosa study data. Three sets of variables were included in the analysis, temporal variables, cyber task questionnaire responses and self-reported experience levels. Three temporal variables included were the median latency, frequency and duration of subject performance captured for interactions with real systems during Day 1 experiment trials. Media latency values varied from 260 to 16,297 seconds, median frequency values varied from 161 to 1,201,900, and median duration values varied from 732 to 23,109. Cyber task questionnaire responses to the level of confusion, self-doubt, confidence, surprise, and frustration were captured in a post-event survey. These values were recorded from 1 to 5, with 1 being the lowest rate and 5 being the highest rate depending on the question answered. Self-reported experience levels across eight areas captured in a pre-event survey were included: Cyber security, Network penetration, Host penetration, Network reconnaissance, Incidence response, Generalized defense practice, Network protocol reverse engineering and Binary reverse engineering. The values were recorded on a range from 1 to 5, with 1 being novice and 5 being expert.

A. PRINCIPAL COMPONENT ANALYSIS

PCA was conducted using JMP® Pro Version 16.0.0 with "Robust" Estimation method across 5 groupings, individual AI, AU, PI and PU experiment groups and a combined group of all. Using Scree Plot analysis and number of Eigenvalues greater than 1, five components were selected for retention. PCA explores the relationships between variables by focusing on total variance of observed variables to transform these variables into a reduced set of uncorrelated variables while retaining as much information from the original variable set as possible (Dunteman, 1989). While primarily a data dimensionality reduction technique, PCA also helps explore the structure of the underlying data by identifying influential variables and variable clusters for further understanding of the phenomenon being studied (Dunteman, 1989).

1. Day 1 Analysis

a. Absent-Informed

For PCA conducted using data from the Day 1 Absent-Informed group, the first principal component is composed of entirely self-reported experience levels with loadings in excess of 0.50 for all. The highest loading was for Network penetration (0.89), Cyber security (0.85), Incidence response (0.83), Network reconnaissance (0.82), Host Penetration (0.81), Binary reverse engineering (0.69), and Network protocol reverse engineering (0.67). This principal component is termed "Technical Competence" to identify the key role that experience in fields related to cyber reconnaissance has in accounting for total variance observed. The second principal component is composed of technical performance measures of Frequency (0.70) and Duration (0.51) and self-reported cyber task questionnaire responses for Confusion (0.66), Self-Doubt (0.82), Confidence (-0.76), and Frustration (0.72). Confidence is negatively correlated with the other measures in this component. This principal component is termed "Uncertainty" to account for the negative correlation between both Frequency and Duration and Confidence. The third principal component is composed of loadings for Latency (0.63), Duration (-0.60), Confusion (0.52), Surprise (0.55), and Network protocol reverse engineering (0.50). Duration and Latency are negatively correlated, with positive correlation between Latency, Confusion, Surprise and Network protocol reverse engineering. This principal component is termed "Caution" to account for these loadings. The four principal component has loadings for Latency (0.54), Surprise (-0.59) and Binary reverse engineering (0.50). Surprise has a negative correlation accounting for variance compared with Latency and Binary reverse engineering. This component is termed "Prudence" in acknowledgement that delayed system interaction (Latency) is potentially impacted by additional time spent analyzing network data previously obtained. Finally, the fifth component is comprised of negative loadings for Frequency (-0.48) and Confusion (-0.30) and positive loadings for Surprise (0.39) and frustration (0.49).This component is termed "Trepidation" from the combination of surprise and frustration potentially creating anxiety for the attacker. The Loading Matrix for loadings greater than 0.30 is provided in Table 73 with the full loading

matrix available in Appendix C. A visual depiction of the Eigenvalues, Score Plot and Loading Plot are provided in Figure 66 with the Scree Plot provided in Figure 67.

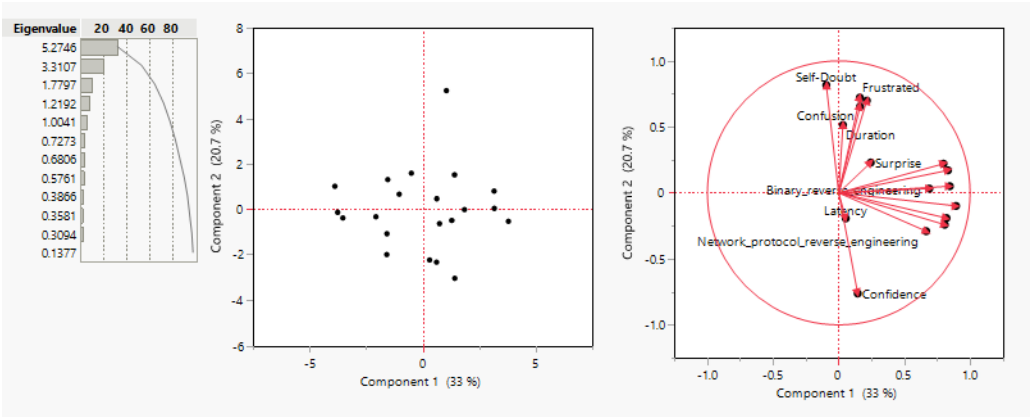


Figure 66. Day 1 Absent-Informed PCA Summary Plots

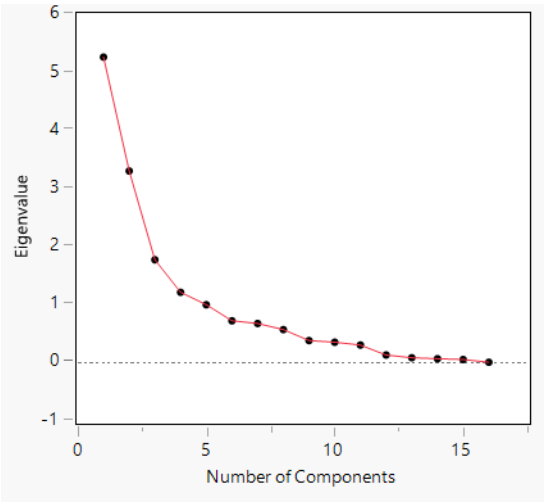


Figure 67. Day 1 Absent-Informed PCA Scree Plot

Table 73. Day 1 Absent-Informed PCA

Variable	**Principal 1** Technical Competence	**Principal 2** Uncertainty	**Principal 3** Caution	**Principal 4** Prudence	**Principal 5** Trepidation
Latency			0.63	0.54	
Frequency		0.70			-0.48
Duration		0.51	-0.60		

Variable	Principal 1 Technical Competence	Principal 2 Uncertainty	Principal 3 Caution	Principal 4 Prudence	Principal 5 Trepidation
Confusion		0.66	0.52		-0.30
Self-Doubt		0.82			
Confidence		-0.76			
Surprise			0.55	-0.59	0.39
Frustrated		0.72			0.49
Cyber_security	0.85				
Network_penetration	0.89				
Host_penetration	0.81				
Network_reconnaissance	0.82				
Incidence_response	0.83				
Generalized_defense_practice	0.80				
Network_protocol_reverse_engineering	0.67		0.50		
Binary_reverse_engineering	0.69			0.50	
Percentage of Total Variance Explained	33.0%	20.7%	11.1%	7.6%	4.5%

b. Absent-Uninformed

For PCA conducted on data from the Day 1 Absent-Informed group, the first principal component is composed self-reported experience levels, negative loadings for Confusion (-0.35) and Self-Doubt (-0.49) along with a positive loading for Confidence (0.62), which account for 34% of total variance observed in the data. This principal component is also termed "Technical Competence" to identify the key role of experience in accounting for total observed variance along with the correlation of this experience with Confidence and negative correlation with Confusion and Self-Doubt. Component 2 is comprised of loadings from all variable groupings (technical performance, emotional state, and self-reported experience levels. Unique to the Absent-Informed condition, the temporal variables of latency (-0.67), frequency (0.66), and duration (0.41) are present in the second component. This principal component is titled "Probing" on account of the positive loadings for frequency and duration and the negative loading for latency. Principle component three is comprised of positive loadings from Confusion (0.72), Surprise (0.44) and Frustrated (0.68) and negative loading on Duration (-0.69) which could be related to caution being exercised by subjects in this group. Component Four is titled "Alertness" on account of positive loadings on Self-Doubt (0.54) and Surprise (0.66) with negative loading on Frustrated (-0.35). Component Five again sees loadings on all three temporal variables, with negative loadings on Duration (-0.35) and Confusion (-0.31) and positive loading on

Self-Doubt (0.34), Latency (0.50), and Frequency (0.53) resulting in a label of "Trepidation." The Loading Matrix for loadings greater than 0.30 is provided in Table 74 with the full loading matrix available in Appendix C. A visual depiction of the Eigenvalues, Score Plot and Loading Plot are provided in Figure 68 with the Scree Plot provided in Figure 69.

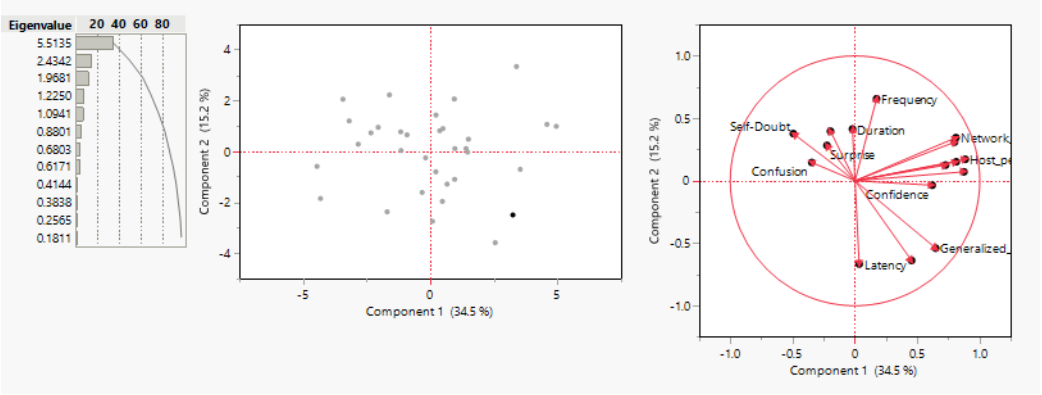


Figure 68. Day 1 Absent-Uninformed PCA Summary Plots

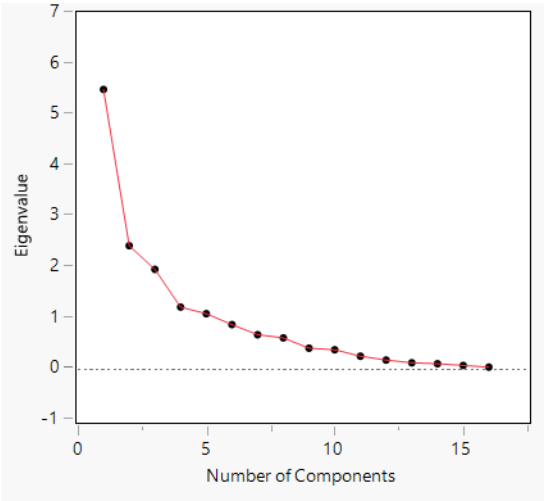


Figure 69. Day 1 Absent-Uninformed PCA Scree Plot

Table 74. Day 1 Absent-Uninformed PCA

	Principal 1 Technical Competence	**Principal 2** Probing	**Principal 3** Caution	**Principal 4** Alertness	**Principal 5** Trepidation
Latency		-0.67			0.50
Frequency		0.66			0.53
Duration		0.41	-0.69		-0.35
Confusion	-0.35		0.72		-0.31
Self-Doubt	-0.49	0.38		0.54	0.34
Confidence	0.62				
Surprise			0.44	0.66	
Frustrated		0.40	0.68	-0.35	
Cyber_security	0.81			0.36	
Network_penetration	0.81	0.35			
Host_penetration	0.88				
Network_reconnaissance	0.87				
Incidence_response	0.45	-0.64			
Generalized_defense_practice	0.65	-0.54		0.31	
Network_protocol_reverse_engineering	0.73				
Binary_reverse_engineering	0.80	0.31			
Percentage of Total Variance Explained	34.5%	15.2%	12.3%	7.7%	6.8%

c. Present-Informed

For PCA conducted on data from the Day 1 Present-Informed group, the first principal component is composed primarily of self-reported experience levels with additional loadings from Confidence (0.35) and Duration (-0.32). The loadings for self-reported experience levels ranged from a high value of 0.87 for Network penetration to a low value of 0.52 for Generalized defense practice. This principal component is also termed "Technical Competence" to identify the key role of experience in accounting for total observed variance along with the correlation of this experience with Confidence. The second principal component is composed of the technical performance measure of Frequency (0.40) and responses for Confusion (0.64), Self-Doubt (0.77), Confidence (-0.48), Surprise (0.70) and Frustrated (0.81). Confidence has a negative loading compared with the other measures in this component. This principal component is termed "Uncertainty" to account for this negative loading. The third principal component is titled "Probing" on account of the positive loadings for Frequency (0.61), Duration (0.72) and Confusion (0.38) and a negative loading for Latency (-0.50) and Self-Doubt (-0.31). The fourth principal component is titled "Investigation" to account for the positive loadings for

Frequency (0.57), Confidence (0.38), Cyber security (0.38), Incidence response (0.49), and Generalized Defense Practice (0.36), and negative loading for Host penetration (-0.41). The fifth component is a combination of Latency (0.64), Confusion (0.44), Confidence (0.37) and Network Protocol Reverse Engineering (0.31). This complicated combination is titled "Trepidation" to account for these disparate loadings. The Loading Matrix for loadings greater than 0.30 is provided in Table 75 with the full loading matrix available in Appendix C. A visual depiction of the Eigenvalues, Score Plot and Loading Plot are provided in Figure 70 with the Scree Plot provided in Figure 71.

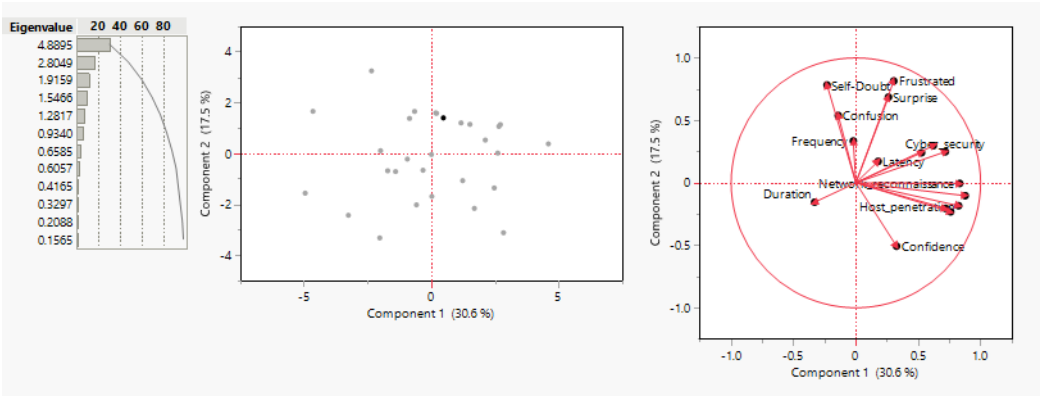


Figure 70. Day 1 Present-Informed PCA Summary Plots

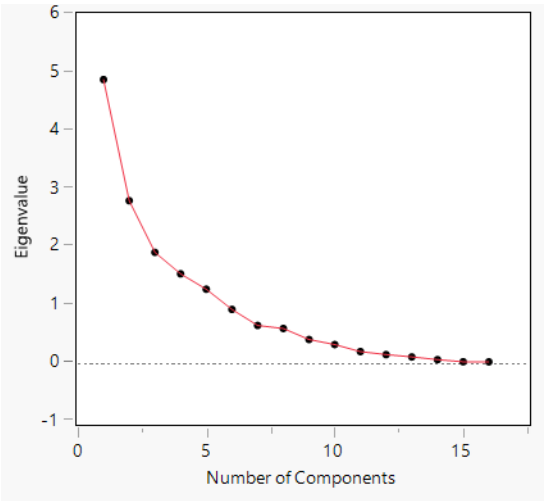


Figure 71. Day 1 Present-Informed PCA Scree Plot

Table 75. Day 1 Present-Informed PCA

	Principal 1 Technical Competence	**Principal 2** Uncertainty	**Principal 3** Probing	**Principal 4** Investigation	**Principal 5** Trepidation
Latency			-0.50		0.64
Frequency		0.40	0.61	0.57	
Duration	-0.32		0.72		
Confusion		0.64	0.38		0.44
Self-Doubt		0.77	-0.31		
Confidence	0.35	-0.48		0.38	0.37
Surprise		0.70			
Frustrated		0.81			
Cyber_security	0.72			0.38	
Network_penetration	0.87				
Host_penetration	0.84			-0.41	
Network_reconnaissance	0.83				
Incidence_response	0.66			0.49	
Generalized_defense_practice	0.52		-0.33	0.36	
Network_protocol_reverse_engineering	0.79				0.31
Binary_reverse_engineering	0.74		0.41		
Percentage of Total Variance Explained	30.6%	17.5%	12.0%	9.7%	8.0%

d. Present-Uninformed

For PCA conducted on data from the Day 1 Present-Uninformed group, the first principal component is composed of all variables examined apart from Frequency, Confusion and Self-Doubt, with a high absolute loading value of 0.87 for both Host penetration and Network reconnaissance and low absolute loading value of -0.35 for Frustrated. All loadings are positive, with the exception of Surprise (-0.42), Frustrated (-0.35), and Latency (-0.40) which may suggest limited apprehension on behalf of the subjects in this condition group. This principal component is termed "Technical Competence" to identify the key role that experience in fields related to cyber reconnaissance has in accounting for total variance observed. The second component is labelled "Uncertainty" to account for the positive loadings for all Cyber Task Questionnaire scores except for Confidence (-0.32), which has a negative loading. The third principal component is titled "Apprehension" on account of the positive loadings for Latency (0.72), Self-Doubt (0.38) and Confusion (0.48), and negative loadings for Duration (-0.32) and Surprise (-0.47). The fourth principal component is titled "Probing" on account of the positive loadings for Frequency (0.73) and negative loadings for Network

protocol reverse engineering (-0.58) and Binary reverse engineering (-0.42). Finally, component five is labelled "Trepidation" to account for the negative loading for Duration (-0.41) and positive loading for Surprise (0.35), Incidence response (0.57) and Generalized defense practice (0.57). The Loading Matrix for loadings greater than 0.30 is provided in Table 76 with the full loading matrix available in Appendix C. A visual depiction of the Eigenvalues, Score Plot and Loading Plot are provided in Figure 72 with the Scree Plot provided in Figure 73.

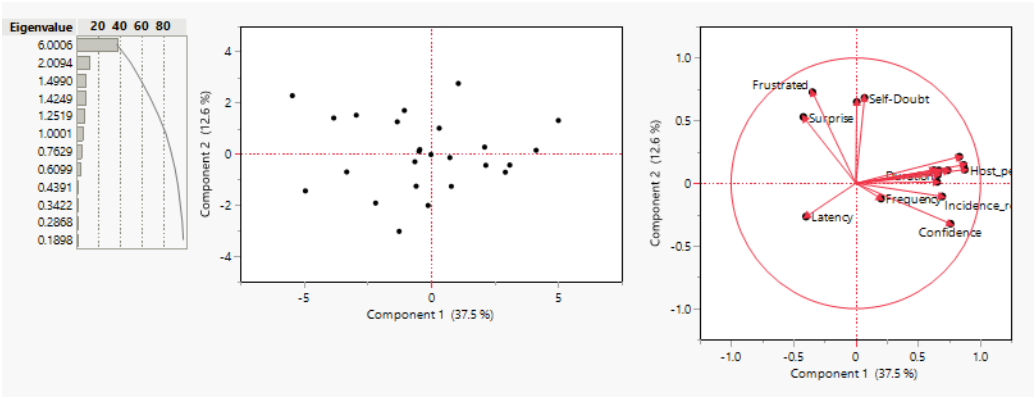


Figure 72. Day 1 Present-Uninformed PCA Summary Plots

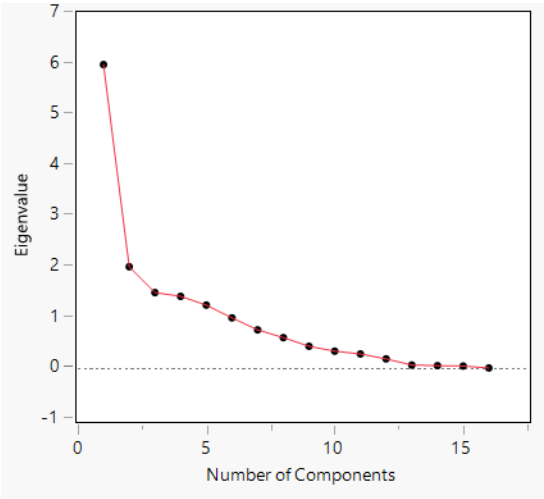


Figure 73. Day 1 Present-Uninformed PCA Scree Plot

Table 76. Day 1 Present-Uninformed PCA

	Principal 1 Technical Competence	**Principal 2** Uncertainty	**Principal 3** Apprehension	**Principal 4** Probing	**Principal 5** Trepidation
Latency	-0.40		0.72		
Frequency				0.73	
Duration	0.66		-0.32		-0.41
Confusion		0.65	0.48		
Self-Doubt		0.68	0.38		
Confidence	0.76	-0.32			
Surprise	-0.42	0.53	-0.47		0.35
Frustrated	-0.35	0.73			
Cyber_security	0.63		0.34	0.33	
Network_penetration	0.83				
Host_penetration	0.87				
Network_reconnaissance	0.87				
Incidence_response	0.69				0.57
Generalized_defense_practice	0.65				0.57
Network_protocol_reverse_engineering	0.67			-0.58	
Binary_reverse_engineering	0.74			-0.42	
Percentage of Total Variance Explained	37.5%	12.6%	9.4%	8.9%	7.8%

e. Combined Absent-Informed, Absent-Uninformed, Present-Informed and Present-Uninformed

PCA was applied to the technical performance, Cyber TSQ scores and self-reported experience levels for subjects across all experimental conditions for Day 1. The Loading Matrix for loadings greater than 0.30 is provided in Table 77 with the full loading matrix available in Appendix C. A visual depiction of the Eigenvalues, Score Plot and Loading Plot are provided in Figure 74 with the Scree Plot provided in Figure 75. This information is provided for comparison purposes without additional analysis.

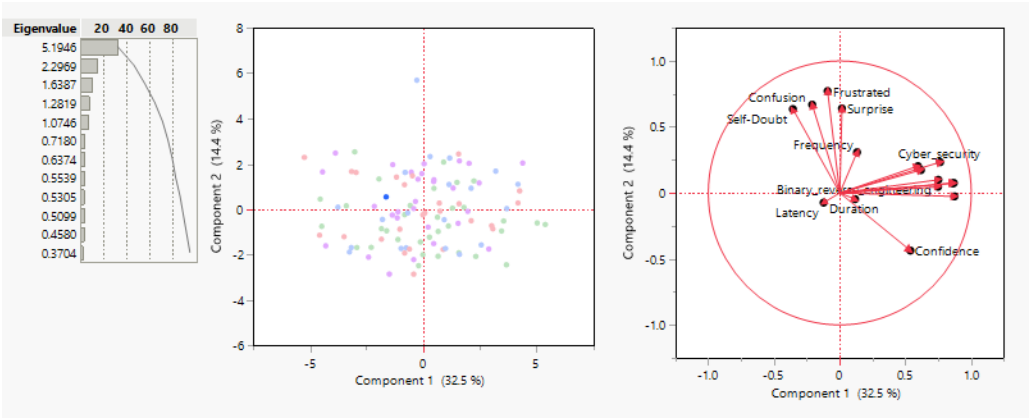


Figure 74. Day 1 Combined PCA Summary Plot

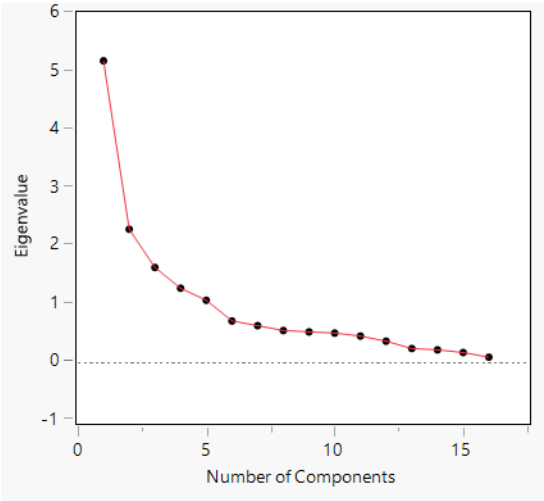


Figure 75. Day 1 Combined PCA Scree Plot

Table 77. Day 1 Combined PCA

	Prin1	Prin2	Prin3	Prin4	Prin5
Latency			0.69		0.48
Frequency		0.31	-0.48	0.38	0.55
Duration			-0.80		
Confusion		0.67			
Self-Doubt	-0.36	0.63			
Confidence	0.54	-0.44			
Surprise		0.64			-0.52
Frustrated		0.77			
Cyber_security	0.76				
Network_penetration	0.87				
Host_penetration	0.87				
Network_reconnaissance	0.86				

	Prin1	Prin2	Prin3	Prin4	Prin5
Incidence_response	0.59		0.32	0.53	
Generalized_defense_practice	0.62		0.38	0.51	
Network_protocol_reverse_engineering	0.75			-0.34	
Binary_reverse_engineering	0.75			-0.32	0.31
Percentage of Total Variance Explained	32.5%	14.4%	10.2%	8.0%	6.7%

2. Day 2 Analysis

a. Absent-Informed/Absent-Uninformed (AIAU)

For PCA conducted on data from the Day 2 AIAU group, the first principal component is composed of all self-reported experience fields with Network penetration having the highest value of 0.92 and Network protocol reverse engineering the lowest value of 0.60 resulting in the component being titled "Technical Competence." The component loadings for the Day 2 AIAU group are very similar to the Day 1 AI group with slight decreases in loading values. Component two loadings include negative loadings for Duration (-0.32) and Confidence (-0.37) along with positive loadings for Confusion (0.91), Self-Doubt (0.79), Surprise (0.70), and Frustrated (0.63) which suggests a degree of "Uncertainty," which provides the name for this component. The third component incudes loadings from all three variable groups, with negative loadings for Frequency (-0.79), Self-Doubt (-0.36), Incidence response (-0.34), and Generalized defense practice (-0.48) and positive loadings for Confidence (0.67), Surprise (0.44), Host Penetration (0.33), and Network protocol reverse engineering. The combination of negative and positive loadings result in the title "Caution" for this component. Component four is also comprised of variables across all three groups (technical performance, Cyber TSQ scores, and self-reported experience) with negative loadings for Latency (-0.70), Host penetration (-0.33), and Network reconnaissance (-0.39) along with positive loadings for Duration (0.41), Confidence (0.37), Frustrated (0.46) and Binary reverse engineering (0.37). This collection of variable loadings is titled "Assurance." Finally, component five is composed of a positive loading for Duration (0.74) and a negative loading for Network protocol reverse engineering (-0.58), resulting in a title of "Exploration." The Loading Matrix for loadings greater than 0.30 is provided in Table 78 with the full loading matrix available in

Appendix C. A visual depiction of the Eigenvalues, Score Plot and Loading Plot are provided in Figure 76 with the Scree Plot provided in Figure 77.

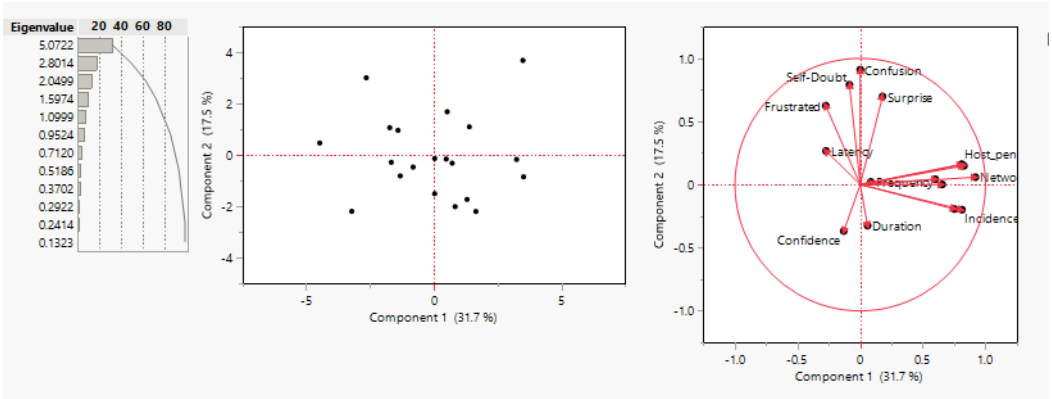


Figure 76. Day 2 Absent-Informed/Absent-Uninformed PCA Summary Plots

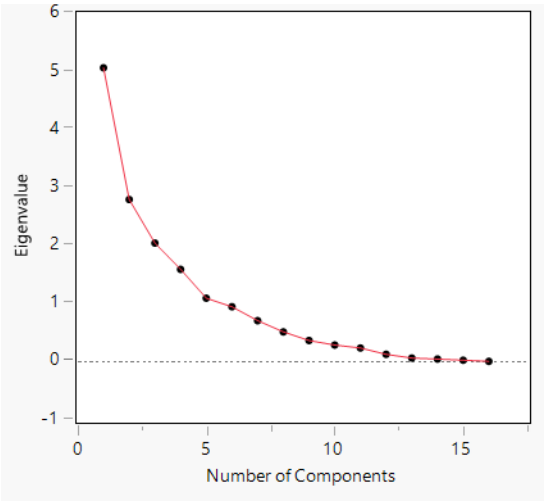


Figure 77. Day 2 Absent-Informed/Absent-Uninformed PCA Scree Plot

Table 78. Day 2 Absent-Informed/Absent-Uninformed PCA

Variable	**Principal 1** Technical Competence	**Principal 2** Uncertainty	**Principal 3** Caution	**Principal 4** Assurance	**Principal 5** Exploration
Latency				-0.70	
Frequency			-0.79		
Duration		-0.32		0.41	0.74
Confusion		0.91			
Self-Doubt		0.79	-0.36		

Variable	Principal 1 Technical Competence	Principal 2 Uncertainty	Principal 3 Caution	Principal 4 Assurance	Principal 5 Exploration
Confidence		-0.37	0.67	0.37	
Surprise		0.70	0.44		
Frustrated		0.63		0.46	
Cyber_security	0.83				
Network_penetration	0.92				
Host_penetration	0.81		0.33	-0.33	
Network_reconnaissance	0.81			-0.39	
Incidence_response	0.81		-0.34		
Generalized_defense_practice	0.75		-0.48		
Network_protocol_reverse_engineering	0.60		0.31		-0.58
Binary_reverse_engineering	0.65			0.37	
Percentage of Total Variance Explained	31.7%	17.5%	12.8%	10.0%	6.9%

b. Absent-Uninformed/Present-Uninformed

For PCA conducted on the Day 2 AUPU group, the first principal component contains loadings for all self-reported experience levels and three of five Cyber TSQ scores, omitting Confusion and Confidence. The loadings for the self-reported experience levels are all positive, while all loadings for the Cyber TSQ scores are negative. This component is also titled "Technical Competence" to reflect these loadings. Comparison this with Day 1 AU group PCA results, similar loadings are found for self-reported experience levels, but the Day 1 AU group lacked loadings greater than 0.30 for Surprise and Frustrate while the Day 2 AUPU group lacked loadings greater than 0.30 for Confidence and Surprise. Component two is composed of negative loadings for Latency (-0.59) and positive loadings for Duration (0.58), Confusion (0.65), Self-Doubt (0.79), and Frustrated (0.70), which suggests "Exploration," the title given to this component. For the third component, negative loadings are present for Confidence (-0.35), Network penetration (-0.31) and Host penetration (-0.31) and positive loadings for Confusion (0.33) and Frustrated (0.43) resulting in the title "Uncertainty" provided to this component. The fourth component has a negative loading for Latency (-0.42) and positive loadings for Confidence (0.69) and Surprise (0.68) suggesting "Probing," the title for this component. Finally, the fifth component is composed of a positive loading for Latency (0.34) and negative loadings for Frequency (-0.85) and Duration (-0.30), suggestive of "Caution," the title given to this component. The Loading Matrix for loadings greater than 0.30 is provided in Table 79 with the full loading

matrix available in Appendix C. A visual depiction of the Eigenvalues, Score Plot and Loading Plot are provided in Figure 78 with the Scree Plot provided in Figure 79.

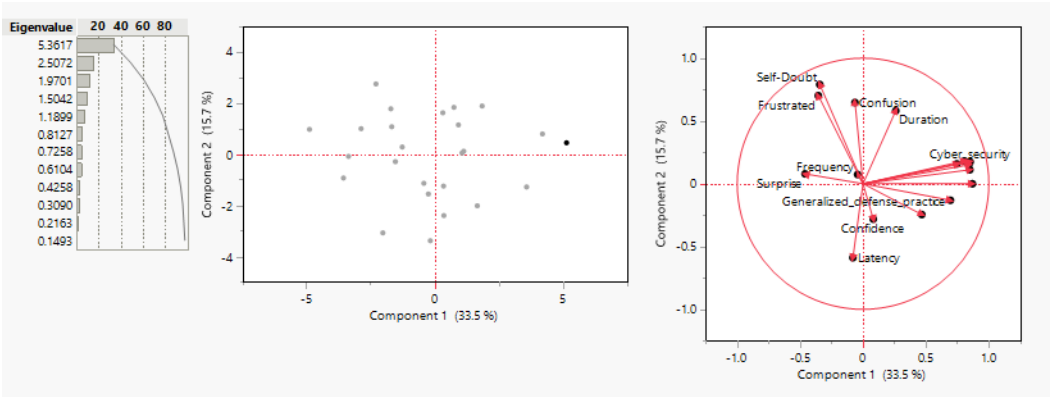


Figure 78. Day 2 Absent-Uninformed/Present-Uninformed PCA Summary Plots

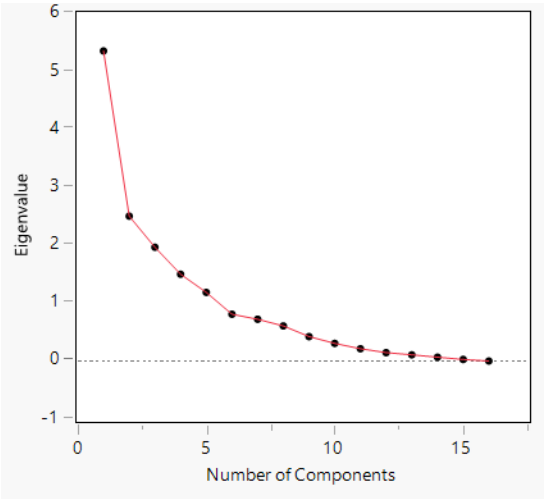


Figure 79. Day 2 Absent-Uninformed/Present-Uninformed PCA Scree Plot

Table 79. Day 2 Absent-Uninformed/Present-Uninformed PCA

	Principal 1 Technical Competence	**Principal 2** Exploration	**Principal 3** Uncertainty	**Principal 4** Probing	**Principal 5** Caution
Latency		-0.59		-0.42	0.34
Frequency					-0.85
Duration		0.58			-0.30
Confusion		0.65	0.33		
Self-Doubt	-0.34	0.79			

	Principal 1 Technical Competence	**Principal 2** Exploration	**Principal 3** Uncertainty	**Principal 4** Probing	**Principal 5** Caution
Confidence			-0.35	0.69	
Surprise	-0.46			0.68	
Frustrated	-0.36	0.70	0.43		
Cyber_security	0.85				
Network_penetration	0.81		-0.31		
Host_penetration	0.85		-0.31		
Network_reconnaissance	0.87				
Incidence_response	0.47		0.74		
Generalized_defense_practice	0.69		0.61		
Network_protocol_reverse_engineering	0.75			-0.33	
Binary_reverse_engineering	0.84				
Percentage of Total Variance Explained	33.5%	15.7%	12.3%	9.4%	7.4%

c. Present-Informed/Absent-Uninformed (PIAU)

The first component of the Day 2 PIAU condition group is composed of nearly all variables across technical performance measures, Cyber TSQ scores and self-reported experience levels. Exceptions are for Frequency, Confusion and Surprise with loadings less than 0.30. Highest loadings are present for technical performance measures leading to this component being titled "Technical Competence," but the breadth of variables included in this component is notable. Component two likewise has loadings across all variable groups, with positive loadings for all variables. With the highest variable loadings present for Confusion (0.90) and Surprise (0.78) this component is titled "Uncertainty." The third component features negative loadings for Duration (-0.42) and Confidence (-0.63) and positive loadings for Latency (0.34), Self-Doubt (0.35), Frustrated (0.57), Network protocol reverse engineering (0.38), and Binary reverse engineering (0.51). These loadings suggest "Caution," the title given to this component. The fourth component is composed of negative loadings for both Confidence (-0.32) and Surprise (-0.31) along with positive loadings for Latency (0.34), Incidence response (0.32) and Generalized defense practice (0.69). These loadings suggest "Trepidation," which is the title given to this component. Finally, component five is composed of all positive loadings for Latency (0.33), Frequency (0.37), Self-Doubt (0.41), and Network reconnaissance (0.33). The combination of these variables is titled "Probing" on account of the technical performance measures and Cyber TSQ scores. The Loading Matrix for loadings greater than 0.30 is provided in Table 80 with the full loading matrix available in Appendix

C. A visual depiction of the Eigenvalues, Score Plot and Loading Plot are provided in Figure 80 with the Scree Plot provided in Figure 81.

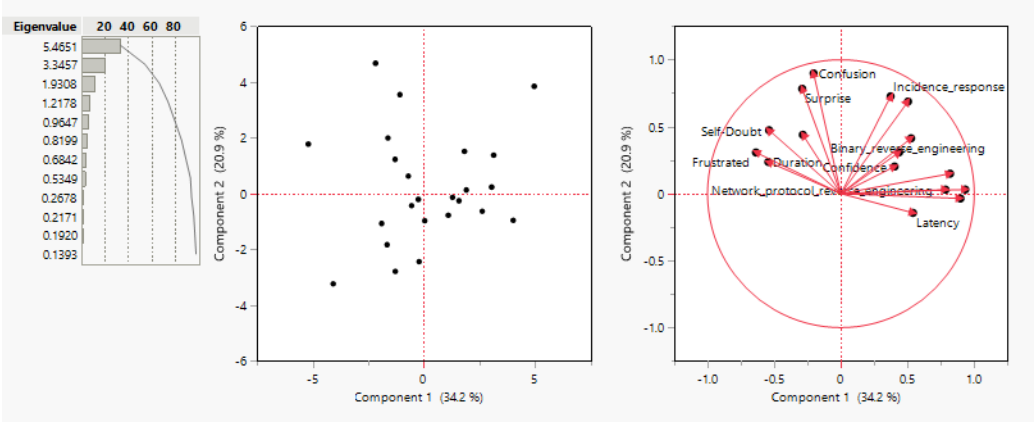


Figure 80. Day 2 Present-Informed/Absent-Uninformed PCA Summary Plots

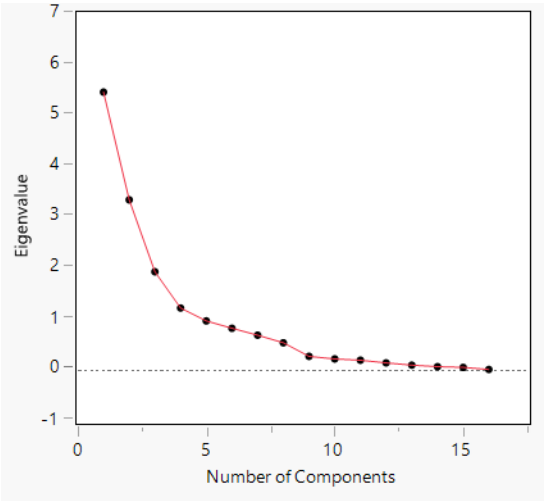


Figure 81. Day 2 Absent-Informed/Absent-Uninformed PCA Scree Plot

Table 80. Day 2 Present-Informed/Absent-Uninformed PCA

	Principal 1 Technical Competence	**Principal 2** Uncertainty	**Principal 3** Caution	**Principal 4** Investigation	**Principal 5** Probing
Latency	0.54		0.34	0.34	0.33
Frequency		0.44			0.37
Duration	-0.54		-0.42		
Confusion		0.90			
Self-Doubt	-0.54	0.47	0.35		0.41

	Principal 1 Technical Competence	**Principal 2** Uncertainty	**Principal 3** Caution	**Principal 4** Investigation	**Principal 5** Probing
Confidence	0.40		-0.63	-0.32	
Surprise		0.78		-0.31	
Frustrated	-0.64	0.31	0.57		
Cyber_security	0.50	0.69			
Network_penetration	0.93				
Host_penetration	0.90				
Network_reconnaissance	0.82				0.33
Incidence_response	0.37	0.73		0.32	
Generalized_defense_practice	0.44	0.31		0.69	
Network_protocol_reverse_engineering	0.78		0.38		
Binary_reverse_engineering	0.53	0.41	0.51		
Percentage of Total Variance Explained	34.2%	20.9%	12.1%	7.6%	6.0%

d. Present-Uninformed/Absent-Uninformed (PUAU)

For PCA conducted on the Day 2 PUAU group, the first principal component is composed of loadings greater than 0.30 from nearly every variable with the exception of Latency, Duration, Incidence response, and Generalized defense practice. Negative loadings are present for Frequency (-0.40), Confusion (-0.56), Self-Doubt (-0.69), Surprise (-0.33), and Frustrated (-0.44). The combination of these negative loadings and positive loadings for Confidence and other self-reported experience levels result in the title "Technical Competence" being applied to this component. Component two is composed of all positive loadings spread across technical performance measures of Frequency, all Cyber TSQ scores with the exception of Confidence, which has a loading of less than 0.30, and five of eight self-reported experience levels. This combination of loadings is titled "Probing." The third component is composed of negative loadings for Incidence response (-0.33) and Generalized defense practice (-0.66) along with positive loadings for Frequency (0.38), Duration (0.75), Surprise (0.41), and Frustrated (0.33). This component is titled "Exploration" with the grouping of Frequency and Duration in conjunction with Surprise and Frustrated. Component four includes negative loadings for Cyber security (-0.41), Network penetration (-0.33) and Confusion (-0.41) and positive loadings for Latency (0.54), Confidence (0.32), Surprise (0.47) and Incidence response (0.55). These loadings suggest a degree of "Caution." Finally, component five is composed of negative loadings for Latency (-0.35), Frustrated (-0.47), Network protocol reverse engineering (-0.42), and Binary reverse engineering and positive

loadings for Frequency (0.43), Duration (0.40) and Incidence response (0.40), indicative of a level of "Assurance" in naming this component. The Loading Matrix for loadings greater than 0.30 is provided in Table 81 with the full loading matrix available in Appendix C. A visual depiction of the Eigenvalues, Score Plot and Loading Plot are provided in Figure 82 with the Scree Plot provided in Figure 83.

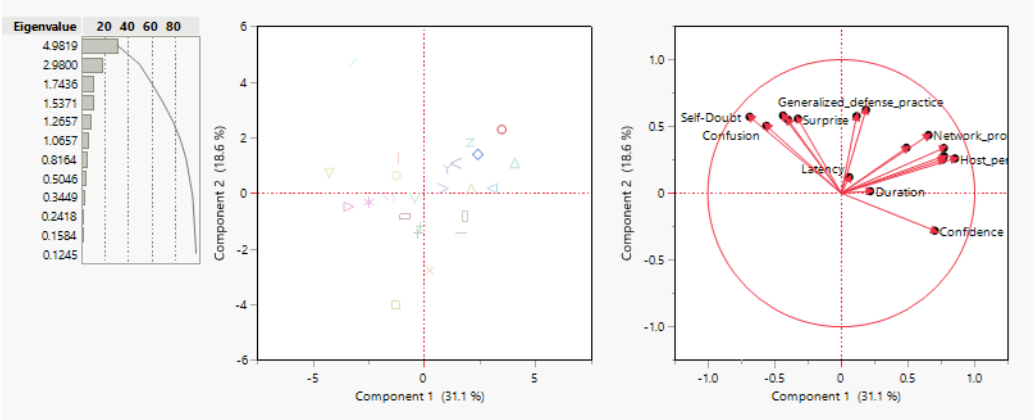


Figure 82. Day 2 Present-Uninformed/Absent-Uninformed PCA Summary Plots

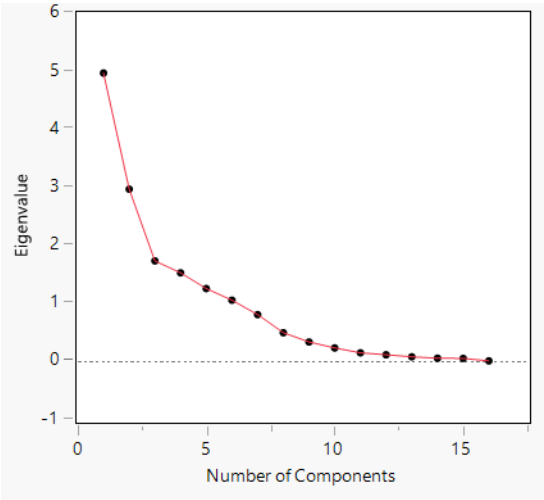


Figure 83. Day 2 Present-Uninformed/Absent-Uninformed PCA Scree Plot

Table 81. Day 2 Present-Uninformed/Absent-Uninformed PCA

	Principal 1 Technical Competence	**Principal 2** Probing	**Principal 3** Exploration	**Principal 4** Caution	**Principal 5** Assurance
Latency				0.54	-0.35
Frequency	-0.40	0.55	0.38		0.43
Duration			0.75		0.40
Confusion	-0.56	0.50		-0.41	
Self-Doubt	-0.69	0.57			
Confidence	0.70			0.32	
Surprise	-0.33	0.56	0.41	0.47	
Frustrated	-0.44	0.58	0.33		-0.47
Cyber_security	0.49	0.34		-0.41	
Network_penetration	0.77			-0.33	
Host_penetration	0.85				
Network_reconnaissance	0.77	0.34			
Incidence_response		0.58	-0.33	0.55	0.40
Generalized_defense_practice		0.62	-0.66		
Network_protocol_reverse_engineering	0.65	0.44			-0.42
Binary_reverse_engineering	0.76				-0.32
Percentage of Total Variance Explained	31.1%	18.6%	10.9%	9.6%	7.9%

e. Combined AIAU, AUPU, PIAU, PUAU

PCA was applied to the technical performance, Cyber TSQ scores and self-reported experience levels for subjects across all experimental conditions for Day 2. The Loading Matrix for loadings greater than 0.30 is provided in Table 82 with the full loading matrix available in Appendix C. A visual depiction of the Eigenvalues, Score Plot and Loading Plot are provided in Figure 84 with the Scree Plot provided in Figure 85.

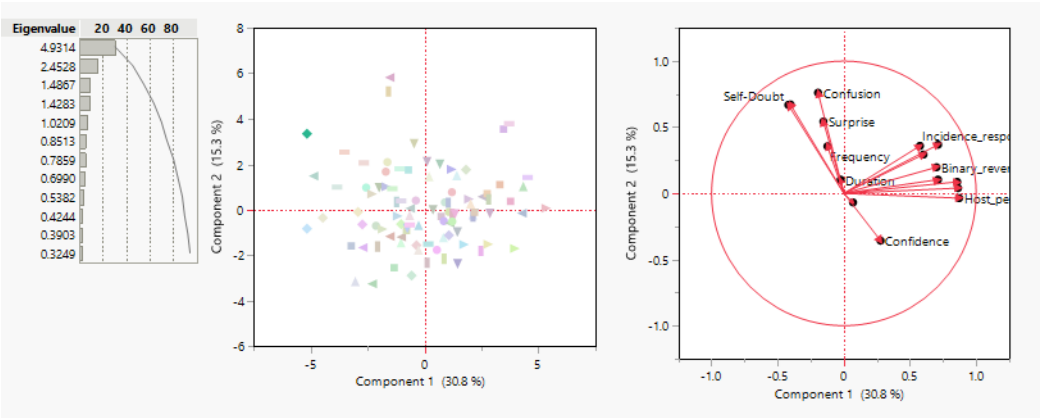


Figure 84. Day 2 Combined PCA Summary Plots

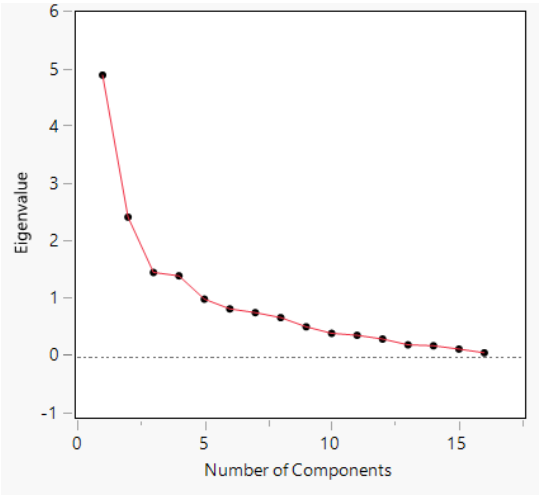


Figure 85. Day 2 Combined PCA Scree Plot

Table 82. Day 2 Combined PCA

	Prin1	Prin2	Prin3	Prin4	Prin5
Latency			-0.67		0.42
Frequency		0.36		-0.46	
Duration			0.68		
Confusion		0.76			
Self-Doubt	-0.40	0.67			
Confidence		-0.36	0.55		0.47
Surprise		0.54	0.31		0.57
Frustrated	-0.42	0.67			
Cyber_security	0.71	0.37			
Network_penetration	0.86				
Host_penetration	0.87				
Network_reconnaissance	0.86				
Incidence_response	0.57	0.36		-0.55	0.31
Generalized_defense_practice	0.60			-0.51	
Network_protocol_reverse_engineering	0.71			0.43	
Binary_reverse_engineering	0.70			0.41	
Percentage of Total Variance Explained	30.8%	15.3%	9.3%	8.9%	6.4%

B. EXPLORATORY FACTORY ANALYSIS

EFA explores the relationships between variables by focusing on common variance captured by these variables with the goal of identifying latent, unobserved variables to study the structure of the underlying data (Henson & Roberts, 2006). The use of EFA for analysis of subject data captured during the Tularosa study within each experiment condition group

and in aggregate seeks to improve conceptual understanding of correlations between measures of technical performance, cognitive state, and experience, a key outcome of EFA highlighted by Fabrigar et al. in their evaluation of EFA in psychological research (Fabrigar et al., 1999). EFA was conducted with JMP® Pro Version 16.0.0 using the "Principal Axis" factor method with "Principal Components" for prior communality. Five factors were used with oblique Promax rotation strategy. Oblique rotation was used to focus on correlation among factors with Promax rotation providing the highest cumulative percentage of variance explained from the oblique rotation methods available within JMP. Full pattern structure matrix, communalities, total variance explained by the factors, initial EVs and variance explained by each factor after rotation are included in the Appendix.

EFA conducted across the experimental conditions for both Day 1 and Day 2 data is largely congruent with the PCA performed on the same data for loadings of the first component and factor, but present more significant deviations in loadings for subsequent components and factors. It is important to highlight EFA aims to capture the maximum common variance, whereas PCA aims to obtain the maximum total variance (Alavi et al., 2020). Also, EFA assumes the latent constructs identified are independent of the loadings identified, whereas PCA assumes the components identified are dependent on the loadings generated (Coltman et al., 2008). The congruence of experience loadings between the first component and first latent factor supports observations from operational practice on the critical role of technical expertise in the conduct of penetration testing against networks and systems.

1. Day 1 Analysis

a. Absent-Informed

The first factor identified for the Absent-Informed condition contained loadings in excess of 0.50 for all self-reported experience levels, highlighting the importance of specific technical knowledge. This factor is titled "Technical Aptitude" to capture the latent factor being representative of technical expertise, which is a function of aptitude to gain such knowledge. The second factor, titled "Uncertainty" is a combination loadings for frequency (0.87) and confusion (0.63) along with a negative loading for confidence (-0.67). This

suggests a degree of uncertainty with common variance of confidence being negatively correlated with corresponding common variance of frequency and confusion. Likewise in factor three, self-doubt (0.71) and frustration (0.89 have high loadings, with lower loading scores for duration (0.38), incidence response (0.36), generalized defense practice (0.39) and negative loading scores for host penetration (-0.36) and network protocol reverse engineering (-0.32). From these loadings, this factor is labeled "Aggravation." The fourth factor, titled "Deliberation" is a combination of latency (0.87) with negative loading for duration (-0.40) and surprisingly positive loadings for network protocol reverse engineering (0.54) and binary reverse engineering (0.45). This latent factor may represent a problem solving or deliberation process. Factor five, with a strong loading for surprise (0.96) and negative loadings for duration (-0.41) and binary reverse engineering (-0.31) may represent a level of concern shared in the common variance extracted from the data resulting in the title "Consternation." The loadings for all factors and percentage of common variance explained by each factor are illustrated in Table 83.

Table 83. Day 1 Absent-Informed EFA

	Factor 1 Technical Aptitude	**Factor 2** Uncertainty	**Factor 3** Aggravation	**Factor 4** Deliberation	**Factor 5** Consternation
Latency				0.87	
Frequency		0.87			
Duration			0.38	-0.40	-0.41
Confusion		0.83			
Self-Doubt			0.71		
Confidence		-0.67			
Surprise					0.96
Frustrated			0.89		
Cyber_security	0.86				
Network_penetration	0.95				
Host_penetration	0.87		-0.36		
Network_reconnaissance	0.87				
Incidence_response	0.80		0.36		
Generalized_defense_practice	0.71		0.39		
Network_protocol_reverse_engineering	0.50		-0.32	0.54	
Binary_reverse_engineering	0.64			0.45	-0.31
Variance Explained by Each Factor	32.6%	16.3%	16.4%	11.7%	10.2%

b. Absent-Uninformed

Unlike the Absent-Informed condition, the first factor for Absent-Uninformed condition includes a significant loading for Confidence (0.64) along with loadings for all self-reported experience levels apart from a lower loading for "Generalized Defense Practice" (0.41) and Incidence Response, which had a factor loading of less than 0.30. Given the significant loadings of self-reported experience levels, this factor is also titled "Technical Aptitude." The second factor identified contains significant positive loadings for Frustrated (0.88) and Confusion (0.83) with less significant negative loadings for duration (-0.32), Cyber security (-0.36), and Generalized Defense Practice (-0.44). These strong loadings for Confusion and Frustrated suggest a degree of uncertainty, resulting in this factor being titled "Uncertainty." Factor 3 includes a significant positive loading for latency (0.83) and significant negative loading for duration (-0.80) suggests an underlying factor of "Hesitancy." Factor 4 includes significant positive loadings for frequency (0.88) and Self-Doubt (0.44) which suggest influence from an eagerness for "Exploration." Finally, Factor 5 includes significant positive loadings for Self-Doubt (0.71) and Surprise (0.86) reflecting a level of "Consternation." The loadings for all factors and percentage of common variance explained by each factor are provided in Table 84.

Table 84. Day 1 Absent-Uninformed EFA

	Factor 1 Technical Aptitude	**Factor 2** Uncertainty	**Factor 3** Hesitancy	**Factor 4** Exploration	**Factor 5** Consternation
Latency			0.83		
Frequency				0.88	
Duration		-0.32	-0.80		
Confusion		0.83		-0.32	
Self-Doubt				0.44	0.71
Confidence	0.63				
Surprise					0.86
Frustrated		0.88			
Cyber_security	0.77	-0.36			
Network_penetration	0.92				
Host_penetration	0.91				
Network_reconnaissance	0.88				
Incidence_response			0.34	-0.47	
Generalized_defense_practice	0.41	-0.44	0.32	-0.34	
Network_protocol_reverse_engineering	0.75				
Binary_reverse_engineering	0.86			0.32	
Variance Explained by Each Factor	33.5%	15.7%	12.2%	11.9%	11.1%

c. *Present-Informed*

The first factor identified for the Present-Informed condition contains significant loadings for Network penetration (0.77), Host penetration (0.95), Network protocol reverse engineering (0.97), and Binary reverse engineering (0.85) with a lesser loading for Network reconnaissance (0.30). This factor is also titled "Technical Aptitude" to capture a latent factor representative of a capacity to acquire technical knowledge. In contrast to the Absent-Informed condition, loadings greater than 0.50 were only observed for four self-reported experience areas. The second factor, with significant loadings for Cyber security (0.77), Network reconnaissance (0.67), Incidence response (0.90) and Generalized defense practice (0.70) and lesser loadings for Surprise (0.33) and negative loading for Confusion (-0.38), suggests a level of "Assurance." The third factor, titled "Uncertainty" is a combination of Cyber TSQ scores with positive loadings for Confusion (0.47), Self-Doubt (0.70), Surprise (0.67), and Frustrated (0.88) with a negative loading for Confidence (-0.70). Factor 4, with loadings for Frequency (0.94), Duration (0.46), and Confusion (0.58) suggests interest in "Exploration" resulting in the factor title. Finally, Factor 5 suggests a degree of "Hesitancy" with positive loadings for Latency (0.88) and Self-Doubt (0.34) and negative loading for Duration (-0.62). The loadings for all factors and percentage of common variance explained by each factor are illustrated in Table 85.

Table 85. Day 1 Present-Informed EFA

	Factor 1 Technical Aptitude	**Factor 2** Assurance	**Factor 3** Uncertainty	**Factor 4** Exploration	**Factor 5** Hesitancy
Latency					0.88
Frequency				0.94	
Duration				0.46	-0.62
Confusion		-0.38	0.47	0.58	
Self-Doubt			0.70		0.34
Confidence			-0.70		
Surprise		0.33	0.67		
Frustrated			0.88		
Cyber_security		0.77			
Network_penetration	0.77				
Host_penetration	0.95				
Network_reconnaissance	0.30	0.67			
Incidence_response		0.90			
Generalized_defense_practice		0.70			

	Factor 1 Technical Aptitude	**Factor 2** Assurance	**Factor 3** Uncertainty	**Factor 4** Exploration	**Factor 5** Hesitancy
Network_protocol_reverse_engineering	0.97				
Binary_reverse_engineering	0.85				
Variance Explained by Each Factor	26.1%	23.4%	16.4%	11.7%	9.6%

d. Present-Uninformed

The first factor identified for the Present-Uninformed condition contained loadings across technical performance measures, Latency (-0.52) and Duration (0.81), Cyber TSQ responses, Confidence (0.48) and Surprise (-0.34), and self-reported experience levels, Cyber security (0.30), Network penetration (0.85), Host penetration (0.90), Network reconnaissance (0.80), Network protocol reverse engineering (0.60), and Binary reverse engineering (0.82). Factor 2 includes positive loadings from Confidence (0.31), Cyber security (0.35), Incidence response (0.87), and Generalized defense practice and negative loading for Frustrated (-0.42) indicating a level of "Assurance," the title for this Factor. Factor 3 includes loadings for Confusion (0.84), Self-Doubt (0.79), Frustrated (0.34) and Cyber security (0.40) resulting in a title "Uncertainty" for this Factor. The fourth factor included a negative loading for Latency (-0.69) and positive loadings for Surprise (0.81) and Frustrated (0.63), which suggests a level of "Caution," which provides the title for this factor. Finally, Factor 5 includes loadings for both Frequency (0.79) and Duration (0.35), which indicate "Persistence," the title given to this Factor. Unique among Day 1 EFA across all experiment condition groups, Factor 1 included loadings across technical performance measures, Cyber TSQ and self-reported experience levels. The loadings for all factors and percentages of common variance explained by each factor as displayed in Table 86.

Table 86. Day 1 Present-Uninformed EFA

	Factor 1 Technical Aptitude	**Factor 2** Assurance	**Factor 3** Uncertainty	**Factor 4** Caution	**Factor 5** Persistence
Latency	-0.52			-0.69	
Frequency					0.79
Duration	0.81				0.35
Confusion			0.84		
Self-Doubt			0.79		

	Factor 1 Technical Aptitude	**Factor 2** Assurance	**Factor 3** Uncertainty	**Factor 4** Caution	**Factor 5** Persistence
Confidence	0.48	0.31			
Surprise	-0.34			0.81	
Frustrated		-0.42	0.34	0.63	
Cyber_security	0.30	0.35	0.40		
Network_penetration	0.85				
Host_penetration	0.90				
Network_reconnaissance	0.80				
Incidence_response		0.87			
Generalized_defense_practice		0.91			
Network_protocol_reverse_engineering	0.60				-0.61
Binary_reverse_engineering	0.82				-0.34
Variance Explained by Each Factor	34.8%	23.3%	11.3%	12.5%	9.2%

e. Combined Absent-Informed, Absent-Uninformed, Present-Informed and Present-Uninformed

EFA was applied to the technical performance, Cyber TSQ scores and self-reported experience levels for subjects across all experimental conditions for Day 1 as presented in Table 87. This information is provided for comparison purposes without additional analysis.

Table 87. Day 1 Combined EFA

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**
Latency				0.87	
Frequency					0.86
Duration				-0.73	0.50
Confusion			0.74		
Self-Doubt			0.65		
Confidence	0.55		-0.43		
Surprise			0.62		-0.36
Frustrated			0.83		
Cyber_security	0.42	0.54			
Network_penetration	0.78				
Host_penetration	0.87				
Network_reconnaissance	0.60	0.35			
Incidence_response		0.91			
Generalized_defense_practice		0.92			
Network_protocol_reverse_engineering	0.87				
Binary_reverse_engineering	0.92				
Variance Explained by Each Factor	29.7%	20.2%	14.9%	10.4%	7.8%

2. Day 2 Analysis

a. Absent-Informed/Absent-Uninformed (AIAU)

The first factor identified for the AIAU condition contained loadings in excess of 0.50 for only self-reported experience levels of Cyber security (0.57), Network penetration (0.78), Host penetration (0.93), and Network reconnaissance (0.98), in contrast to the Day 1 Absent-Informed condition where all self-reported experience levels had loadings in excess of 0.50. A positive loading for Surprise (0.52) and negative loading for Frustrated (-0.37) were also identified. This factor is also titled "Technical Aptitude" given the preponderance of loadings involving self-reported experience levels. The second factor contains a large negative loading for Latency (-0.79) along with loadings greater than 0.50 for Incidence response (0.51), Generalized defense practice (0.67), Network protocol reverse engineering (0.57), and Binary reverse engineering (0.64) along with lesser loadings for Cyber security (0.47) and Network penetration (0.32). This factor is titled "Assurance" to highlight the negative loading on Latency. Factor 3 includes significant loadings for Confusion (0.89), Self-Doubt (0.85), Surprise (0.62), and Frustrated (0.79) resulting in a title of "Uncertainty." With a significant positive loading for Frequency (0.78) and negative loading for Confidence (-0.82) and lesser loadings for Self-Doubt (0.40), Incidence response (0.35), Generalized defense practice (0.42) and Network protocol reverse engineering (-0.37), factor 4 is titled "Exploration" accounting for the loadings on Frequency and Confidence. Factor 5 includes significant positive loading for Duration (0.89) and negative loadings with Latency (-0.33) and Network protocol reverse engineering (-0.48) indicating "Persistence." The loadings for all factors and percentages of common variance explained by each factor are illustrated in Table 88.

Table 88. Day 2 Absent-Informed/Absent-Uninformed EFA

	Factor 1 Technical Aptitude	**Factor 2** Assurance	**Factor 3** Uncertainty	**Factor 4** Exploration	**Factor 5** Persistence
Latency		-0.79			-0.33
Frequency				0.78	
Duration					0.89
Confusion			0.89		
Self-Doubt			0.85	0.40	

	Factor 1 Technical Aptitude	**Factor 2** Assurance	**Factor 3** Uncertainty	**Factor 4** Exploration	**Factor 5** Persistence
Confidence				-0.82	
Surprise	0.52		0.62		
Frustrated	-0.37		0.79		
Cyber_security	0.57	0.47			
Network_penetration	0.78	0.32			
Host_penetration	0.93				
Network_reconnaissance	0.98				
Incidence_response	0.39	0.51		0.35	
Generalized_defense_practice		0.67		0.42	
Network_protocol_reverse_engineering		0.57		-0.37	-0.48
Binary_reverse_engineering	0.30	0.64			
Variance Explained by Each Factor	26.9%	21.1%	17.1%	13.6%	8.3%

b. Absent-Uninformed/Present-Uninformed (AUPU)

The first factor identified for the AUPU condition contained loadings in excess of 0.50 for only self-reported experience levels of Cyber security (0.65), Network penetration (0.90), Host penetration (0.89), Network reconnaissance (0.74), Network protocol reverse engineering (0.79), and Binary reverse engineering along with negative loadings for Frequency (-0.31) and Surprise (-0.42). This factor again is titled "Technical Aptitude" to account for the significant loadings of self-reported experience levels. The second factor is titled "Assurance" on account of a large negative loading for Self-Doubt (-0.58) and larger loadings for Incident response (0.95) followed by Generalized defense practice (0.83) then Cyber security (0.49), Network reconnaissance (0.40) and finally Latency (0.31). Factor 3 loadings included technical performance measures of Latency (-0.33) and Duration (0.52) along with Cyber TSQ scores for confusion (0.73), Self-Doubt (0.70), Confidence (-0.44) and Frustrated (0.89) resulting in a factor title of "Uncertainty" to account for these loadings. Negative loadings for Latency (-0.44) and Confusion (-0.37) along with positive loadings for Confidence (0.78) and Surprise (0.64) comprise Factor 4, which is titled "Composure" to account for the positive loadings of both Confidence and Surprise and negative loading for Latency. Finally, Factor 5 is comprised solely of technical performance factors with a negative loading for Latency (-0.40) and positive loadings for Frequency (0.89) and Duration (0.42) suggesting "Exploration," the title for this factor.

The loadings for all factors and percentages of common variance explained by each factor are illustrated in Table 89.

Table 89. Day 2 Absent-Uninformed/Present-Uninformed EFA

	Factor 1 Technical Aptitude	**Factor 2** Assurance	**Factor 3** Uncertainty	**Factor 4** Composure	**Factor 5** Exploration
Latency		0.31	-0.33	-0.44	-0.40
Frequency	-0.31				0.89
Duration			0.52		0.42
Confusion			0.73	-0.37	
Self-Doubt		-0.58	0.70		
Confidence			-0.44	0.78	
Surprise	-0.42			0.64	
Frustrated			0.89		
Cyber_security	0.65	0.49			
Network_penetration	0.90				
Host_penetration	0.89				
Network_reconnaissance	0.74	0.40			
Incidence_response		0.95			
Generalized_defense_practice		0.83			
Network_protocol_reverse_engineering	0.79			-0.34	
Binary_reverse_engineering	0.95				
Variance Explained by Each Factor	30.9%	18.6%	15.5%	9.8%	9.0%

c. Present-Informed/Absent-Uninformed (PIAU)

The first factor identified for the PIAU condition contained four loadings greater than 0.30 for self-reported experience levels, the fewest loadings for self-reported experience levels for all Day 1 and Day 2 conditions. These loadings range from 0.86 for Binary reverse engineering, to 0.84 for Network protocol reverse engineering, 0.74 for Host penetration and 0.53 for Network penetration. Factor 1 also included negative loadings for Duration (-0.41) and Self-Doubt (-0.30) The heavy weighting of self-reported experience levels result in this factor being titled "Technical Aptitude." The second factor includes only positive loadings across technical performance measures, Cyber TSQ scores and a single self-reported experience area. Surprise has the highest loading for this factor of 0.81, followed by Confusion (0.73), Frequency (0.70), Self-Doubt (0.61), and Cyber security. This factor is titled "Uncertainty" to account for these loadings. The third factor includes loadings for Self-Doubt (0.67), Confidence (-0.87), Frustrated (0.83), Cyber

security (-0.40), Network penetration (-0.33), and Network reconnaissance (-0.31). The negative loadings for Confusion and the self-reported experience levels may indicate a level of "Disorientation," the title given to this factor. Factor four includes loadings for Latency (0.78), Duration (-0.56), Self-Doubt (0.32), Confidence (-0.36) Network penetration (0.46), and Network reconnaissance (0.62). These loadings may indicate a degree of wariness, given the high loading value for Latency and negative loading values for Duration and Confidence. Finally, factor five includes only two loadings, Generalized defense practice, which has an overloaded loading of 1.02 and Incidence response with a loading of 0.77. These loadings may indicate latent "Vigilance," hence the title for this factor. The loadings for all factors and percentages of common variance explained by each factor are illustrated in Table 90.

Table 90. Day 2 Present-Informed/Absent-Uninformed EFA

	Factor 1 Technical Aptitude	**Factor 2** Uncertainty	**Factor 3** Disorientation	**Factor 4** Wariness	**Factor 5** Vigilance
Latency				0.78	
Frequency		0.70		-	
Duration	-0.41			-0.56	
Confusion		0.73			
Self-Doubt	-0.30	0.61	0.67	0.32	
Confidence			-0.87	-0.36	
Surprise		0.81			
Frustrated			0.83		
Cyber_security		0.61	-0.40		
Network_penetration	0.53		-0.33	0.46	
Host_penetration	0.74				
Network_reconnaissance			-0.31	0.62	
Incidence_response					0.77
Generalized_defense_practice					1.02
Network_protocol_reverse_engineering	0.84				
Binary_reverse_engineering	0.86				
Variance Explained by Each Factor	23.3%	19.6%	22.0%	21.1%	17.0%

d. Present-Uninformed/Absent-Uninformed (PUAU)

Loadings for the first factor of the PUAU condition are more representative of EFA loadings for Day 1 and Day 2 groups, with loadings greater than 0.30 for all self-reported experience levels except for Incidence response. These loadings result in a title of

"Technical Aptitude" for this factor. Factor 2 includes positive loadings for Frequency (0.50), Confusion (0.90), Self-Doubt (0.72), and Frustrated (0.89) and a negative loading for Confidence (-0.76). The combination of loadings for Frequency and Cyber TSQ scores result in a label of "Probing" being applied to this factor. Factor 3 includes only loadings for Incidence response (0.95) and Generalized defense practice (0.81), similar to factor 5 of the PIAU condition, resulting in the same title "Vigilance." Factor 4 includes loadings for Frequency (0.67), Duration (0.86) and Surprise (0.57), leading to the title of "Exploration" to account for this combination of loadings. Finally, factor 5 includes positive loadings for Latency (0.67), Surprise (0.38), Frustrated (0.34), and Network protocol reverse engineering (0.63) and a negative loading for Frequency (-0.37), suggesting "Caution" the title for this factor. The loadings for all factors and percentages of common variance explained by each factor are illustrated in Table 91.

Table 91. Day 2 Present-Uninformed/Absent-Uninformed EFA

	Factor 1 Technical Aptitude	**Factor 2** Probing	**Factor 3** Vigilance	**Factor 4** Exploration	**Factor 5** Caution
Latency					0.67
Frequency		0.50		0.67	-0.37
Duration				0.86	
Confusion		0.90			
Self-Doubt		0.72			
Confidence		-0.76			
Surprise				0.57	0.38
Frustrated		0.89			0.34
Cyber_security	0.72				
Network_penetration	0.88				
Host_penetration	0.85				
Network_reconnaissance	0.79				
Incidence_response			0.95		
Generalized_defense_practice	0.33		0.81		
Network_protocol_reverse_engineering	0.64				0.63
Binary_reverse_engineering	0.81				
Variance Explained by Each Factor	27.9%	23.4%	13.1%	11.4%	10.4%

e. Combined AIAU, AUPU, PIAU, PUAU

EFA was applied to the technical performance, Cyber TSQ scores and self-reported experience levels for subjects across all experimental conditions for Day 2 as presented in

Table 92. This information is provided for comparison purposes without additional analysis.

Table 92. Day 2 Combined EFA

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency				0.80	
Frequency			0.37	-0.51	
Duration				-0.62	
Confusion		0.78			
Self-Doubt		0.83			
Confidence		-0.52			0.65
Surprise		0.49			0.72
Frustrated		0.83			
Cyber_security	0.45		0.54		
Network_penetration	0.83				
Host_penetration	0.86				
Network_reconnaissance	0.64		0.34		
Incidence_response			0.93		
Generalized_defense_practice			0.83		
Network_protocol_reverse_engineering	0.87				
Binary_reverse_engineering	0.89				
Variance Explained by Each Factor	27.7%	18.0%	18.6%	8.9%	7.5%

THIS PAGE INTENTIONALLY LEFT BLANK

VII. THEORETICAL MODELS OF DEFENSIVE CYBER DECEPTION

Analysis of the empirical data from the Tularosa study inspired the development of two categories of theoretical models for exploration of cyber deception impact on attacker behavior, vector-based and quantum-theory based. The vector-based model was constructed to investigate the non-additive effect generated by the combination of the technical presence of cyber deception and correct awareness of its presence. Development of this model resulted in a connection with quantum cognition, an emerging field exploring human decision-making using principles from Quantum Probability Theory (QPT). While these models provide limited and approximate models of the complex phenomenon of cyber deception, they were developed to aid in generation of new hypotheses for future experimentation and improvement in techniques to assess the operational effectiveness of the deployment of defensive cyber deception systems.

A. VECTOR-BASED CONCEPTUAL MODELS

In the development of these theoretical models for cyber deception under the conditions of psychological deception and decoy-based technical deception, the temporal variable of latency was selected as a basis. This variable was selected as it provided the most significant differentiator between experimental groups. Latency is typically a measure of delay, although for this research latency was defined as the time between experiment trial start and event observation. As a measure of delay, the observed latency values enable comparison between groups for the impact of varying conditions on increasing delay in decision-making.

1. OODA Loop-based Model

The Observe, Orient, Decide, Act (OODA) Loop was developed by Colonel (Retired) John Boyd to explore decision-making in combat operations. In exploring the impact of delay on decision making, the OODA Loop provides an outline for the unobservable cognitive processes of a cyber-attacker prior to the initiation of sending the first packet to a target system, which is directly measured by the latency temporal variable.

Figure 86 provides a depiction of the adapted OODA Loop for the Absent-Uninformed Group. Modifications from the original OODA Loop are show in red, with "Information from Real Systems Passively Available" added as an input into the Observe phase, along with "Task Briefing" added as an input into the Outside Information input into the Observe phase. Subjects in the Tularosa study were able to passively observe traffic on the cyber range network.

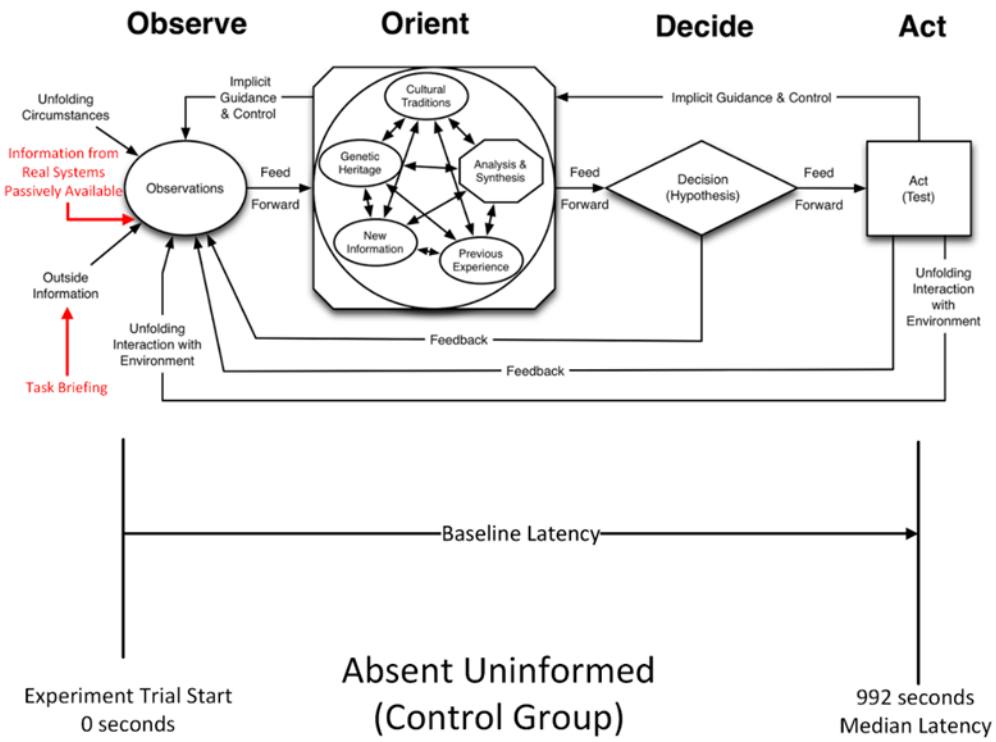


Figure 86. Absent-Uninformed OODA Loop Model. Adapted from Boyd (1995).

Subjects also needed to Orient themselves to the Kali Linux system provided for conducting reconnaissance to provide actionable intelligence about the target network as outlined in the Task Briefing provided at the beginning of each experiment trial. Based on the information gathered from the Observe and Orient phases, subjects executed a decision-making process, which ultimately resulted in the performance of an action. In this case, the action was sending a packet to a target system. Time spent in each of these phases cannot

be measured individually, however the duration of the combination of these phases can be discerned by the elapsed time from the start of the experiment trial and the observed latency. The real system median latency value for each group, with over 1,300 system interactions captured provides an optimal measurement to capture the overall duration of these phases.

As depicted in Figure 86, the median latency observed for the Absent-Uninformed control group was 992 seconds elapsed from the start of the experiment. Figure 87 depicts the Absent-Informed condition, which includes "Deception may be present." The addition of psychological deception boosts median latency observed to 1127 seconds elapsed from the start of the experiment, an increase of 135 seconds over the control group.

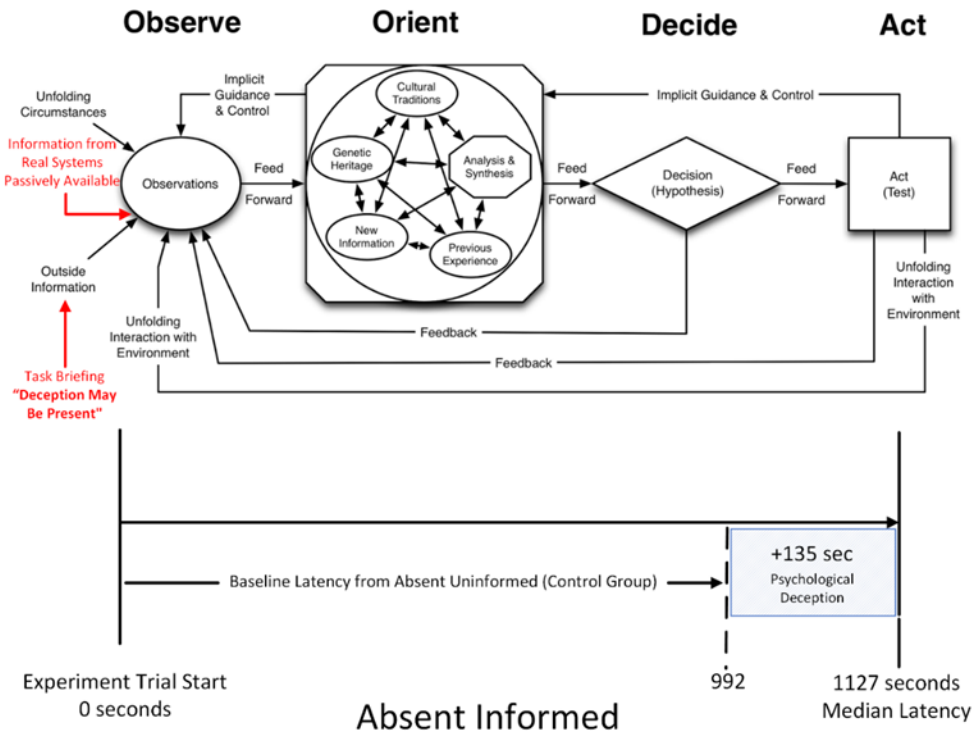


Figure 87. Absent-Informed OODA Loop Model. Adapted from Boyd (1995).

Figure 88 depicts the Present-Uninformed condition, which includes technical deception in the form of 50 decoy systems present on the cyber range network. The addition of

technical deception boosts median latency observed to 1068 seconds elapsed from the start of the experiment, an increase of 76 seconds over the control group.

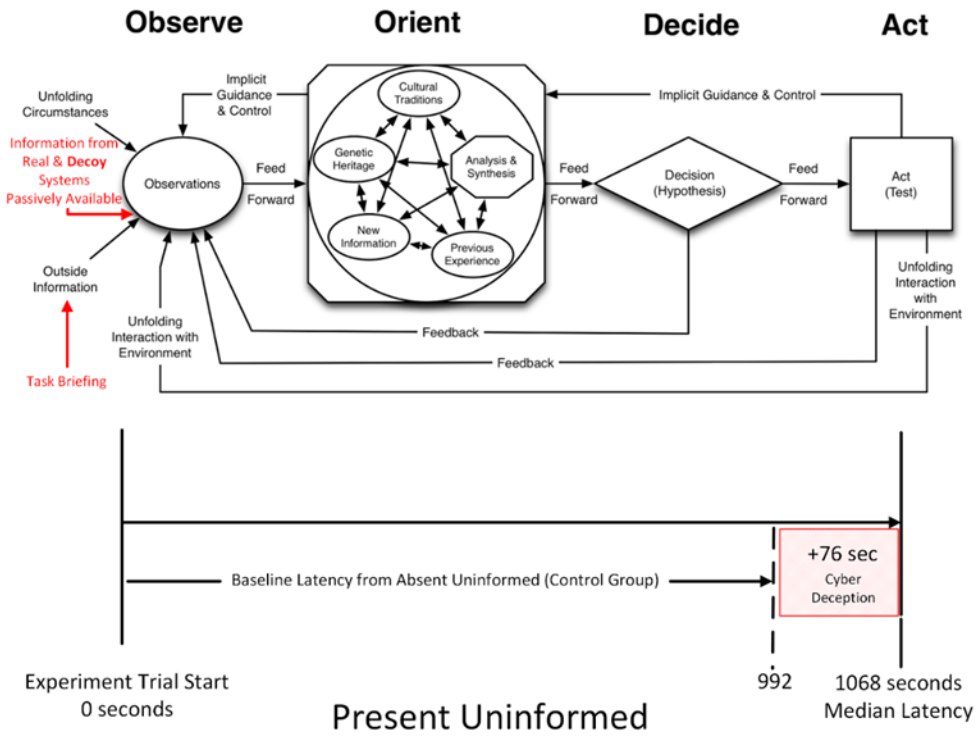


Figure 88. Present-Uninformed OODA Loop Model. Adapted from Boyd (1995).

The anticipated median latency value for the Present-Informed condition, which combines both psychological deception and technical deception was the additive combination of the increase in median latency or 211 seconds from the combination of 135 seconds and 76 seconds respectively. However, what was observed for the Present-Informed group was an increase of 377 seconds, or nearly 1.8 times the expected value as illustrated in Figure 89.

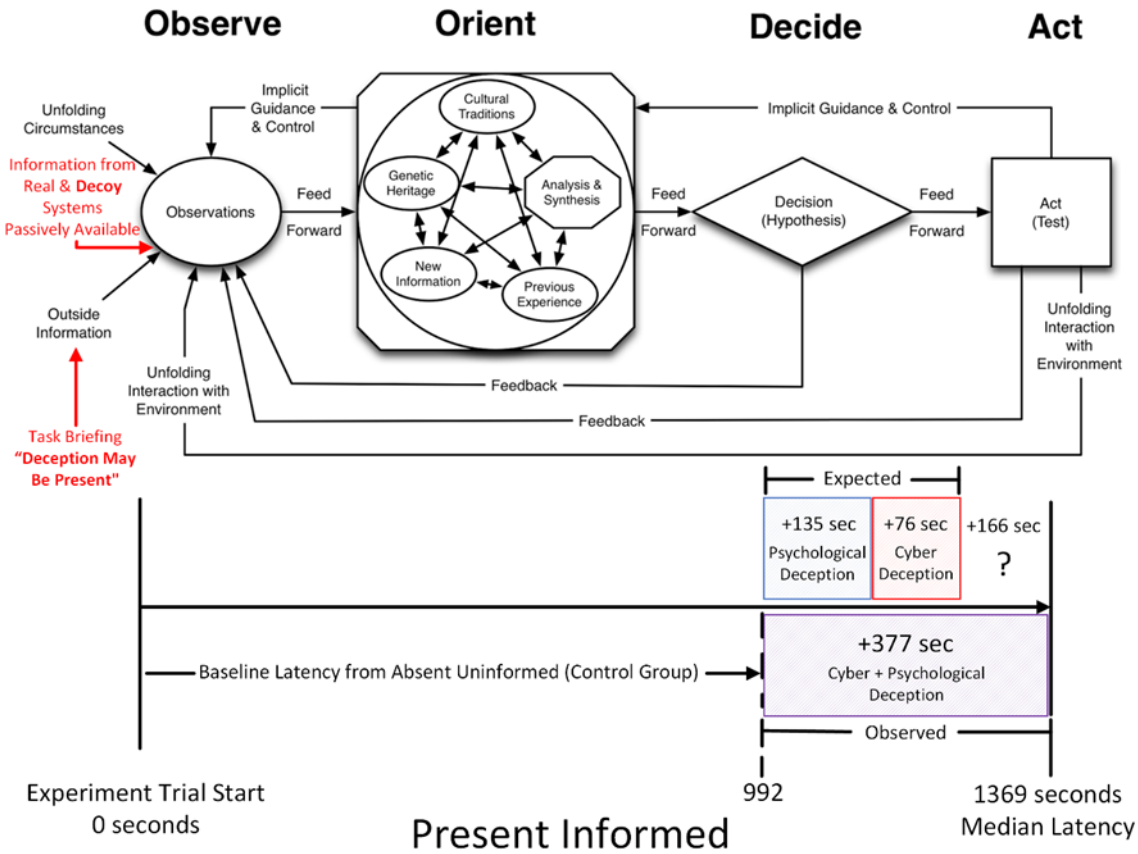


Figure 89. Present-Informed OODA Loop Model. Adapted from Boyd (1995).

2. Typology of Misperception Based-Model

A second perspective on modeling cyber-attacker decision-making under uncertainty was provided by the General Theory of Deception offered by Whaley. He stated, "deception is one form or mode of perception" and created the Typology of Misperception depicted in Figure 90 (Whaley, 1982, p. 180). Whaley (1982) describes the process of perception in five steps, which are roughly analogous to the Observe, Orient, Decide steps of Boyd's OODA Loop.

> 1) The environment continuously transmits a chaotic cascade or spectrum of discrete data (the 'information bits' of communications theorists). 2) Our sensors (intrinsic such as eye and ear as well as extrinsic such as seeing-eye dogs or radar sets) detect certain portions of some of these spectra. 3) These bits and scraps of received data are transmitted (with slight delay but often

> considerable distortion) to the brain. 4) The brain discards most of these data but processes some immediately and stores it in memory. 5) The brain then develops hypotheses about the environment by drawing inferences from new as well as stored data. (p. 181)

Perception in cyberspace is accomplished through passive and active reconnaissance to collect information about the target environment. Passive reconnaissance is conducted by subjects monitoring network traffic without interaction with other systems on the network. The passive nature of this reconnaissance would require recording or otherwise actively monitoring subject interaction with the Kali Linux system to determine the duration of this activity. In contrast, active reconnaissance is conducted by sending packets to other systems on the network. As such, the start of active reconnaissance is easily identified from the first packet sent to a target system by the subject.

Transitioning from perception to misperception, Whaley (1982) references the work of British neuropsychologist R.L. Gregory, who suggested:

> "illusions are failed hypotheses" and suggests that they occur whenever either 1) our physiological perceptual input mechanisms malfunction, or 2) our cognitive hypothesis-generating strategies are inappropriate. The first circumstance accounts for all physiological illusions, the second for all psychological ones. (p. 182)

Flipping Whaley's Typology of Misperception upside down, elements of misperception that impact the unobservable cognitive processes of a cyber-attacker prior to initiation of active reconnaissance could be considered as the summation of delays resulting from failed hypotheses generated from unintentional misrepresentation, self-deception, and illusion. For the Absent Uninformed control group, this summation of delays combined with delay caused by subjects orienting themselves to the Kali Linux system used could be considered as a baseline value for observed latency. Using this baseline value, the impact of deliberate deception can be assessed, with the underlying assumption that median latency for the control group is also representative of the experimental groups.

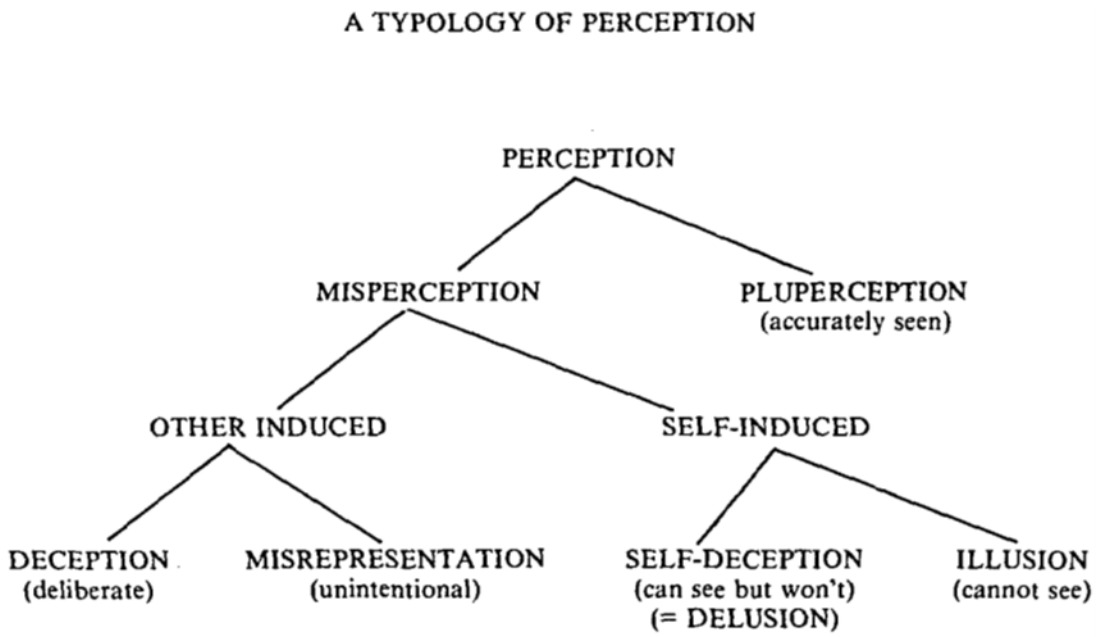


Figure 90. A Typology of Misperception. Source: Whaley (1982).

In this research based-on network traffic data captured during the Tularosa study, initial active reconnaissance was measured using latency, the seconds elapsed from the experiment start to the first packet sent to a target system on the cyber range network. With latency captured for subject activity where deception was both present and not present, empirical data can be fitted to this typology. Not included in Whaley's Typology is a consideration for system orientation by cyber-attackers but given all subjects in the Tularosa study used the same system configuration, the time for subjects to orient themselves to the Kali Linux system used to interact with target systems on the cyber range network could be considered as a constant value for all groups.

The Absent-Uninformed group demonstrated the lowest median latency of 992 seconds and the Present-Informed demonstrated the highest median latency of 1369 seconds. With deception not present for the Absent-Uninformed group, median time to initiate active reconnaissance (perception) could be considered as the summation of components for the impact of self-induced misperception, unintentional misrepresentation, and system orientation on non-observable cognitive processes prior to a packet being sent to a target system. With the data available it is not possible to assess what proportion each

of the components contribute to the overall time to initiate active reconnaissance. However, the data does provide a baseline value for evaluation of the impact of deliberate deception on active reconnaissance initiation. A depiction of the summation of components for the Absent-Uninformed condition is provided in Figure 91, for the Absent-Informed condition in Figure 92, for the Present-Uninformed condition in Figure 93 and for the Present-Informed condition in Figure 94. As with the OODA Loop based model, the observed median latency for the Present-Informed condition is an increase of 377 seconds instead of an increase of 211 seconds as would be expected by the additive combination of psychological and cyber deception.

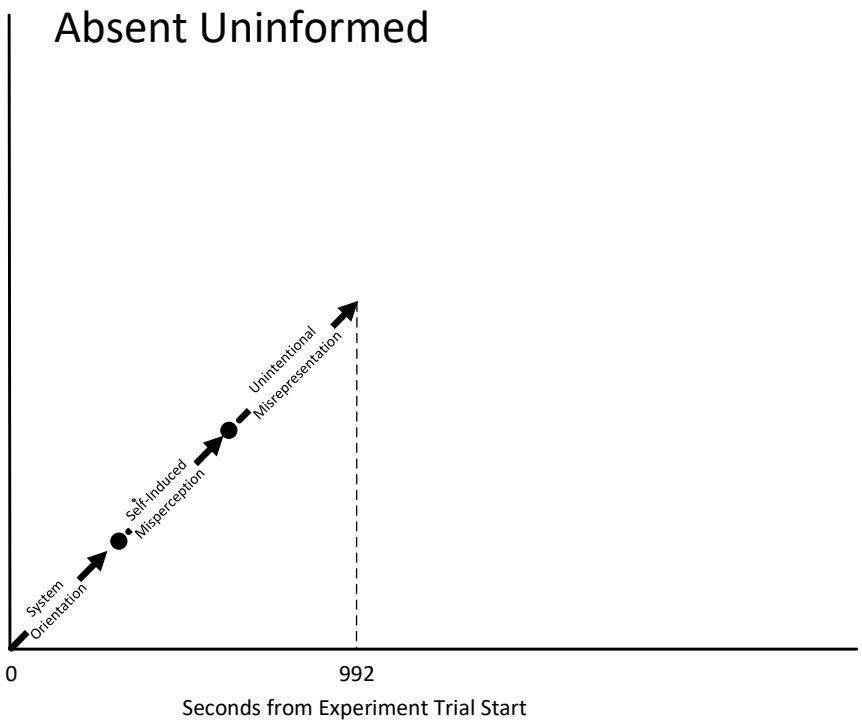


Figure 91. Absent-Uninformed Perception Components

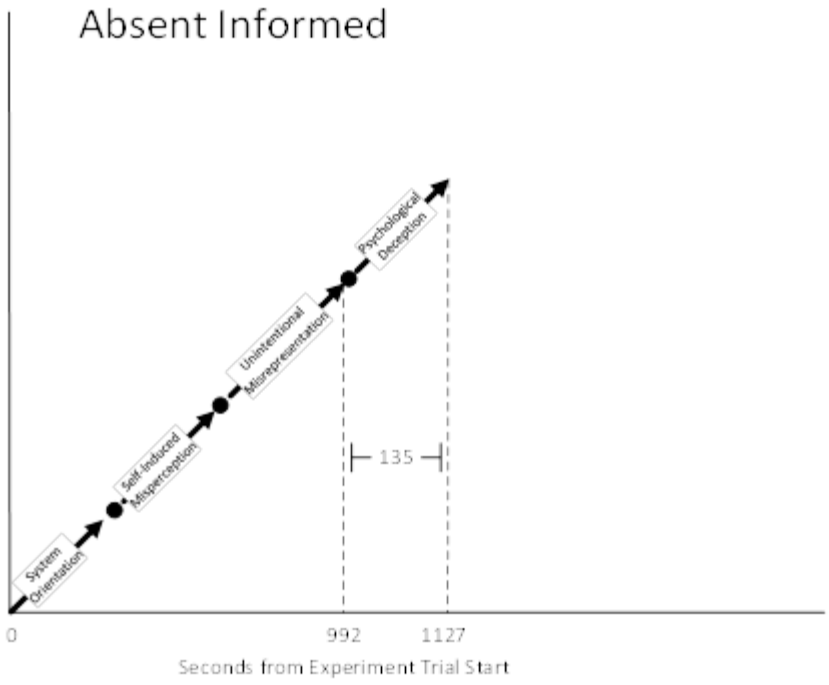


Figure 92. Absent-Informed Perception Components

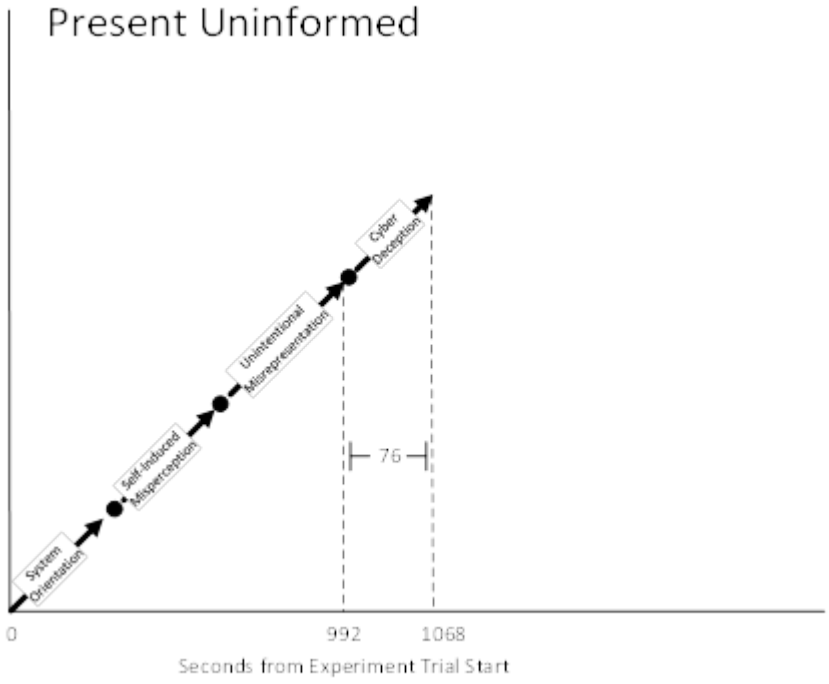


Figure 93. Present-Uninformed Perception Components

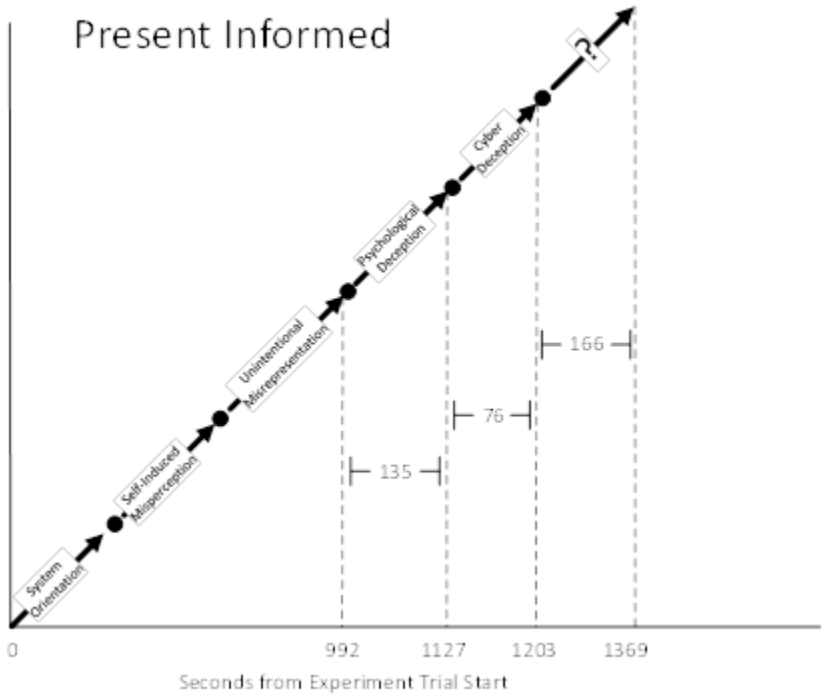


Figure 94. Present-Informed Perception Components

B. DEMPSTER-SHAFER THEORY APPLICATION

The first theory investigated to help explain the non-additive increase in median latency result observed was the Dempster-Shafer theory (DST). DST was developed by Arthur Dempster and Glenn Shafer as a mathematical theory for reasoning with uncertainty of evidence (Sentz & Ferson, 2002). Using functions for *Belief*, *Plausibility* and Basic probability assignment, evidence from multiple sources or observations can combined to quantify uncertainty for a given hypothesis (Sentz & Ferson, 2002). DST has previously been applied to the field of cyber deception in a 2012 article by Hayatle et al. in which they describe how DST could be leveraged for combining evidence to achieve a relatively high confidence in identifying honeypot systems (Hayatle et al., 2012). DST has also been applied to risk assessment for Information Systems security using evidence accumulation to account for risk (Sun et al., 2006).

Analyzing subject activity during the Tularosa study through the lens of DST was used to investigate for the Absent-Uninformed and Present-Informed groups depicted in Figures 91 and 94. For these groups subject activity steadily increases over time,

presumably on account of evidence accumulation for the presence or lack of cyber deception. For the Absent-Informed and Present-Uninformed groups, however, the utility of DST is less clear. A reasonable assumption would offer that the greater accumulation of evidence, even under psychological deception or cyber deception, would result in steady or increasing rates of network activity as observed for the Absent-Uninformed and Present-Informed groups. However, graphs of aggregate subject activity for the Absent-Informed and Present-Uninformed groups revealed patterns of oscillation. For these groups, aggregate subject network activity fluctuates over the elapsed experiment time across graphs for average packets per minute, median packets per minute, and percentage of subject activity across multiple thresholds. A depiction of the Day 1 Real System Average Packets Per Minute with the peaks and troughs of network activity marked is provided in Figure 95. For the Absent-Informed group, there is a peak in network activity at approximately 1,500 seconds from the experiment start, a trough in activity at 4,500 seconds, another peak in activity at 8,500 seconds, another trough at 13,000 seconds, and a peak in activity at 18,000 seconds before decreasing through the end of the experiment.

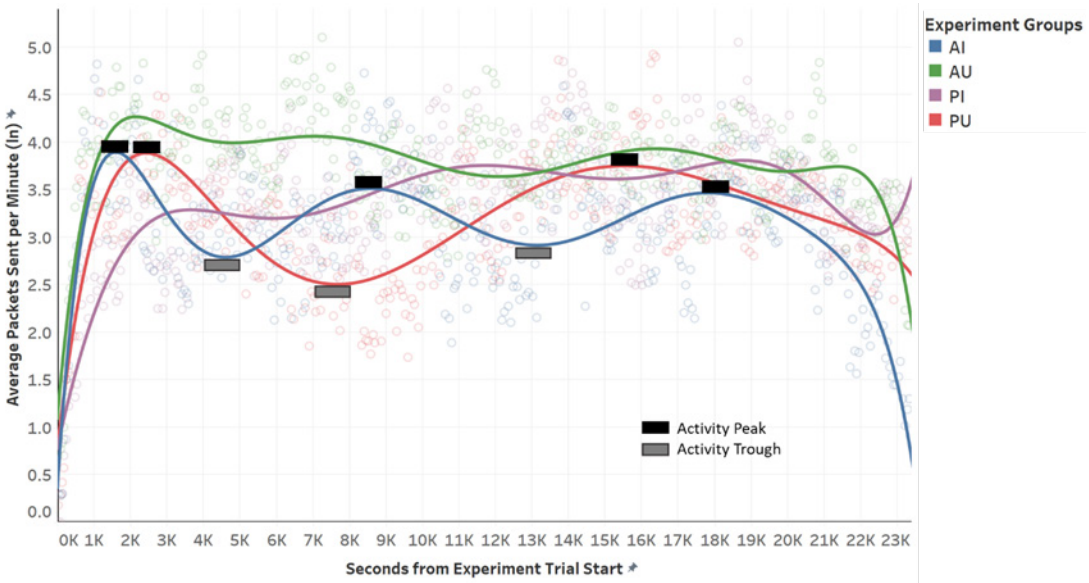


Figure 95. Day 1 Real System Average Packets Per Minute Activity Peaks and Troughs

On the same graph for the Present-Uninformed group a peak in network activity was observed at approximately 2,500 seconds from the experiment start, a trough in activity at 7,500 seconds, and another peak in activity at 15,500 seconds before decreasing through the end of the experiment.

While the *Belief* and *Plausibility* measures provided by DST offer a suitable explanation for network behavior observed for the Absent-Uninformed and Present-Informed groups, with gradual decreases in activity and increases in activity respectively, the same cannot be said for the Absent-Informed and Present-Uninformed groups. For these groups, variations in network activity over time cannot be explained by the accumulation of evidence over time. In hindsight, this outcome should not have come as a surprise given that the basis for DST, a generalization of Bayesian probability, is the Dempster Rule of Combination. This rule is centered on a conjunctive operation with a belief function "based on conjunctive pooled evidence" (Sentz & Ferson, 2002, p. 16). Considerable criticism has been levied against the utility of DST when a conflict in information between sources exists, which is present for the Absent-Informed and Present-Uninformed groups (Sentz & Ferson, 2002). In exploring potential explanations for the oscillation observed, a different approach was needed.

C. QUANTUM PROBABILITY THEORY-BASED MODEL

This section of the dissertation explores the phenomenon of cyber deception through quantum models of cognition. It suggests that principles offered by quantum cognition for understanding decision-making under uncertainty, which are based on QPT, may provide new insights into the complex phenomenon of cyber deception.

1. Introduction to Quantum Probability Theory

Following the development of vector-based models and assessment of DST, other theories for human decision-making under uncertainty were explored. This investigation exposed a vigorous, ongoing discourse on the limitations of Bayesian statistics in explaining human cognitive processes. Classical Probability Theory (CPT), for instance, leverages Bayesian statistics to provide powerful techniques for exploring event probabilities; however, persistent challenges have been levied against CPT axioms in

accounting for the conjunction fallacy and other situations involving cognitive interference (Pothos & Busemeyer, 2022).

Trueblood and Busemeyer (2011) suggest four reasons to consider QPT-based approaches in modeling human decision making over CPT-based approaches:

> (a) judgment is not a simple readout from a preexisting or recorded state, but instead it is constructed from the current context and question; from this first point it then follows that (b) making a judgment changes the context which disturbs the cognitive system; and the second point implies that (c) changes in context produced by the first judgment affects the next judgment producing order effects, so that (d) human judgments do not obey the commutative rule of classic probability theory. (p. 1519)

With the uncertainty and complexity inherent in performing reconnaissance and exploitation activities against remote network systems, subjects in the Tularosa study likely experienced one or more of these situations.

2. Interference Effect

Given the complexity of cognitive processes involved in the conduct of cyber-attack behavior, determining the probability an attacker accurately or incorrectly identified a target is exceedingly challenging, but the application of QPT axioms provide a novel approach to delve into this complexity (Al-Shamisi, 2014; Pothos & Busemeyer, 2022). QPT-derived models present a new direction to explain psychological processes in specific situations. These situations include where interference between different possibilities exist, where inference is context-sensitive, and where intermediate judgements or choices can influence the mental state for subsequent judgements (Pothos & Busemeyer, 2022). Cognitive interference is typically observed in situations which the decision maker cannot form a simultaneous cognitive representation of the phenomenon of interest.

The interference effect occurs when there is a violation in the law of total probability, with the probability of a single event estimated alone being different than when the original event is measured along with another event i.e., $p(A) \neq p(A \cap B) + p(A \cap \sim B)$ (Busemeyer & Wang, 2015). QPT offers that prior to making a complex decision, an individual may be in a state of superposition, with all possible decision outcomes having potential to occur, representing uncertainty or conflict that is often experienced when

making a decision (Busemeyer & Wang, 2015). A state of superposition allows all possible outcomes to be considered, with potential for these outcomes to cause interference between them, similar to the phenomenon of wave interference in physics, which can be destructive or constructive, and described with the axioms of QPT (Busemeyer & Wang, 2015).

Busemeyer and Wang (2015) suggest that the categorization experiment depicted in Figure 96 provides evidence of the interference effect, where requiring an explicit categorization (categorization-then-decision) versus not requiring an explicit categorization (decision-alone) concerning a face being a "good guy" or "bad guy" changes the total probability for the subject deciding to withdraw or attack. By making subjects categorize the face prior to making this decision, the total probability of an attack decision changes from 0.69 to 0.59, a significant impact by merely making the subject perform an explicit intermediate judgement.

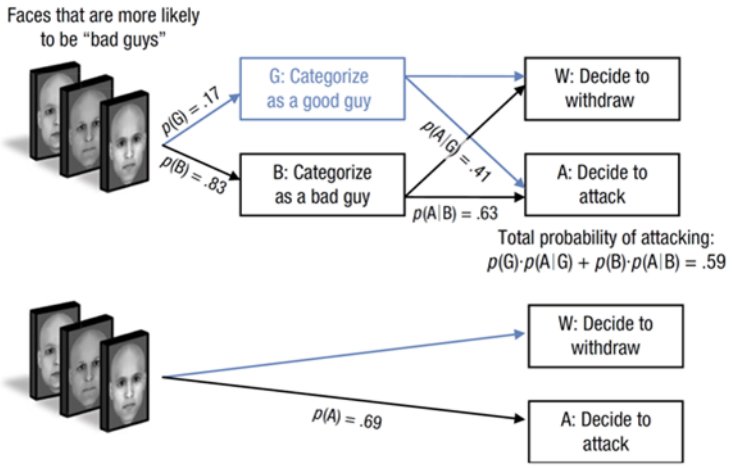


Figure 96. Illustration of the categorization-decision interference effect. Source: Busemeyer & Wang (2015).

As described in the Busemeyer and Wang study (2015) and more recent work by Zhen, Busemyer & Nosofsky (2023), for the Tularosa study subjects in the Absent-Uninformed and Present-Uninformed groups would be similar to the decision-alone condition, remaining in a state of superposition, considering all possible outcomes, with potential for these outcomes to cause interference between each other until a decision is

made to send a packet to a target system. In contrast, subjects in the Absent-Informed and Present-Informed groups would be similar to the categorization-then-decision condition, having been advised in the Task Briefing that "deception may be present." Subjects in the in the informed condition would likely characterize passively available information from systems on the cyber range environment as being "good" (generated by real system) or "bad" (generated by a decoy system) prior to deciding to send a packet to target systems.

Unlike the Busemeyer and Wang study (2015), which captured the probability of attacking or withdrawing, the network traffic from Tularosa study captured the timing of attacker interaction with target systems. However, a similar impact is observed, with higher median latency observed for the categorization then decisions conditions (Absent-Informed and Present-Informed) compared to the decision-alone conditions (Absent-Uninformed and Present-Uninformed), which is analogous to the decreased attack probability identified by Busemeyer and Wang (2015). Increased latency, time elapsed from experiment start to initial system interaction, represents a decreased probability for initial system interaction for a given time interval. For the Tularosa study, there is a 135 second increase in median latency between the Absent-Informed group (1127 seconds) and the Absent-Uninformed group (992 seconds) and a 301 second increase in median latency between the Present-Informed group (1369 seconds) and the Present-Uninformed group (1068 seconds). The 2.2x increase between the groups where cyber deception is Absent (135 seconds) and where cyber deception is Present (301 seconds) is comparable to the 2x increase in the number of target systems from the Absent cyber range network, which had 50 target systems, and the Present cyber range network, which had 100 target systems.

3. Order Effect

QPT-based models can also help explain the impact of the order of information provided in decision-making. The order effect is a well-documented phenomenon in psychology where the order in which questions are asked or information is provided impacts the outcome (Pothos, 2022). CPT is challenged in explaining the order effect because of the commutative nature of conjunction in CPT where P(A&B) = P(B&A) (Pothos, 2022). In the Tularosa study information was manipulated through the use of

psychological deception, information regarding the presence of deception, and cyber deception, the presence or absence of decoy systems on the cyber range environment. For the subjects in the Absent-Informed and Present-Informed groups, the order of these information inputs was explicit. In the task briefing provided to these groups, subjects were informed that "deception may be present" which occurred prior to their access to the cyber range environment and potential identification of decoy systems (Ferguson-Walter et al., 2021). This order of information is depicted in Figure 97.

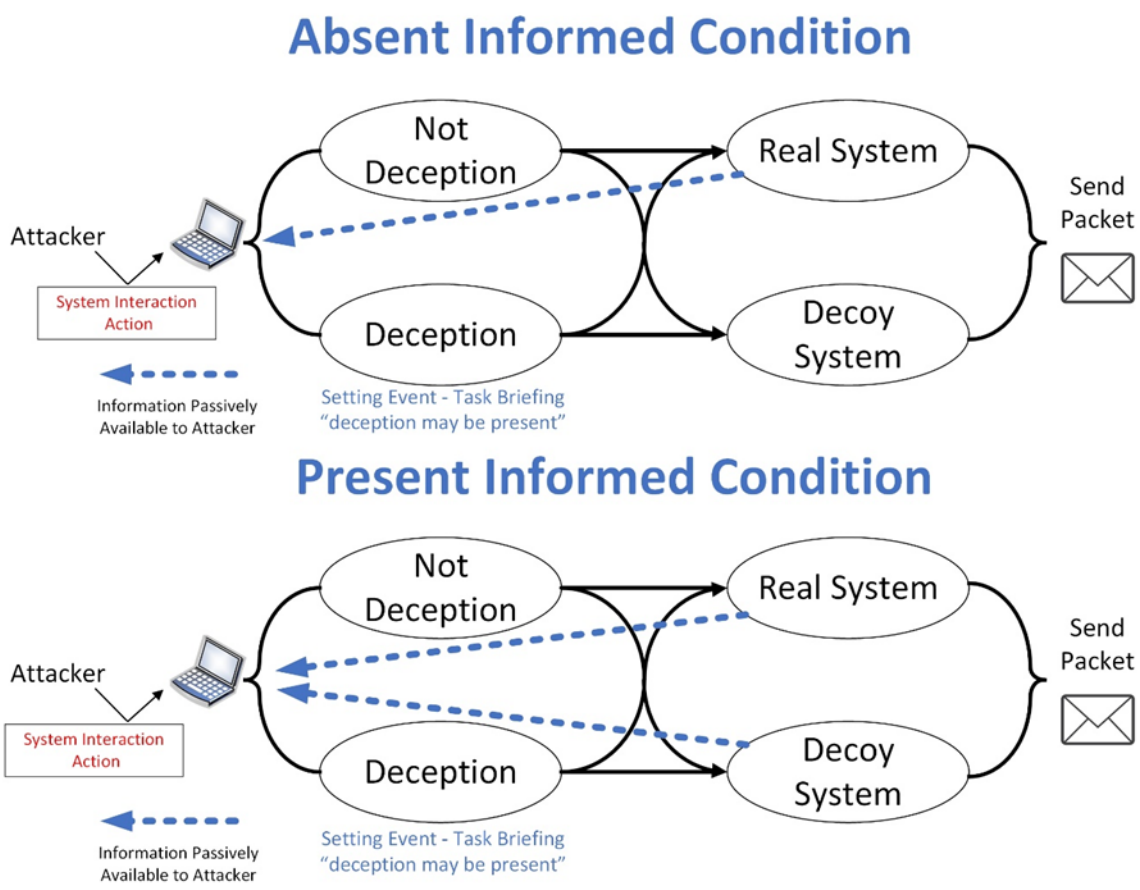


Figure 97. Presumed Information Order for AI and PI Groups

For the Absent-Uninformed and Present-Uninformed groups, information regarding the presence of deception was not explicitly provided. While it is not possible to ascertain if a subject considered the presence of deception prior to interaction with systems on the cyber

range environment, the presumed order of information considered for the Absent-Uninformed and Present Uninformed groups is depicted in Figure 98.

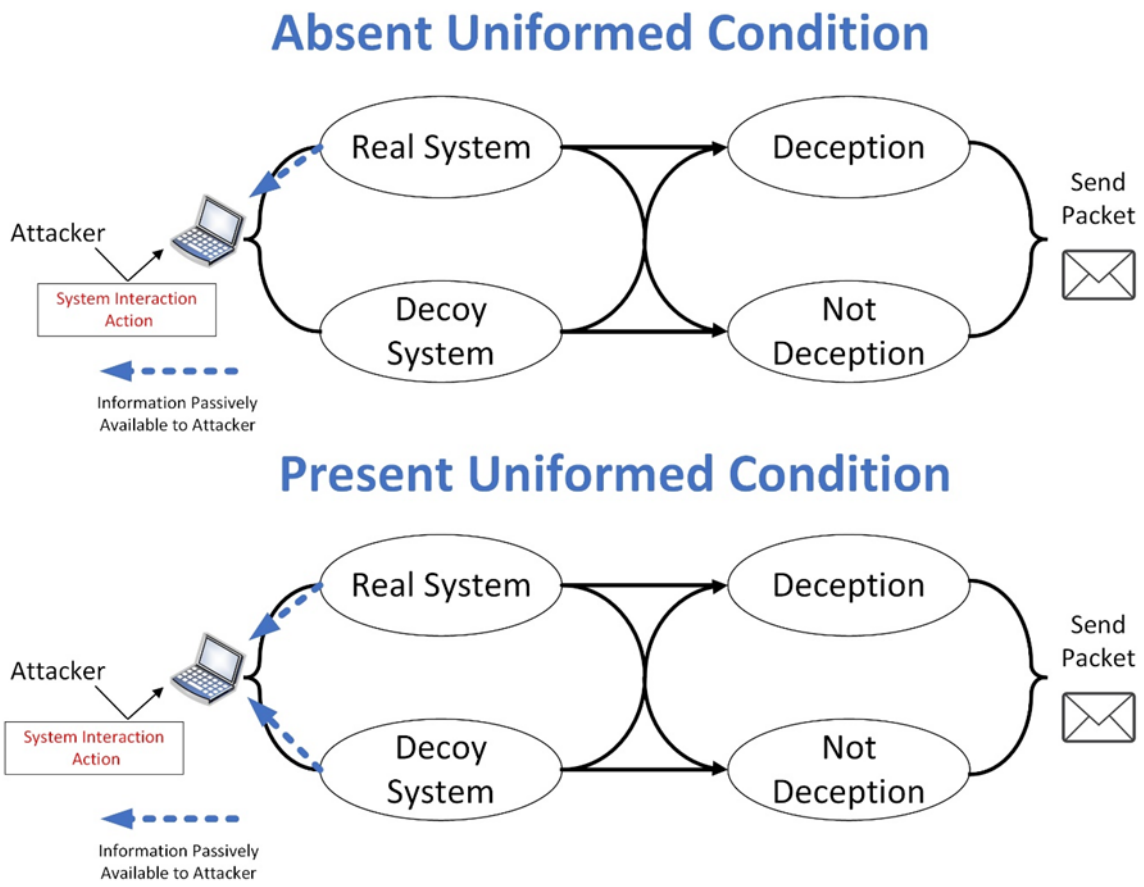


Figure 98. Presumed Information Order for AU and PU Groups

Trueblood and Busemeyer offer a quantum model example similar to the information ordering effect described in Figures 97 and 98 using a 1998 study that examined the impact of the ordering of clinical data on the resulting diagnosis for a Urinary Tract Infection (UTI) (Trueblood & Busemeyer, 2011). In this study, physicians were presented with physical examination and laboratory information in different orders for the diagnosis of an infection with the physical examination information providing evidence of an infection and the lab information did not provide evidence of an infection (Trueblood & Busemeyer, 2011). Physicians presented with the physical examination information first

followed by the lab information had a 9% lower probability (p = .03) of diagnosing an infection compared with the physicians who were presented the lab information first followed by the physical examination information (Trueblood & Busemeyer, 2011). In the quantum-based model developed by Trueblood and Busemeyer (2011), the infection diagnosis is presented as two complementary hypotheses, present or absent, and the physical examination as positive evidence for an infection being present and the lab information as being negative evidence for an infection. For this medical decision-making task, the state vectors for the diagnosis beliefs were represented using the equation in Figure 99.

$$\omega = \begin{bmatrix} \omega_{h_1,e_1} \\ \omega_{h_1,e_2} \\ \omega_{h_2,e_1} \\ \omega_{h_2,e_2} \end{bmatrix} = \begin{bmatrix} \text{amplitude for UTI present and positive evidence} \\ \text{amplitude for UTI present and negative evidence} \\ \text{amplitude for UTI absent and positive evidence} \\ \text{amplitude for UTI absent and positive evidence} \end{bmatrix}$$

Figure 99. Belief State Vector for Medical Decision-Making Task. Source: Trueblood and Busemeyer (2011, p. 1523).

Adapting this equation to the ordering effect depicted in Figures 97 and 98 provides the belief state vector for the experimental conditions for the Tularosa study, with the complementary hypotheses of trustworthy information and not trustworthy information and initial evidence of real system or decoy system, generating a new set of amplitudes adapted from Trueblood and Busemeyer as illustrated in Figure 100.

$$\omega = \begin{bmatrix} \omega_{h_1,e_1} \\ \omega_{h_1,e_2} \\ \omega_{h_2,e_1} \\ \omega_{h_2,e_2} \end{bmatrix}$$

$$=$$

$$\begin{bmatrix} amplitude\ for\ deception\ present\ (informed)\ and\ positive\ evidence\ (decoy) \\ amplitude\ for\ deception\ present\ (informed)\ and\ negative\ evidence\ (real) \\ amplitude\ for\ deception\ absent\ (uninformed)\ and\ positive\ evidence\ (real) \\ amplitude\ for\ deception\ absent\ (uninformed)\ and\ negative\ evidence\ (decoy) \end{bmatrix}$$

$$=$$

$$\begin{bmatrix} Trustworthy\ Information\ and\ then\ Decoy\ System \\ Not\ Trustworthy\ Information\ and\ then\ Real\ System \\ Real\ System\ and\ then\ Trustworthy\ Information \\ Decoy\ System\ and\ then\ Not\ Trustworthy\ Information \end{bmatrix}$$

$$=$$

$$\begin{bmatrix} Present - Informed \\ Absent - Informed \\ Absent - Uninformed \\ Present - Uninformed \end{bmatrix}$$

Figure 100. Belief State Vector for Tularosa Study Decision-Making Task. Adapted from Trueblood & Busemeyer (2011).

4. QPT State Vector Representation

Visually the probability amplitude for sending or not sending a packet as a result of these two question outcomes can be illustrated with three sets of subspaces, one

depicting the decision between a real system or a decoy system, a second decision on if the information provided is not trustworthy (psychological deception) or trustworthy (not psychological deception) and a third depicting the decision to send or not send a packet. Without the uncertainty generated by these two questions, calculation of the probability of sending a packet would be calculated by projecting a ray from the state vector onto the "Send" vector to calculate length *b* as depicted in Figure 101. Length *b* is then squared to provide the probability of sending a packet, or p(Send).

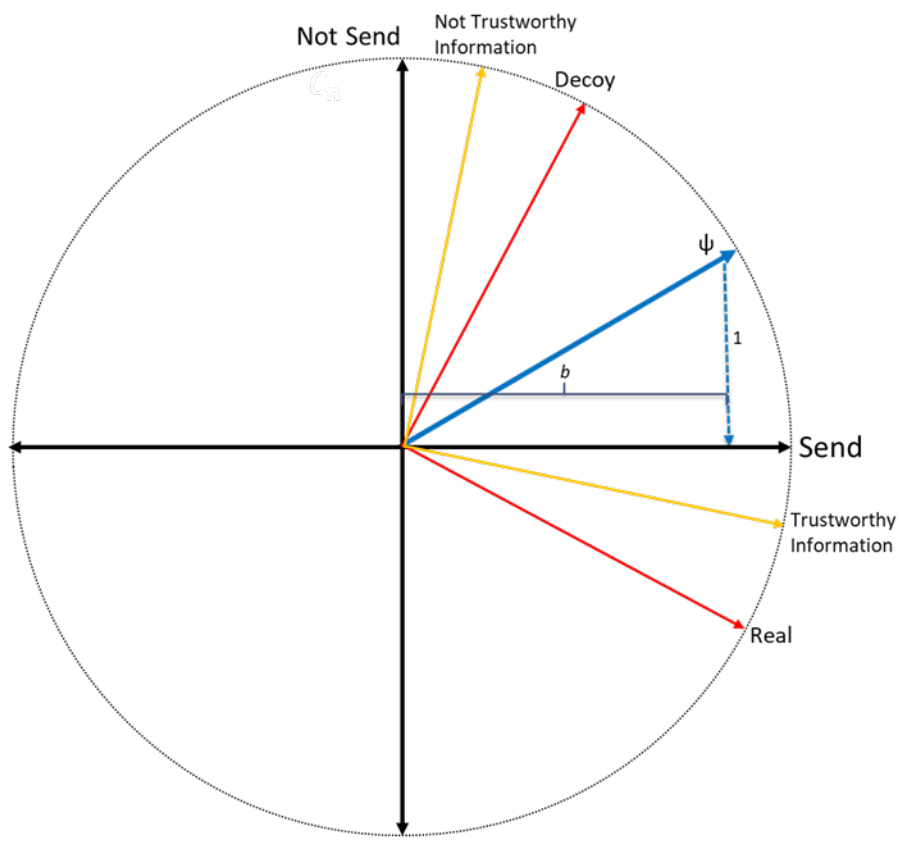


Figure 101. QPT Representation of Sending Packet Without Uncertainty

For the Absent-Uninformed condition, computing the amplitude for sending a packet to a real system with the decision-making path of real system and then Trustworthy Information is calculated by projecting the state vector onto the "Real" ray, then projecting a second line on to the "Trustworthy Information" ray and finally to the "Send" ray to

calculate length *a* as depicted in Figure 102. Length *a* is then squared to provide the probability of sending a packet, or p(Send).

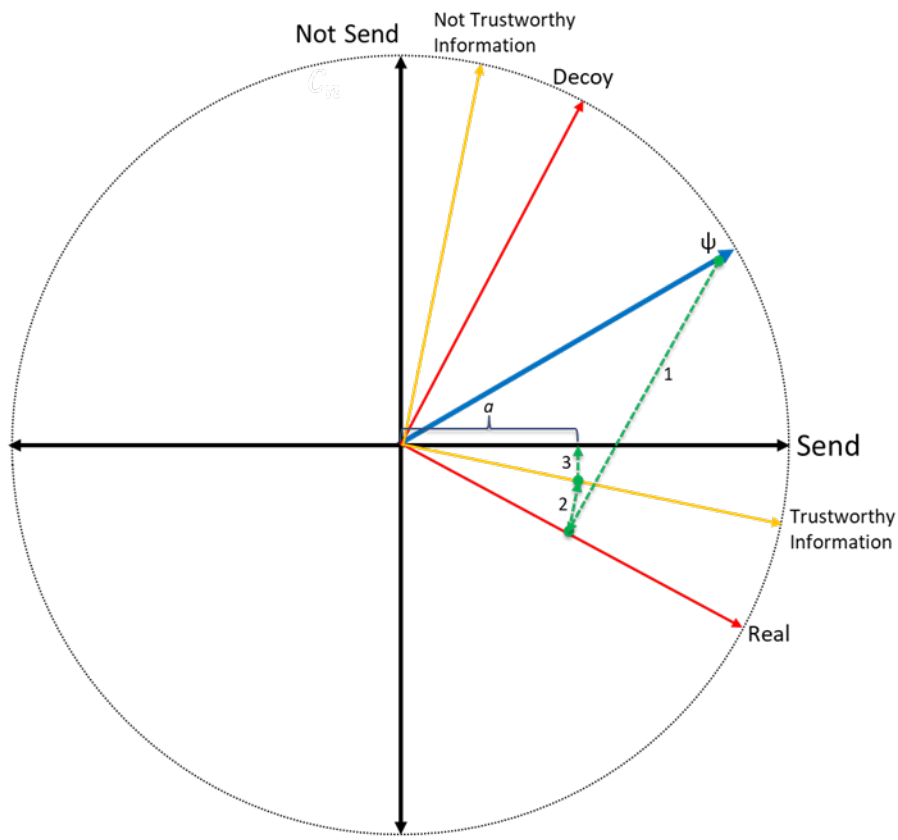


Figure 102. Absent-Uninformed QPT Representation

Similarly for the Absent-Informed condition computing the amplitude for sending a packet to a real system with the decision-making path of Not Trustworthy Information then real system is calculated by projecting the state vector onto the "Not Trustworthy Information" ray, then projecting a second line on to the "Real" ray and finally to the "Send" ray to calculate length *a* as depicted in Figure 103. Length *a* is then squared to provide the probability of sending a packet, or p(Send).

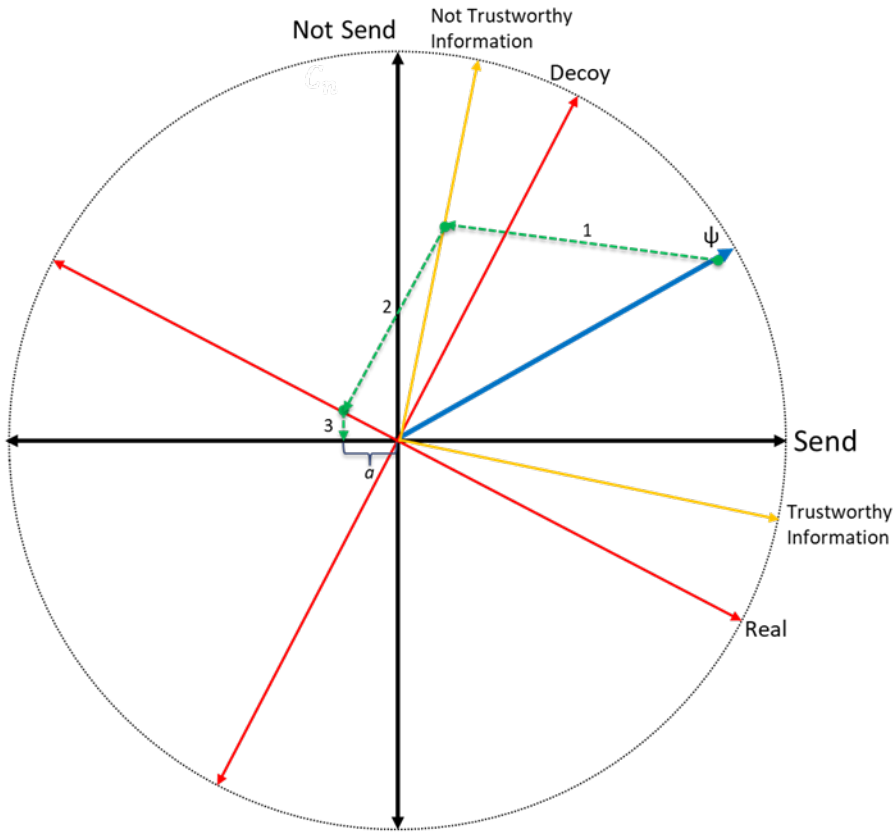


Figure 103. Absent-Informed QPT Representation

Modeling the order effect for the Present-Uninformed and Present-Informed groups are performed in a similar manner as depicted in Figures 102 and 103. For the Present-Uninformed condition computing the amplitude for sending a packet to a decoy system with the decision-making path of decoy system and then Not Trustworthy Information is calculated by projecting the state vector onto the "Decoy" ray, then projecting a second line on to the "Not Trustworthy Information" ray and finally to the "Send" ray to calculate length *a* as depicted in Figure 104**.** Length *a* is then squared to provide the probability of sending a packet, or p(Send).

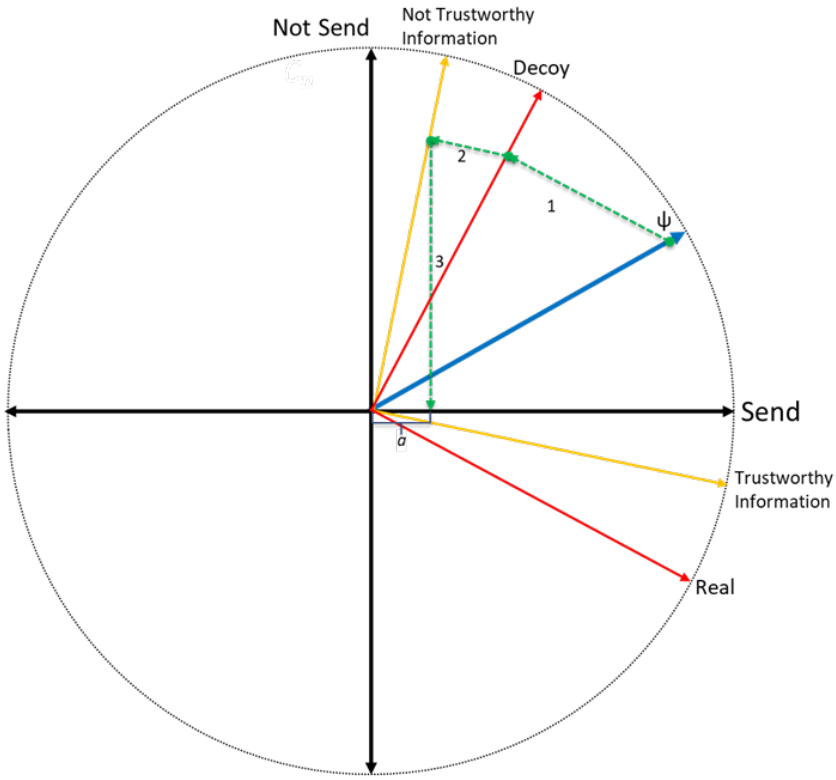


Figure 104. Present-Uninformed QPT Representation

Finally for the Present-Informed condition computing the amplitude for sending a packet to a decoy system with the decision-making path of Trustworthy Information then decoy system is calculated by projecting the state vector onto the "Trustworthy Information" ray, then projecting a second line on to the "Decoy" ray and finally to the "Send" ray to calculate length *a* as depicted in Figure 105. Length *a* is then squared to provide the probability of sending a packet, or p(Send).

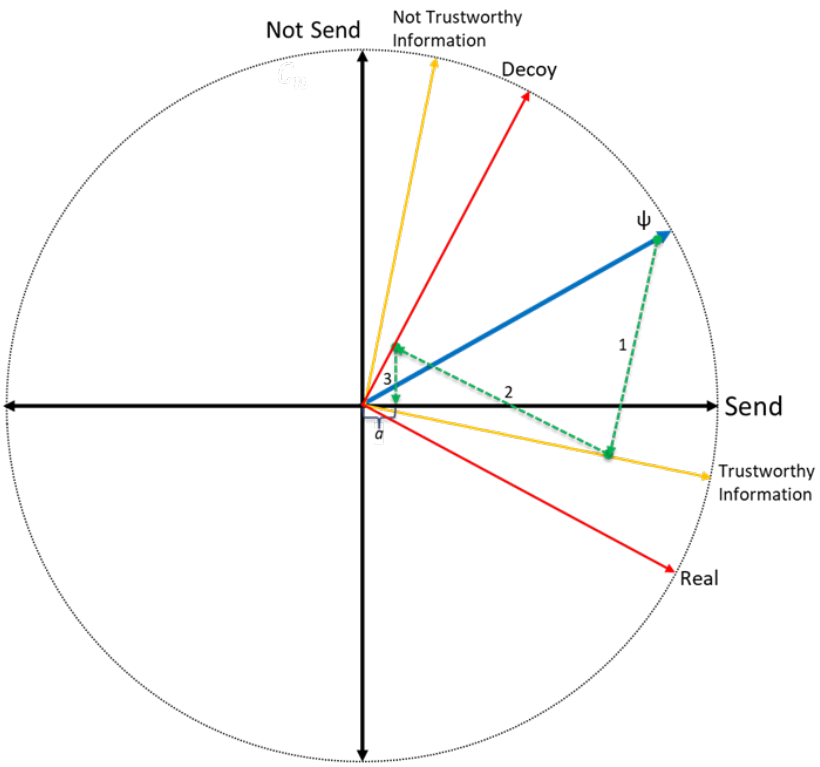


Figure 105. Present-Informed QPT Representation

While QPT provides a framework for investigating decision-making under uncertainty, prior to a decision being made, the state vector is a superposition of subspace possibilities (Pothos & Busemeyer, 2013). As such only the tendency or potential for a decision can be described (Pothos & Busemeyer, 2013). The key insight provided by QPT is that the decision that resolves the superposition of the state vector is a construct of the question and current context. This superposition state also represents the ontic uncertainty describing an individual's internal uncertainty on different decisions that could be made in a specific moment (Kvam et al., 2021). A QPT-based open system model has been developed to account for the interplay between epistemic uncertainty and ontic uncertainty and its impact on expressed preferences which combine aspects of both quantum and classical dynamics (Kvam, Busemeyer & Pleskac, 2021). The analysis of the network traffic generated during Tularosa study in this research provides a starting point for further exploration of the phenomenon of cyber deception using the open system model developed by Kvam, Busemeyer and Pleskac (2021) to generate new insights.

VIII. CONCLUSION

A. KEY FINDINGS

Analysis of network traffic generated by human subjects during the Tularosa study supports previous assertions that decoy-based deception is significantly more effective against attackers when the presence is known, with statistically significant differences for latency, frequency and duration between the control group with the Absent-Uninformed condition and experimental group with the Present-Informed condition. This analysis supports previous assertions by Heuer, Yuill et al. and Ferguson-Walter et al. that the combination of cyber and psychological deception has the greatest impact on the timing of cyber-attacker behavior (Heuer, 1982; Yuill et al., 2007; Ferguson-Walter et al., 2021). Additionally, the first ever analysis of Day 2 data captured during the Tularosa study suggests an enduring impact on cyber-attackers following an experience involving both psychological and cyber deception, with the original Present-Informed group having the highest median latency for Day 2 despite the absence of deception. This analysis revealed distinct temporal patterns within the aggregated network traffic for each of the four experimental groups, which were present in multiple graphs of the network traffic aggregated by group. These patterns provide additional empirical evidence on the impact of varying conditions of psychological and technical deception on cyber-attacker behavior. This research also provides the first application of factor analysis to the Tularosa study data, using both PCA and EFA to explore the variance observed, which highlighted the key role of technical experience in explaining the variance observed. Finally, a quantum probability theory-based model was developed to help explore findings from statistical and graphical analysis. While exploratory, quantum probability theory provides a new direction to help explain human decision-making under conditions experienced by subjects during the Tularosa study, with potential impact from both interference effect and ordering effect phenomena.

B. REASSESSMENT OF TULAROSA STUDY HYPOTHESES

Three hypotheses offered by Ferguson-Walter et al. (2021) previously described were reassessed using the extracted temporal variables from Day 1 to evaluate the effectiveness of the presence and awareness of cyber deception using a multitude of measures including decoy alerts, exploitation attempts, keystroke counts and byte counts. Table 93 provides a listing of all comparison groups for each variable for each hypothesis where the p-value from Dunn's Post Hoc Test with Benjamini-Hochberg correction was less than 0.05. Table 93 and Figure 106 provide a concise summary of hypotheses supported by each temporal variable, for real systems interaction, decoy systems interaction and real and decoy systems interactions for the reconnaissance phase of the cyber kill chain using the same framework offered by Ferguson-Walter et al. (2021).

Table 93. Day 1 Summary of Statistical Analysis Results

Hypothesis H1: Cyber and psychological deception impedes attackers. Comparison Groups: Control (AU) versus Experimental (AI, PU, PI)

Real System Interactions:

Metric	Data Source	Lower Median	Higher Median	p-value
Latency	Network Capture	Absent-Uninformed	Absent-Informed	p = 0.0022 **
Latency	Network Capture	Absent-Uninformed	Present-Informed	p = 0.0017 **
Frequency	Network Capture	Absent-Uninformed	Absent-Informed	p = 0.0000 ***
Frequency	Network Capture	Absent-Uninformed	Present-Informed	p = 0.0000 ***
Duration	Network Capture	Present-Informed	Absent-Uninformed	p = 0.0043 **
Duration	Network Capture	Present-Uninformed	Absent-Uninformed	p = 0.0006 ***

Present-Informed condition resulted in 6.75% reduction in real system interaction compared with Absent-Uninformed groups.

Hypothesis H2: Cyber deception tools are effective even if an attacker is aware of their use. Comparison Groups: Uninformed (AU, PU) versus Informed (PI)

Real System Interactions:

Metric	Data Source	Lower Median	Higher Median	p-value
Latency	Network Capture	Absent-Uninformed	Present-Informed	p = 0.0017 **
Frequency	Network Capture	Absent-Uninformed	Present-Informed	p = 0.0000***
Frequency	Network Capture	Present-Uninformed	Present-Informed	p = 0.0000 ***
Duration	Network Capture	Present-Informed	Absent-Uninformed	p = 0.0043 **

Present-Informed condition resulted in 6.75% reduction in real system interaction compared with Absent-Uninformed group and 7% reduction compared with Present-Uninformed group.

Decoy System Interactions:

Metric	Data Source	Lower Median	Higher Median	p-value
Latency	Network Capture	Present-Uninformed	Present-Informed	p = 0.0100 **
Frequency	Network Capture	Present-Uninformed	Present-Informed	p = 0.0000 ***

Real and Decoy System Combined Interactions:

Metric	Data Source	Lower Median	Higher Median	p-value
Latency	Network Capture	Present-Uninformed	Present-Informed	p = 0.0020 **
Frequency	Network Capture	Present-Uninformed	Present-Informed	p = 0.0000 ***

Hypothesis H3: Cyber deception is effective if the attacker merely believes it may be in use, even if it is not. Comparison Groups: Control (AU) versus Psychological Deception (AI)

Metric	Data Source	Lower Median	Higher Median	p-value
Latency	Network Capture	Absent-Uninformed	Absent-Informed	p = 0.0022 **
Frequency	Network Capture	Absent-Uninformed	Absent-Informed	p = 0.0000 ***

Summary of Findings: Significant differences are indicated as ***p < .001, ** p < 0.01, * p < 0.05

Figure 106 provides a concise summary of reconnaissance progress results from analysis performed on network traffic generated by subjects during the Tularosa study.

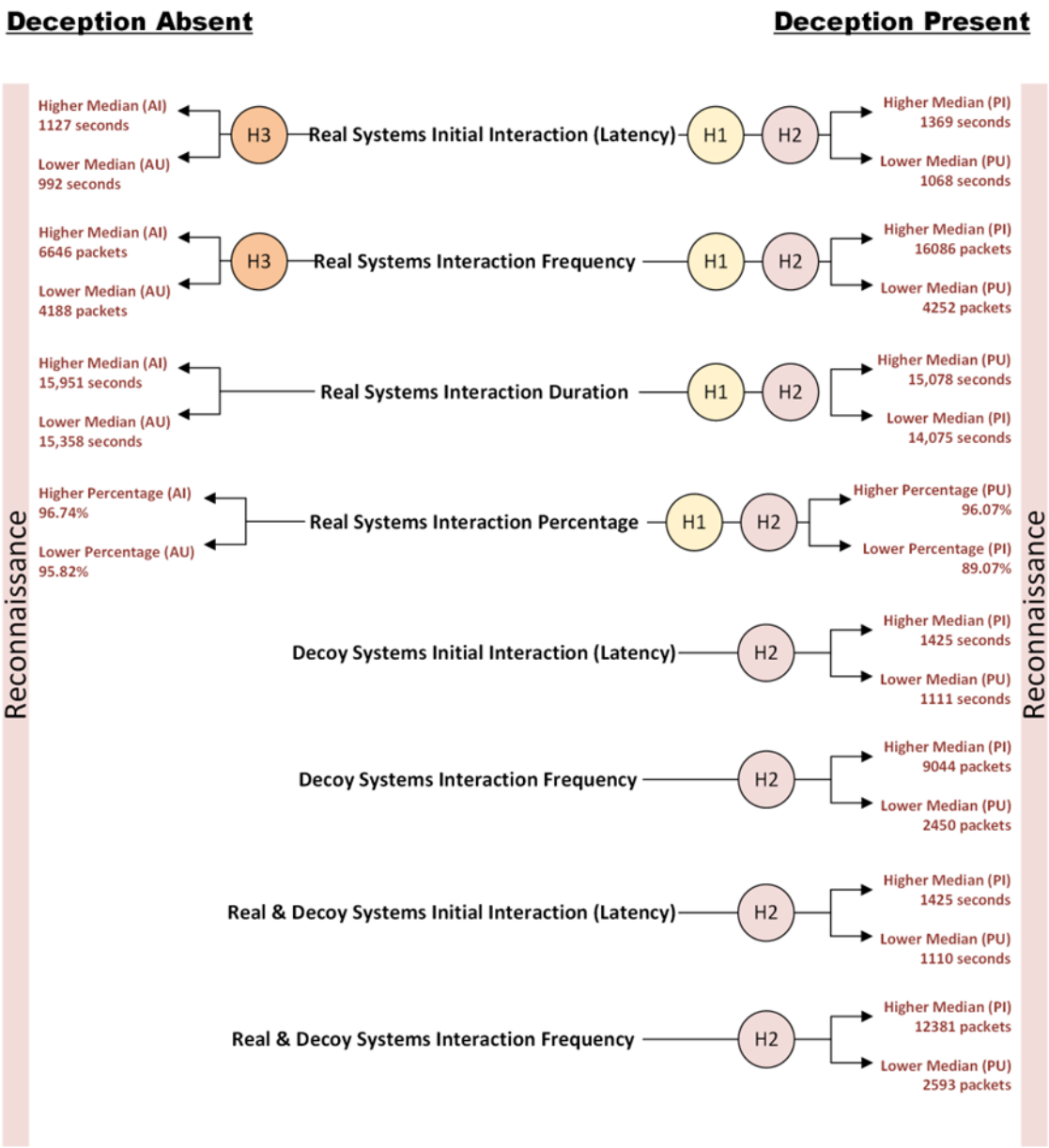


Figure 106. Reconnaissance Progress Results Summary. Adapted from Ferguson-Walter et al. (2021).

C. MEASURES OF SUCCESS

In "Examining the Efficacy of Decoy-based and Psychological Cyber Deception" Ferguson-Walter et al. (2021) provide a framework for measures of "success" using metrics for forward progress, wasted effort, and altered perception. This framework and definitions

provided are used for further assessment of the temporal variables for both Day 1 and Day 2 activity.

1. Day 1 Analysis

a. Forward Progress

Forward progress was measured by Ferguson-Walter et al. as strategic gains obtained as a cyber actor advance within a target network, using the Cyber Kill Chain® model to define cyber-attack stages (2021). For this research exploring temporal variables, only stage one, reconnaissance was considered. Success from a defender's perspective for reconnaissance results from wasted cyber-attacker resources in the form of time or effort. All condition group comparison statistics use the Kruskal-Wallis test, with group pairs using Dunn's post hoc test with a Benjamini-Hochberg correction as described in Chapter V.

(1) Increased Latency

Analysis of latency for all real system interactions revealed that the Present-Informed group had the highest median latency (PI, median = 1369) of all groups ($H(3) = 43.48$, $p < 0.001$). The Dunn's post hoc test demonstrated a significant difference between the Absent-Uninformed group (AU, median = 992) and Present-Informed group at the $p = 0.0017$ level and between the Absent-Uninformed and Absent-Informed group (AI, median = 1127) at the $p = 0.0022$ level. These statistically significant findings indicate that cyber-attackers delayed initial interaction with real systems under conditions of psychological deception and the combination of psychological and cyber deception, wasting cyber-attacker time resources as they contemplated information regarding deception.

(2) Decreased Duration

Analysis of duration for all real system interactions revealed that the Present-Informed group had the lowest median latency (PI, median = 14,075) of all groups ($H(3) = 17.91$, $p < 0.0015$). The Dunn's post hoc test revealed a statistically significant difference between the Absent-Uninformed (AU, median = 15,358) and Present-Informed group at the $p = 0.0043$ level and between the Absent-Uninformed and Present-Uninformed group

at the p = 0.006 level. These statistically significant findings indicate that cyber-attackers reduced overall length of interaction with real systems under conditions of cyber deception and the combination of psychological and cyber deception. This decrease in duration is desirable from a cyber-defender point as decreased length of interaction with a system reduces a cyber-attacker's opportunity to identify vulnerabilities. While the Tularosa study was conducted with a strict, artificial time limit, decreased duration of activity by a similarly time-limited cyber-attacker is beneficial to the defender by reducing opportunities for cyber-attacker reconnaissance to identify potential vulnerabilities.

(3) Reduced System Interaction

Analysis of the overall percentage of real system interactions revealed that the Present-Informed group experienced a nearly 7% reduction in the percentage of real system interactions. The Present-Informed group displayed the lowest overall percentage of interactions with of only 89.07% of real systems present having subject interaction compared to 95.82% in the Absent-Uninformed group, 96.07% in the Present-Uninformed group and 96.74% in the Absent-Informed group. This finding suggests that the combination of psychological and cyber deception results in cyber-attackers performing reconnaissance activity more cautiously, thus reducing overall interaction with systems during a time-constrained scenario similar to conditions presented during the Tularosa study.

(4) Delayed Activity

Analysis of the subject group network activity using Figures 27 and 28, the Present-Informed group displayed significant reduction in both average and median packets per minute compared to Absent-Informed, Absent-Uninformed and Present-Uninformed groups until approximately 3,500 seconds from the start of the experiment trial. Present-Informed group average and median packets continued to increase until approximately 12,000 seconds from the start of the experiment trial, unlike the Absent-Informed and Present-Uninformed groups which oscillated following initial spikes in average and median packets occurring at 1,500 and 2,000 seconds respectively from the start of the experiment trial. These graphs highlight delayed network activity by cyber-attackers under

the condition of both psychological and cyber deception, which was not observed for the use of these deception techniques individually.

b. Attacker Resources Expended

(1) Increased Frequency

Analysis of the frequency for all real system interactions revealed that the Present-Informed group had the highest median frequency (PI, median = 16,086) of all groups ($H(3) = 150.22$, $p = 0.000$). The Dunn's post hoc test revealed a statistically significant difference between the Absent-Uninformed (AU, median = 4,188) and Present-Informed group at the $p = 0.000$ level and between the Absent-Uninformed and Absent-Informed group (AI, median = 6,646) at the $p = 0.000$ level. These statistically significant findings indicate that cyber-attackers increase the number of packets sent to real systems under conditions of psychological deception and the combination of psychological and cyber deception. While an increase in the number of packets sent to a target system provides a cyber-attacker with additional information from these interactions, increased levels of network activity provide cyber-defenders more opportunities to identify cyber-attacker reconnaissance.

c. Altered Perception

In the Present-Informed group, both average packets, median packets and percentage of subjects sending packets from 2 to 403 packets per minute increase monotonically as depicted in Figures 27, 28, 31, 32, 33, 34, 35 and 36 respectively. However, for the Present-Uninformed and Absent-Informed groups, oscillations occur for both average packets, median packets and percentage of subjects sending packets illustrated in the same figures. The oscillations in subject activity for the Present-Uninformed and Absent-Informed groups may indicate an impact on the attacker's cognitive and/or emotional state over time for these groups, potentially supporting Hypothesis H4. The incongruity between information provided about the cyber deception presence and actual presence of cyber deception could potentially explain the oscillating patterns of subject activity.

The difference in behavior patterns between Absent-Informed and Present-Informed groups is also noteworthy. Given subjects in both groups start at the same state of uncertainty it would be reasonable to expect that subject activity between the two groups would exhibit similar initial behavior patterns. However, subject activity by the Present-Informed group is significantly delayed compared to the Absent-Informed group. This difference in behavior patterns may be explained by the congruence of information about deception and the presence of cyber deception. Subjects conducting passive reconnaissance of the network traffic on the cyber range network may have identified the presence of one or more decoy systems potentially resulting in initial caution, which subsided over time as depicted in average and median packets sent per minute by subjects as illustrated in Figures 27 and 28.

Cyberpsychology methods may have the impact of increasing model uncertainty by causing cyber-attackers to question the suitability of the conceptual frame initially applied to model the cyber range network. For the Absent-Informed group in Day 1, this doubt of conceptual frame suitability appears to occur at approximately 1,500 seconds from the start of the experiment trial with the peak of subject activity illustrated in peak packets for both the average and median depicted in Figures 28 and 29, as well as the percentage of subjects reaching packet thresholds specified in Figures 31, 32, 33, 34, 35 and 36. For the Present-Uninformed group in Day 1, this doubt of conceptual frame suitability appears to occur at approximately 2,000 seconds from the start of the experiment trial as depicted in Figures 27 and 28, as well as the percentage of subjects reaching packet thresholds illustrated in Figures 31, 32, 33, 34, 35 and 36.

d. Factor Analysis

The use of PCA provided a reduction from 16 variables analyzed to 5 principal components, which accounted for more than 70% of the total observed variance across all experiment groups for both Day 1 and Day 2 experiment trials. PCA of the Day 1 experiment groups highlighted that self-reported experience levels comprise the majority of variables included in the first principal component, which accounted for over 30% of total observed variance across each group. Self-reported experience levels had component

loading levels averaging 0.75 and exceeding 0.50 for all self-reported experience levels across experiment groups, with a single exception of *Incidence response* for the Absent-Uninformed group, with a loading of 0.45. For the Absent-Uninformed and Present-Uninformed groups, self-reported *Confidence* had a loading of 0.619 and 0.759 respectively. In contrast, self-reported *Confidence* had a loading of less than 0.30 for the Absent-Informed group and a loading of 0.350 for the Present-Informed group, which suggests the absence of the *Confidence* variable in explaining shared variance for these two groups. Differences in variable loadings for components across each experiment group suggest varying impact from the manipulation of cyber and psychological deception as reflected by shared total variance explained by these loadings. EFA also highlights significant loadings for self-reported experience levels in explaining common variance for the first factor "Technical Aptitude" across experiment group groups as well. Differences in variable loading for factors across each experiment condition group were also observed with EFA.

2. Day 2 Analysis

a. Forward Progress

Research for this dissertation provided the first analysis of data from Day 2. Forward progress for Day 2 analysis uses the same definition and statistical techniques described previously for Day 1 analysis. A key consideration for the analysis of Day 2 data is the impact of repeated testing given subjects performed the same penetration testing task assigned on Day 1 on a cyber network range with the same configuration of real systems.

(1) Increased Latency

Analysis of all real system interactions for Day 2 experiment trials displayed a reduction in median latency for all groups, with the smallest decrease (223 seconds) occurring for the Absent-Uninformed/Present- Uninformed group and the largest decrease occurring for the Absent-Informed/Absent-Uninformed group (457 seconds). The Present-Informed/Absent-Uninformed group continued to have the highest median latency of all groups at 980 seconds ($H(3) = 150.22$, $p < 0.001$). The Dunn's post hoc test revealed there was no statistically significant difference between the Absent-Uninformed/Present-

Uninformed group and the Present-Informed/Absent-Uninformed groups, which suggests an enduring impact on cyber-attacker behavior following an experience with the combination of psychological deception and cyber deception even in the absence of deception in a subsequent penetration testing task.

(2) Decreased Duration

Analysis of all real system interactions for the duration for Day 2 experiment trials revealed that the Absent-Uninformed/Present-Uninformed group had the highest median duration of 15,826 seconds and that the Present-Informed/Absent-Uninformed group had the lowest median duration of 14,368 seconds ($H(3) = 29.18$, $p < 0.001$).). The Dunn's post hoc test revealed a statistically significant difference between the Absent-Uninformed/ Present-Uninformed group and the Absent-Informed/Absent-Uninformed group (15,110 seconds) at the $p = 0.002$ level, the Present-Informed/Absent-Uninformed group (14,619 seconds) at the $p = 0.001$ level, and the Present-Uninformed/Absent-Uninformed group at the $p = 0.002$ level. This suggests an enduring impact from psychological deception, cyber deception and the combination of psychological deception and cyber deception even in the absence of deception in a subsequent penetration testing task.

b. Attacker Resources Expended

(1) Increased Frequency

Analysis of all real system interactions for the frequency for Day 2 experiment trials revealed that the Absent-Informed/Absent-Uninformed group had the highest median frequency of 65,555 packets and that the Absent-Uninformed/Present-Uninformed group had the lowest median frequency of 4,154 packets ($H(3) = 252.43$, $p = 0.000$).). The Dunn's post hoc test revealed a statistically significant difference between the Absent-Uninformed/Present-Uninformed group and the Absent-Informed/Absent-Uninformed group at the $p = 0.000$ level, the Present-Informed/Absent-Uninformed group (12,132 packets) at the $p = 0.000$ level, and the Present-Uninformed/Absent-Uninformed group (12,025 packets) at the $p = 0.000$ level. This suggests an enduring impact from psychological deception, cyber deception and the combination of psychological deception and cyber deception even in the absence of deception in a subsequent penetration testing

task. As previously stated for analysis of Day 1 experiment trial data, increased levels of network activity provide cyber-defenders more opportunities to identify cyber-attacker reconnaissance, while potentially providing cyber-attackers with additional information from more numerous system interactions.

D. CONTRIBUTIONS

This research provides six main contributions to both scientific research and operational practice of the phenomenon of cyber deception: (1) bolstering previous assertions by Heuer (1982), Yuill et al. (2007), and Ferguson-Walter et al. (2021) that decoy-based deception is significantly more effective against attackers when presence is known by using a novel network traffic data-based analytic approach on PCAP data collected during the Tularosa study data; (2) development of an analytic methodology for cyber-attacker behavior leveraging data readily accessible to cyber-defenders (3) first analysis of Day 2 experiment trial data from the Tularosa study; (4) extension of cyber deception research through first application of factor analysis (PCA and EFA) on data collected during the Tularosa study; (5) introduction of a quantum probability theory-based model to advance nascent state of cyber deception theory; and (6) development of an IT artifact for aggregation and analysis of large volumes of network traffic data made openly available to researchers to support future computer network-based experimentation. Empirical assessment of the impact of defensive cyber deception on the timing of cyber attacker behavior against target systems was based on latency, duration and frequency variables extracted from network traffic data captured during the Tularosa study. Over 10,000 unique system interactions by more than 100 subjects across two days of experiments for each of these variables were extracted from the network traffic data and quantitatively analyzed. This analysis provides a comprehensive assessment of the impact of defensive cyber on the timing of cyber attacker behavior leveraging data readily available to network defenders. The first analysis of Day 2 experiment trial data from the Tularosa study suggests an enduring impact on cyber-attacker behavior following an experience with both psychological deception and cyber deception. The factor analysis applied to data from the Tularosa study provides significant insights into the role technical experience has in contributing to both total and common variance of the variables analyzed,

as well as providing an analytic template for future research. Finally, the methodology and IT artifact enables transformation of large volumes (200+ GB) of network traffic data (PCAP) captured during human subjects' experimentation into discrete experimental trial timelines for empirical assessment of their activity based on any attribute collected within the network traffic using a novel network-graph database approach.

E. FUTURE WORK

1. Quantum Random Walk Theory-Based Cyber-Attacker Decision-Making Model

Further exploration of the phenomenon of cyber deception through the "open system" model offered by Kvam, Busemeyer and Pleskac, which incorporate both epistemic and ontic uncertainty using a combination of Markov processes and quantum walk models may provide new opportunities and insights into this complex phenomenon (Kvam et al., 2021). Temporal oscillations in number of packets and percentage of subjects sending packets were observed for the Absent-Informed and Present-Uninformed groups which may indicate an impact on subject activity from both preference fluctuations and reductions in aleatory uncertainty over time. Quantifying the interplay between epistemic and ontic uncertainty in the human decision-making process may provide significant benefits in the development of adaptive cyber defense systems, which will rely on AI and/ or ML-based model to address challenges in both volume of data and required speed of action. The importance of quantifying uncertainty for predicting outcomes in ML applications was highlighted by Hüllermeier and Waegeman to address limitations in using a single probability distribution in modeling uncertainty (Hüllermeier and Waegeman, 2020). Techniques exist for ML methods to incorporate uncertainty into their modeling processes, but first, types of uncertainty need to be distinguished and formalized (Hüllermeier and Waegeman, 2020).

2. Decoy Effectiveness Improvement using Threat and Capability Coevolutionary Wargame Model

Subject interaction with decoy systems during the Tularosa study provides valuable information on how cyber-attackers interact with lightweight, low-fidelity decoys that were

used in the study. While responses from the decoys were automatically scripted depending on the inputs provided, subsequent activity by the human cyber-attacker varied greatly. This pattern of activity from initial input sent to the decoy by the cyber-attacker, the response from the decoy and reaction by the cyber-attacker could be analyzed algorithmically to identify decoy responses that elicit increased interaction with the decoys. Coevolutionary algorithms have been developed to model adversarial interaction across a wide range of environments from enterprise logistics to cyber-defense (Zhao et al., 2021; Rush et al., 2015).

In a pilot effort in support of funded research for the Cyber Agreements for Resilient Machines through Augmented AI (CARMA-AI) project, a small subset of network traffic generated by human subjects during the Tularosa study network traffic was examined using the "Threat and Capability Coevolutionary Wargame (TCCW)" model by Dr. Ying Zhao (Y. Zhao, personal communication, October 1, 2022). In this model a series of tournaments between the cyber-attacker and decoy were modeled based on the traffic statistics including number of packets sent between source and destination, protocols used, and the timing of sent packets (Y. Zhao, personal communication, October 1, 2022). The measure of effectiveness for these tournaments was based on the number of timestamps accumulated. (Y. Zhao, personal communication, October 1, 2022). Initial analysis conducted demonstrated the potential to improve the effectiveness of decoys by modifying the type and volume of packets sent by decoys in response to cyber-attacker interaction. Further work is needed to analyze the entire corpus of network traffic data captured during the Tularosa study using the TCCW model and conduct experimentation to validate the improved effectiveness of decoy responses recommended by the model, but application of coevolutionary algorithms provides a promising opportunity for the development of adaptive cyber defenses highlighted by Wang and Lu critically needed for proactive defense (Wang, 2018).

3. Further Factor Analysis

As this research does not investigate causal relationships between observed and latent variables identified from the Tularosa study, future work could construct models

derived from this research to determine how well data from the Tularosa study or other cyber deception experimentation would fit to the hypothesized models constructed using Confirmatory Factor Analysis (CFA) or Structural Equation Modeling (SEM) to generate additional insights. Additional survey data was collected from subjects participating in the Tularosa study which was not analyzed in the Factor Analysis conducted in research for this research. Investigation of this data using PCA, EFA, CFA or SEM could also yield new insights into the role of variables in accounting for variance observed.

4. Additional Tularosa Study Data Integration

While this research focused on temporal variables extracted from host and network data collected during the Tularosa study, temporal physiological data including heart-rate information, motion-based activity, peripheral skin temperature and galvanic skin response was also collected. Analysis of the data using a similar methodology as proposed in this dissertation may yield additional insights on the phenomenon of cyber deception through the exploration of relationships between captured network traffic, recording of subject activity, collected survey data and physiological data. Researchers investigating socio-technical phenomena occurring in and through cyberspace are frequently challenged by consolidating and integrating a cacophony of data collected from multiple disparate sources. The network graph-based methodology developed to aggregate and normalize timestamped and non-timestamped data during the Tularosa study could also be leveraged to quickly integrate disparate sources of data from previous or future experimentation.

APPENDIX A. PROGRAMMING CODE

This Appendix contains the programming code used to extract and transform raw data from the Tularosa study. Classes are not utilized in the Python coding within the Jupyter Notebooks because there is no need to maintain data state between function calls and all scripting is relatively short in length.

A. PCAP TRANSFORMATION

1. Split Oversize PCAP (250+ MB) into 200 MB files

```
tcpdump -r <input file> -w <output file) - C 200
```

The .pcap file extension is then appended to the split files

```
find . -type f  ! -name "*.*" -exec mv {}  {}.pcap \;
```

2. PCAP to CSV Conversion

```
for file in *.pcap; do tshark -r ${file%.*}.pcap -T fields -e frame.number -e
frame.time_epoch -e frame.len -e eth.src -e eth.dst -e ip.src -e ip.dst -e ip.proto -e
_ws.col.Protocol -e tcp.dstport -e tcp.stream -e udp.stream -e tcp.srcport -e udp.dstport
-e udp.srcport -e ip.flags -E header=y -E separator=',' -E quote=d -E occurrence=f >
${file%.*}.csv; done
```

B. PYTHON CODE USED TO IMPORT CSV PCAP DATA INTO NEO4J

Notes:

This Jupyter Notebook is configured for use on a Windows system. Slight modifications are needed for use on Linux systems, which are provided as comments in this section..

The Python programming used in this Jupyter Notebook is linear and straightforward to promote accessibility and use across a wide audience of users.

Neo4j Configuration Recommendations – Change the default settings for the following items: – dbms.memory.heap.initial_size=2G – dbms.memory.heap.max_size=4G – dbms.memory.pagecache.size=2G

For Windows systems, add the following items to the neo4j configuration settings

dbms.tx_log.rotation.retention_policy=keep_none

apoc.import.file.enabled=true

apoc.export.file.enabled=true

For Linux systems

Add to your neo4j.conf file

dbms.tx_log.rotation.retention_policy=keep_none

Add to your apoc.conf file

apoc.import.file.enabled=true

apoc.export.file.enabled=true

In each import function, the x = 'CALL apoc.periodic.iterate("CALL apoc.load.json('file:///' command, one / needs to be removed from file:///

This modification is included as a comment in each import function. Deleting the # in front of the line provided will uncomment the line so it will be used by the system.

Notes on Neo4j Configuration changes – Log.rotation.retention_policy setting needed to minimize size of transaction logs which can exceed tens of GB due to number of transactions that occur during import of data. – Apoc.import is needed to allow data to be imported into neo4j – Apoc.export is needed to allow files containing query results to be exported

General Notes – DO NOT have filenames with spaces in them. You will get this error – "ClientError: {code: Neo.ClientError.Procedure.ProcedureCallFailed} {message: Failed to invoke procedure *apoc.load.csv*: Caused by: java.nio.file.InvalidPathException: Illegal char."

```
#This block imports the python "neo4j," "time" and "os" modules used to import the data into N
eo4j.
#These modules need to be installed within the Python environment before use.

from neo4j import GraphDatabase
import time
import os

#### Connect to Neo4j Database

#Note this is only intended to be used on your local system because the username and password a
re being passed in cleartext
#Additional information on secure authentication can be found at https://neo4j.com/developer/pyt
hon/

uri = input('Please type in the URL of your Neo4j instance and port. Hit enter for Default – bolt://
localhost:7687') or 'bolt://localhost:7687'
username = input('Please type in the name of your Neo4j instance. Hit enter for default – neo4j')
or 'neo4j'
password = input('Please type in the password for your Neo4j instance. Hit enter for default – pas
sword') or 'password'

driver = GraphDatabase.driver(uri, auth=(username, password))
```

```
#This block specifies the filename and environment based on the subject whose data is being imp
orted

subject = input('Please type the SubjectID you want to enter – ')
print(subject)

if subject == 'S01':
   filename_day1 = 'S01_Day1_AU'
   filename_day2 = 'S01_Day2_PU'
   environment_day1 = '1'
   environment_day2 = '3'

elif subject == 'S02':
   filename_day1 = 'S02_Day1_AU'
   filename_day2 = 'S02_Day2_PU'
   environment_day1 = '1'
   environment_day2 = '3'

elif subject == 'S03':
   filename_day1 = 'S03_Day1_AI'
   filename_day2 = 'S03_Day2_AU'
   environment_day1 = '2'
   environment_day2 = '1'

elif subject == 'S04':
```

```
    filename_day1 = 'S04_Day1_AI'
    filename_day2 = 'S04_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '1'

elif subject == 'S05':
    filename_day1 = 'S05_Day1_AI'
    filename_day2 = 'S05_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '1'
elif subject == 'S06':
    filename_day1 = 'S06_Day1_AU'
    filename_day2 = 'S06_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S07':
    filename_day1 = 'S07_Day1_AU'
    filename_day2 = 'S07_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S08':
    filename_day1 = 'S08_Day1_AU'
    filename_day2 = 'S08_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S09':
    filename_day1 = 'S09_Day1_AU'
    filename_day2 = 'S09_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S10':
    filename_day1 = 'S10_Day1_AU'
    filename_day2 = 'S10_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S11':
    filename_day1 = 'S11_Day1_AU'
    filename_day2 = 'S11_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S12':
    filename_day1 = 'S12_Day1_AU'
    filename_day2 = 'S12_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S13':
    filename_day1 = 'S13_Day1_AU'
    filename_day2 = 'S13_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
```

```
elif subject == 'S14':
    filename_day1 = 'S14_Day1_PU'
    filename_day2 = 'S14_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S15':
    filename_day1 = 'S15_Day1_PU'
    filename_day2 = 'S15_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S16':
    filename_day1 = 'S16_Day1_PU'
    filename_day2 = 'S16_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'

elif subject == 'S17':
    filename_day1 = 'S17_Day1_PU'
    filename_day2 = 'S17_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S18':
    filename_day1 = 'S18_Day1_PU'
    filename_day2 = 'S18_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S19':
    filename_day1 = 'S19_Day1_PU'
    filename_day2 = 'S19_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S20':
    filename_day1 = 'S20_Day1_PI'
    filename_day2 = 'S20_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S21':
    filename_day1 = 'S21_Day1_PI'
    filename_day2 = 'S21_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S22':
    filename_day1 = 'S22_Day1_PI'
    filename_day2 = 'S22_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S23':
    filename_day1 = 'S23_Day1_PI'
    filename_day2 = 'S23_Day2_AU'
    environment_day1 = '3'
```

```
    environment_day2 = '2'
elif subject == 'S24':
    filename_day1 = 'S24_Day1_PI'
    filename_day2 = 'S24_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S25':
    filename_day1 = 'S25_Day1_PI'
    filename_day2 = 'S25_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S26':
    filename_day1 = 'S26_Day1_PI'
    filename_day2 = 'S26_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S27':
    filename_day1 = 'S27_Day1_AI'
    filename_day2 = 'S27_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S28':
    filename_day1 = 'S28_Day1_AI'
    filename_day2 = 'S28_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S29':
    filename_day1 = 'S29_Day1_AI'
    filename_day2 = 'S29_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S30':
    filename_day1 = 'S30_Day1_AI'
    filename_day2 = 'S30_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S31':
    filename_day1 = 'S31_Day1_AI'
    filename_day2 = 'S31_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S32':
    filename_day1 = 'S32_Day1_AI'
    filename_day2 = 'S32_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S33':
    filename_day1 = 'S33_Day1_AU'
    filename_day2 = 'S33_Day2_PU'
    environment_day1 = '1'
```

```
    environment_day2 = '3'
elif subject == 'S34':
    filename_day1 = 'S34_Day1_AU'
    filename_day2 = 'S34_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S35':
    filename_day1 = 'S35_Day1_AU'
    filename_day2 = 'S35_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S36':
    filename_day1 = 'S36_Day1_AU'
    filename_day2 = 'S36_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S37':
    filename_day1 = 'S37_Day1_AU'
    filename_day2 = 'S37_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S38':
    filename_day1 = 'S38_Day1_AU'
    filename_day2 = 'S38_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S39':
    filename_day1 = 'S39_Day1_AU'
    filename_day2 = 'S39_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S40':
    filename_day1 = 'S40_Day1_AU'
    filename_day2 = 'S40_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S41':
    filename_day1 = 'S41_Day1_AU'
    filename_day2 = 'S41_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S42':
    filename_day1 = 'S42_Day1_AU'
    filename_day2 = 'S42_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S43':
    filename_day1 = 'S43_Day1_PU'
    filename_day2 = 'S43_Day2_AU'
    environment_day1 = '3'
```

```
    environment_day2 = '2'
elif subject == 'S44':
    filename_day1 = 'S44_Day1_PU'
    filename_day2 = 'S44_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S45':
    filename_day1 = 'S45_Day1_PU'
    filename_day2 = 'S45_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S46':
    filename_day1 = 'S46_Day1_PU'
    filename_day2 = 'S46_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S47':
    filename_day1 = 'S47_Day1_PU'
    filename_day2 = 'S47_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S48':
    filename_day1 = 'S48_Day1_PU'
    filename_day2 = 'S48_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S49':
    filename_day1 = 'S49_Day1_PU'
    filename_day2 = 'S49_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S50':
    filename_day1 = 'S50_Day1_PU'
    filename_day2 = 'S50_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S51':
    filename_day1 = 'S51_Day1_PU'
    filename_day2 = 'S51_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S52':
    filename_day1 = 'S52_Day1_PU'
    filename_day2 = 'S52_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S53':
    filename_day1 = 'S53_Day1_PI'
    filename_day2 = 'S53_Day2_AU'
    environment_day1 = '3'
```

```
    environment_day2 = '2'
elif subject == 'S54':
    filename_day1 = 'S54_Day1_PI'
    filename_day2 = 'S54_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S55':
    filename_day1 = 'S55_Day1_PI'
    filename_day2 = 'S55_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S56':
    filename_day1 = 'S56_Day1_PI'
    filename_day2 = 'S56_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S57':
    filename_day1 = 'S57_Day1_PI'
    filename_day2 = 'S57_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S58':
    filename_day1 = 'S58_Day1_PI'
    filename_day2 = 'S58_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S59':
    filename_day1 = 'S59_Day1_PI'
    filename_day2 = 'S59_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S60':
    filename_day1 = 'S60_Day1_PI'
    filename_day2 = 'S60_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S61':
    filename_day1 = 'S61_Day1_AI'
    filename_day2 = 'S61_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S62':
    filename_day1 = 'S62_Day1_AI'
    filename_day2 = 'S62_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S63':
    filename_day1 = 'S63_Day1_AI'
    filename_day2 = 'S63_Day2_AU'
    environment_day1 = '1'
```

```
    environment_day2 = '2'
elif subject == 'S64':
    filename_day1 = 'S64_Day1_AI'
    filename_day2 = 'S64_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S65':
    filename_day1 = 'S65_Day1_AI'
    filename_day2 = 'S65_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S66':
    filename_day1 = 'S66_Day1_AI'
    filename_day2 = 'S66_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S67':
    filename_day1 = 'S67_Day1_AI'
    filename_day2 = 'S67_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S68':
    filename_day1 = 'S68_Day1_AI'
    filename_day2 = 'S68_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S69':
    filename_day1 = 'S69_Day1_AI'
    filename_day2 = 'S69_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S70':
    filename_day1 = 'S70_Day1_AI'
    filename_day2 = 'S70_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S71':
    filename_day1 = 'S71_Day1_AU'
    filename_day2 = 'S71_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S72':
    filename_day1 = 'S72_Day1_AU'
    filename_day2 = 'S72_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S73':
    filename_day1 = 'S73_Day1_AU'
    filename_day2 = 'S73_Day2_PU'
    environment_day1 = '1'
```

```
    environment_day2 = '3'
elif subject == 'S74':
    filename_day1 = 'S74_Day1_AU'
    filename_day2 = 'S74_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S75':
    filename_day1 = 'S75_Day1_AU'
    filename_day2 = 'S75_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S76':
    filename_day1 = 'S76_Day1_AU'
    filename_day2 = 'S76_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S77':
    filename_day1 = 'S77_Day1_AU'
    filename_day2 = 'S77_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S78':
    filename_day1 = 'S78_Day1_AU'
    filename_day2 = 'S78_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S79':
    filename_day1 = 'S79_Day1_AU'
    filename_day2 = 'S79_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S80':
    filename_day1 = 'S80_Day1_AU'
    filename_day2 = 'S80_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S81':
    filename_day1 = 'S81_Day1_PU'
    filename_day2 = 'S81_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S82':
    filename_day1 = 'S82_Day1_PU'
    filename_day2 = 'S82_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S83':
    filename_day1 = 'S83_Day1_PU'
    filename_day2 = 'S83_Day2_AU'
    environment_day1 = '3'
```

```
    environment_day2 = '2'
elif subject == 'S84':
    filename_day1 = 'S84_Day1_PU'
    filename_day2 = 'S84_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S85':
    filename_day1 = 'S85_Day1_PU'
    filename_day2 = 'S85_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S86':
    filename_day1 = 'S86_Day1_PU'
    filename_day2 = 'S86_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S87':
    filename_day1 = 'S87_Day1_PU'
    filename_day2 = 'S87_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S88':
    filename_day1 = 'S88_Day1_PU'
    filename_day2 = 'S88_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S89':
    filename_day1 = 'S89_Day1_PU'
    filename_day2 = 'S89_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S90':
    filename_day1 = 'S90_Day1_AI'
    filename_day2 = 'S90_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S91':
    filename_day1 = 'S91_Day1_PI'
    filename_day2 = 'S91_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S92':
    filename_day1 = 'S92_Day1_PI'
    filename_day2 = 'S92_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S93':
    filename_day1 = 'S93_Day1_PI'
    filename_day2 = 'S93_Day2_AU'
    environment_day1 = '3'
```

```
    environment_day2 = '2'
elif subject == 'S94':
    filename_day1 = 'S94_Day1_PI'
    filename_day2 = 'S94_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S95':
    filename_day1 = 'S95_Day1_PI'
    filename_day2 = 'S95_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S96':
    filename_day1 = 'S96_Day1_PI'
    filename_day2 = 'S96_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S97':
    filename_day1 = 'S97_Day1_PI'
    filename_day2 = 'S97_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S98':
    filename_day1 = 'S98_Day1_PI'
    filename_day2 = 'S98_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S99':
    filename_day1 = 'S99_Day1_PI'
    filename_day2 = 'S99_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S100':
    filename_day1 = 'S100_Day1_AU'
    filename_day2 = 'S100_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S101':
    filename_day1 = 'S101_Day1_AU'
    filename_day2 = 'S101_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S102':
    filename_day1 = 'S102_Day1_AU'
    filename_day2 = 'S102_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S103':
    filename_day1 = 'S103_Day1_AU'
    filename_day2 = 'S103_Day2_PU'
    environment_day1 = '1'
```

```
    environment_day2 = '3'
elif subject == 'S104':
    filename_day1 = 'S104_Day1_AU'
    filename_day2 = 'S104_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S105':
    filename_day1 = 'S105_Day1_AI'
    filename_day2 = 'S105_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S106':
    filename_day1 = 'S106_Day1_AI'
    filename_day2 = 'S106_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S107':
    filename_day1 = 'S107_Day1_AI'
    filename_day2 = 'S107_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S108':
    filename_day1 = 'S108_Day1_AI'
    filename_day2 = 'S108_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S109':
    filename_day1 = 'S109_Day1_AI'
    filename_day2 = 'S109_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S110':
    filename_day1 = 'S110_Day1_AU'
    filename_day2 = 'S110_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S111':
    filename_day1 = 'S111_Day1_AU'
    filename_day2 = 'S111_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S112':
    filename_day1 = 'S112_Day1_AU'
    filename_day2 = 'S112_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S113':
    filename_day1 = 'S113_Day1_PU'
    filename_day2 = 'S113_Day2_AU'
    environment_day1 = '3'
```

```
    environment_day2 = '2'
elif subject == 'S114':
    filename_day1 = 'S114_Day1_PU'
    filename_day2 = 'S114_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S115':
    filename_day1 = 'S115_Day1_PI'
    filename_day2 = 'S115_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S116':
    filename_day1 = 'S116_Day1_PI'
    filename_day2 = 'S116_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S117':
    filename_day1 = 'S117_Day1_PI'
    filename_day2 = 'S117_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S118':
    filename_day1 = 'S118_Day1_AI'
    filename_day2 = 'S118_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S119':
    filename_day1 = 'S119_Day1_AI'
    filename_day2 = 'S119_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S120':
    filename_day1 = 'S120_Day1_AU'
    filename_day2 = 'S120_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S121':
    filename_day1 = 'S121_Day1_AU'
    filename_day2 = 'S121_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S122':
    filename_day1 = 'S122_Day1_AU'
    filename_day2 = 'S122_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S123':
    filename_day1 = 'S123_Day1_PU'
    filename_day2 = 'S123_Day2_AU'
    environment_day1 = '3'
```

```
    environment_day2 = '2'
elif subject == 'S124':
    filename_day1 = 'S124_Day1_PU'
    filename_day2 = 'S124_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S125':
    filename_day1 = 'S125_Day1_PI'
    filename_day2 = 'S125_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S126':
    filename_day1 = 'S126_Day1_PI'
    filename_day2 = 'S126_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S127':
    filename_day1 = 'S127_Day1_PI'
    filename_day2 = 'S127_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S128':
    filename_day1 = 'S128_Day1_AI'
    filename_day2 = 'S128_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S129':
    filename_day1 = 'S129_Day1_AI'
    filename_day2 = 'S129_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S130':
    filename_day1 = 'S130_Day1_AI'
    filename_day2 = 'S130_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S131':
    filename_day1 = 'S131_Day1_AI'
    filename_day2 = 'S131_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
elif subject == 'S132':
    filename_day1 = 'S132_Day1_PU'
    filename_day2 = 'S132_Day2_AU'
    environment_day1 = '3'
    environment_day2 = '2'
elif subject == 'S133':
    filename_day1 = 'S133_Day1_AU'
    filename_day2 = 'S133_Day2_PU'
    environment_day1 = '1'
```

```
    environment_day2 = '3'
elif subject == 'S134':
    filename_day1 = 'S134_Day1_AU'
    filename_day2 = 'S134_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S135':
    filename_day1 = 'S135_Day1_AU'
    filename_day2 = 'S135_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S136':
    filename_day1 = 'S136_Day1_AU'
    filename_day2 = 'S136_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S137':
    filename_day1 = 'S137_Day1_AU'
    filename_day2 = 'S137_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S138':
    filename_day1 = 'S138_Day1_AU'
    filename_day2 = 'S138_Day2_PU'
    environment_day1 = '1'
    environment_day2 = '3'
elif subject == 'S139':
    filename_day1 = 'S139_Day1_AI'
    filename_day2 = 'S139_Day2_AU'
    environment_day1 = '1'
    environment_day2 = '2'
```

```
#Note: This creates index for various properties to greatly speed queries


def create_indexes_neo4j():

        with driver.session() as session:
        session.run('CREATE INDEX second_node_key IF NOT EXISTS FOR  (n:Second) ON (n.
Node_Key)')
        session.run('CREATE INDEX second_epoch IF NOT EXISTS FOR (n:Second) ON (n.Epo
chSeconds)')
        session.run('CREATE INDEX NetFlow_node_key IF NOT EXISTS FOR (n:NetFlow) ON
(n.Node_Key)')
        session.run('CREATE INDEX NetFlow_seconds IF NOT EXISTS FOR (n:NetFlow) ON (n
.EpochSeconds)')
        session.run('CREATE INDEX NetFlow_SourceIP IF NOT EXISTS FOR (n:NetFlow) ON (
n.SourceIP)')
```

```
    session.run('CREATE INDEX NetFlow_DestIP IF NOT EXISTS FOR (n:NetFlow) ON (n.
DestIP)')
    session.run('CREATE INDEX NetFlow_Env IF NOT EXISTS FOR (n:NetFlow) ON (n.En
vironment)')
    session.run('CREATE INDEX NetFlow_SubjectID IF NOT EXISTS FOR (n:NetFlow) ON
(n.SubjectID)')
    session.run('CREATE INDEX SystemUUID IF NOT EXISTS FOR (n:system) ON (n.UUI
D)')
    session.run('CREATE INDEX SystemIP IF NOT EXISTS FOR (n:system) ON (n.IP_Addre
ss)')

create_indexes_neo4j()
```

```
# This syntax imports information on subject Day 1 start, end, lunch start and lunch end time
# Included is a 1 hour time offset corrections for Subjects S01-S19 whose experiment trials occur
ed during
# Mountain Daylight Time (MDT) and not Mountain Standard Time (MST) since Coordinated Un
iversal Time (UTC)
# does not observe Daylight Savings Time

def import_day1_into_neo4j(filename):


  with driver.session() as session:

    neodir = session.run('CALL dbms.listConfig() YIELD name, value WHERE name = \'dbms
.directories.neo4j_home\' RETURN value').values()
    path = neodir [0][0] + '/import/'


    path = path.replace("\\","/")
    x = 'CALL apoc.load.csv(\'file:///' + path + filename + '\',{header:true}) YIELD map AS ro
w WITH apoc.create.uuid() AS uuid,row WHERE row.sub_id1 = \'' + subject + '\' MERGE (c:d
ay1 {Node_Key: uuid, Environment:row.env_day1, Day: \'1\', SubjectID: row.sub_id1, Conditio
n: row.cond_day1, StartTime: apoc.date.parse(row.rundate1+row.day1_start, \'s\', \'yyyyMMdd
HH:mm:ss\')+25200, EndTime: apoc.date.parse(row.rundate1+row.day1_end, \'s\', \'yyyyMMdd
HH:mm:ss\')+25200, LunchStartTime: apoc.date.parse(row.rundate1+row.day1_lunch_start, \'s\'
, \'yyyyMMddHH:mm:ss\')+25200, LunchEndTime: apoc.date.parse(row.rundate1+row.day1_lu
nch_end, \'s\', \'yyyyMMddHH:mm:ss\')+25200, Date: date(row.rundate1)})'

    #For Linux systems, use the Cypher syntax below
    #x = 'CALL apoc.load.csv(\'' + path + filename + '\',{header:true}) YIELD map AS row WI
TH apoc.create.uuid() AS uuid,row WHERE row.sub_id1 = \'' + subject + '\' MERGE (c:day1 {
Node_Key: uuid, Environment:row.env_day1, Day: \'1\', SubjectID: row.sub_id1, Condition: ro
w.cond_day1, StartTime: apoc.date.parse(row.rundate1+row.day1_start, \'s\', \'yyyyMMddHH:
mm:ss\')+25200, EndTime: apoc.date.parse(row.rundate1+row.day1_end, \'s\', \'yyyyMMddHH
:mm:ss\')+25200, LunchStartTime: apoc.date.parse(row.rundate1+row.day1_lunch_start, \'s\', \
'yyyyMMddHH:mm:ss\')+25200, LunchEndTime: apoc.date.parse(row.rundate1+row.day1_lunc
```

```
h_end, \'s\', \'yyyyMMddHH:mm:ss\')+25200, Date: date(row.rundate1)})'

        #This prints out the Cypher syntax generated to help troubleshooting if needed. This output c
an be pasted directly into the Neo4j Browser
        #print(x)

        result = session.run(x).data()
        session.run('WITH ["S01","S02","S03","S04","S05","S06","S07","S08","S09","S10","S11
","S12",S13","S14","S15","S16","S17","S18","S19"] AS subjects UNWIND subjects AS x
MATCH (n:day1) WHERE n.SubjectID = x SET n.EndTime = (n.EndTime – 3600),
n.LunchEndTime = (n.LunchEndTime – 3600), n.LunchStartTime =
(n.LunchStartTime – 3600), n.StartTime = (n.StartTime – 3600)')

print("Any Errors generated." + str(result))
 print("Empty Brackets [] are returned if there are no errors")

import_day1_into_neo4j('Tularosa_subjects.csv')


---

# This syntax imports information on subject Day 1 start, end, lunch start and lunch end time
# Included is a 1 hour time offset corrections for Subjects S01-S19 whose experiment trials occur
ed during
# Mountain Daylight Time (MDT) and not Mountain Standard Time (MST) since Coordinated Un
iversal Time (UTC)
# does not observe Daylight Savings Time

def import_day2_into_neo4j(filename):

    with driver.session() as session:

        neodir = session.run('CALL dbms.listConfig() YIELD name, value WHERE name = \'dbms
.directories.neo4j_home\' RETURN value').values()
        path = neodir [0][0] + '/import/'
        path = path.replace("\\","/")

        x = 'CALL apoc.load.csv(\'file:///' + path + filename + '\',{header:true}) YIELD map AS ro
w WITH apoc.create.uuid() AS uuid,row WHERE row.sub_id1 = \'' + subject + '\' MERGE (c:d
ay2 {Node_Key: uuid, Environment:row.env_day2, Day: \'2\', SubjectID: row.sub_id2, Conditio
n: row.cond_day2, StartTime: apoc.date.parse(row.rundate2+row.day2_start, \'s\', \'yyyyMMdd
HH:mm:ss\')+25200, EndTime: apoc.date.parse(row.rundate2+row.day2_end, \'s\', \'yyyyMMdd
HH:mm:ss\')+25200, LunchStartTime: apoc.date.parse(row.rundate2+row.day2_lunch_start, \'s\'
, \'yyyyMMddHH:mm:ss\')+25200, LunchEndTime: apoc.date.parse(row.rundate2+row.day2_lu
nch_end, \'s\', \'yyyyMMddHH:mm:ss\')+25200, Date: date(row.rundate2)})'

        #For Linux systems, use the Cypher syntax below
        #x = 'CALL apoc.load.csv(\'' + path + filename + '\',{header:true}) YIELD map AS row WI
TH apoc.create.uuid() AS uuid,row WHERE row.sub_id1 = \'' + subject + '\' MERGE (c:day2 {
Node_Key: uuid, Environment:row.env_day2, Day: \'2\', SubjectID: row.sub_id2, Condition: ro
w.cond_day2, StartTime: apoc.date.parse(row.rundate2+row.day2_start, \'s\', \'yyyyMMddHH:
```

```
mm:ss\')+25200, EndTime: apoc.date.parse(row.rundate2+row.day2_end, \'s\', \'yyyyMMddHH
:mm:ss\')+25200, LunchStartTime: apoc.date.parse(row.rundate2+row.day2_lunch_start, \'s\', \
'yyyyMMddHH:mm:ss\')+25200, LunchEndTime: apoc.date.parse(row.rundate2+row.day2_lunc
h_end, \'s\', \'yyyyMMddHH:mm:ss\')+25200, Date: date(row.rundate2)})'

    #This prints out the Cypher syntax generated to help troubleshooting if needed. This output c
an be pasted directly into the Neo4j Browser
    #print(x)

    result = session.run(x).data()
    session.run('WITH ["S01","S02","S03","S04","S05","S06","S07","S08","S09","S10","
S11","S12","S13","S14","S15","S16","S17","S18","S19"] AS subjects UNWIND subjects AS x
MATCH (n:day2) WHERE n.SubjectID = x SET n.EndTime = (n.EndTime – 3600), n.LunchEnd
Time = (n.LunchEndTime – 3600), n.LunchStartTime = (n.LunchStartTime – 3600), n.StartTime
= (n.StartTime – 3600)')


    print("Any Errors generated." + str(result))
    print("Empty Brackets [] are returned if there are no errors")

import_day2_into_neo4j('Tularosa_subjects.csv')
```

```
# This imports the Tularosa Host Configuration spreadsheet data for systems on the Experiment
Cyber Range for each Environment used by the Subject

def load_host_config_into_neo4j(filename):
  with driver.session() as session:

    neodir = session.run('CALL dbms.listConfig() YIELD name, value WHERE name = \'dbms
.directories.neo4j_home\' RETURN value').values()
    path = neodir [0][0] + '/import/'
    path = path.replace("\\","/")

    x = 'CALL apoc.periodic.iterate("CALL apoc.load.csv(\'file:///' + path + filename + '\',{hea
der:true})'
    y = """
    y =  x + ' YIELD map AS row","MATCH (d) WHERE row.`Environment` = d.Environment
MERGE (r:system {Node_Key: coalesce(row.`Computer_Name`, \'\'), SubjectID: d.SubjectID,
UUID: d.Node_Key, Day: d.Day, Environment: coalesce(row.`Environment`, \'\'), Validity: coal
esce(row.`Validity`, \'\'), IP_Address: coalesce(row.`IP_Address`, \'\'), OS_Type: coalesce(row.`
os_type`, \'\'), OS: coalesce(row.`Operating_System`, \'\'), Emulation: coalesce(row.`Emulating`
, \'\'), Version: coalesce(row.`Version`, \'\'), Services: coalesce(row.`Services`, \'\'), Open_Ports:
coalesce(row.`Open_Ports (AS DETERMINED BY SANDIA)`, \'\'), Notes: coalesce(row.`Notes
(ground truth or discrepancies)`, \'\'), Host_Script_Results: coalesce(row.`HOST SCRIPT RESU
LTS (CLOCKSKEW / SSH KEYS)`, \'\')}),"  {batchSize:1000, iterateList:true, parallel:false})'

    #This prints out the Cypher syntax generated to help troubleshooting if needed. This output c
an be pasted directly into the Neo4j Browser
    #print(y)
```

```
        #This executes the Cypher syntax and returns any errors
        result = session.run(y).data()
        print("Failed Batches – " + str(result [0]['failedBatches']))
        print("Error Messages – " + str(result [0]['batch']['errors']))
        print("Nodes Created – " + str(result [0]['updateStatistics']['nodesCreated']))
        print("Properties Set – " + str(result [0]['updateStatistics']['propertiesSet']))

        session.close()

load_host_config_into_neo4j("Host_Configuration.csv")
```

```
# The first line connects systems within each Cyber Range Environment to the Router used on the network
# The second and third lines connect the Router of the Cyber Range Environment to its respective Day1 or Day2 node.
# This connection permits queries of the correct Cyber Range Environment system based on the day of the experiment trial

def create_relationships_between_systems_and_days():

    with driver.session() as session:
        session.run("MATCH (n:system),(m:system) WHERE n.UUID = m.UUID AND n.IP_Address = '192.168.5.1' AND n.IP_Address <> m.IP_Address MERGE (m)-[:CONNECTED_TO]->(n)")
        session.run("MATCH (n:day1),(s:system) WHERE n.Node_Key = s.UUID AND n.Environment = s.Environment AND s.Day = '1' AND s.IP_Address = '192.168.5.1' MERGE (n)<-[:DAY1_NETWORK]-(s)")
        session.run("MATCH (n:day2),(s:system) WHERE n.Node_Key = s.UUID AND n.Environment = s.Environment AND s.Day = '2' AND s.IP_Address = '192.168.5.1' MERGE (n)<-[:DAY2_NETWORK]-(s)")

create_relationships_between_systems_and_days()
```

```
# Create Year, Month, Day, Hour time trees for each day for each subject experiment trial

def year_month_day_hour_time_tree():

    with driver.session() as session:
        #This executes the Cypher syntax and returns any errors

        x = 'CALL apoc.periodic.iterate(\'MATCH (d:day1) WITH datetime(apoc.date.toISO8601(d.StartTime, "s")) AS start,d WITH datetime(apoc.date.toISO8601(d.EndTime, "s")) AS end,start,d WITH end,start,d, range(start.year, end.year) AS years, range(start.month,end.month) as months, range(start.hour-6,end.hour+6) AS hours, range(0,59) AS minutes, d.Node_Key AS UUID RETURN start,end,d,years,months,hours,minutes, UUID\',\' FOREACH(year IN years | MERGE (y:Year {year: year, Node_Key: d.Node_Key}) FOREACH(month IN months | CREATE (m:Month {month: month,Node_Key: UUID}) MERGE (y)-[:HAS_MONTH]->(m) FOREACH(day IN ran
```

```
ge(start.day, end.day) | CREATE (d:Day {day: day,Node_Key: UUID}) MERGE (m)-[:HAS_DA
Y]->(d) FOREACH(hour IN hours | CREATE (h:Hour {hour: hour,Node_Key: UUID}) MERGE
(d)-[:HAS_HOUR]->(h))))) RETURN COUNT(*)\', {batchSize:1000, parallel:true, iterateList:tr
ue})'

    result = session.run(x).data()
    print("Failed Batches – " + str(result [0]['failedBatches']))
    print("Error Messages – " + str(result [0]['batch']['errors']))
    print("Nodes Created – " + str(result [0]['updateStatistics']['nodesCreated']))
    print("Properties Set – " + str(result [0]['updateStatistics']['propertiesSet']))


    session.run('CALL apoc.periodic.iterate(\'MATCH (d:day2) WITH datetime(apoc.date.toIS
O8601(d.StartTime, "s")) AS start,d WITH datetime(apoc.date.toISO8601(d.EndTime, "s")) AS
end,start,d WITH end,start,d, range(start.year, end.year) AS years, range(start.month,end.month)
as months, range(start.hour-6,end.hour+6) AS hours, range(0,59) AS minutes, d.Node_Key AS U
UID RETURN start,end,d,years,months,hours,minutes, UUID\',\' FOREACH(year IN years | ME
RGE (y:Year {year: year, Node_Key: d.Node_Key}) FOREACH(month IN months | CREATE (
m:Month {month: month,Node_Key: UUID}) MERGE (y)-[:HAS_MONTH]->(m) FOREACH(d
ay IN range(start.day, end.day) | CREATE (d:Day {day: day,Node_Key: UUID}) MERGE (m)-[:
HAS_DAY]->(d) FOREACH(hour IN hours | CREATE (h:Hour {hour: hour,Node_Key: UUID}
) MERGE (d)-[:HAS_HOUR]->(h))))) RETURN COUNT(*)\', {batchSize:1000, parallel:true, ite
rateList:true})')

year_month_day_hour_time_tree()
```

```
# Start and end times for the subjects included below cross days when coverted to UTC (MST +7)
, 1700 MST = 0000 UTC which requires creation of additional hour nodes to account for this day
crossover

def year_month_day_hour_time_tree_utc_crossover():


  with driver.session() as session:
    session.run('CALL apoc.periodic.iterate(\'UNWIND ["S33","S34","S35","S36","S37","S3
8","S39","S40","S41","S42"] AS x MATCH (d:day1) WHERE d.SubjectID = x WITH d.Node_
Key AS UUID MATCH (z:Day) WHERE z.Node_Key = UUID AND z.day = 13 WITH UUID,z,
range(16,23) AS hours, range(0,59) AS minutes RETURN hours, minutes, UUID,z\',\' FOREAC
H(hour IN hours | CREATE (h:Hour {hour: hour,Node_Key: UUID}) MERGE (z)-[:HAS_HOU
R]->(h))\', {batchSize:1000, parallel:true, iterateList:true})')
    session.run('CALL apoc.periodic.iterate(\'UNWIND ["S33","S34","S35","S36","S37","S3
8","S39","S40","S41","S42"] AS x MATCH (d:day1) WHERE d.SubjectID = x WITH d.Node_
Key AS UUID MATCH (z:Day) WHERE z.Node_Key = UUID AND z.day = 14 WITH UUID,z,
range(0,0) AS hours, range(0,59) AS minutes RETURN hours, minutes, UUID,z\',\' FOREACH(
hour IN hours | CREATE (h:Hour {hour: hour,Node_Key: UUID}) MERGE (z)-[:HAS_HOUR]-
>(h))\', {batchSize:1000, parallel:true, iterateList:true})')

year_month_day_hour_time_tree_utc_crossover()
```

```
# This syntax creates 60 Seconds nodes for each minute node
# Note: This script take several minutes due to the large number of nodes create

def second_time_tree():

    with driver.session() as session:
        session.run('WITH range(0,23) AS hours, range(0,59) AS minutes UNWIND hours AS hour
WITH hour, minutes MATCH (h:Hour {hour: hour}) FOREACH(minute IN minutes | CREATE
(m:Minute {minute: minute}) MERGE (h)-[:HAS_MINUTE]->(m))')
        session.run('Call apoc.periodic.iterate(\'WITH range(0,9) AS minutes, range (0,59) AS seco
nds UNWIND minutes AS minute WITH minute, seconds MATCH (m:Minute {minute: minute}
) RETURN m,seconds\',\'FOREACH(second IN seconds | CREATE (s:Second {second: second}
) MERGE (m)-[:HAS_SECOND]->(s))\', {batchSize:1000, parallel:true, iterateList:true})')
        session.run('Call apoc.periodic.iterate(\'WITH range(10,19) AS minutes, range (0,59) AS se
conds UNWIND minutes AS minute WITH minute, seconds MATCH (m:Minute {minute: minut
e}) RETURN m,seconds\',\'FOREACH(second IN seconds | CREATE (s:Second {second: secon
d}) MERGE (m)-[:HAS_SECOND]->(s))\', {batchSize:1000, parallel:true, iterateList:true})')
        session.run('Call apoc.periodic.iterate(\'WITH range(20,29) AS minutes, range (0,59) AS se
conds UNWIND minutes AS minute WITH minute, seconds MATCH (m:Minute {minute: minut
e}) RETURN m,seconds\',\'FOREACH(second IN seconds | CREATE (s:Second {second: secon
d}) MERGE (m)-[:HAS_SECOND]->(s))\', {batchSize:1000, parallel:true, iterateList:true})')
        session.run('Call apoc.periodic.iterate(\'WITH range(30,39) AS minutes, range (0,59) AS se
conds UNWIND minutes AS minute WITH minute, seconds MATCH (m:Minute {minute: minut
e}) RETURN m,seconds\',\'FOREACH(second IN seconds | CREATE (s:Second {second: secon
d}) MERGE (m)-[:HAS_SECOND]->(s))\', {batchSize:1000, parallel:true, iterateList:true})')
        session.run('Call apoc.periodic.iterate(\'WITH range(40,49) AS minutes, range (0,59) AS se
conds UNWIND minutes AS minute WITH minute, seconds MATCH (m:Minute {minute: minut
e}) RETURN m,seconds\',\'FOREACH(second IN seconds | CREATE (s:Second {second: secon
d}) MERGE (m)-[:HAS_SECOND]->(s))\', {batchSize:1000, parallel:true, iterateList:true})')
        session.run('Call apoc.periodic.iterate(\'WITH range(50,59) AS minutes, range (0,59) AS se
conds UNWIND minutes AS minute WITH minute, seconds MATCH (m:Minute {minute: minut
e}) RETURN m,seconds\',\'FOREACH(second IN seconds | CREATE (s:Second {second: secon
d}) MERGE (m)-[:HAS_SECOND]->(s))\', {batchSize:1000, parallel:true, iterateList:true})')

second_time_tree()
```

```
# This script sets the common UUID used for all nodes associated with each day of the experimen
t trial then sets Unix Epoch Second
# associated with each second

def set_unix_epoch_seconds():

    with driver.session() as session:

        # Set UUID to all second and minute nodes
        session.run('CALL apoc.periodic.iterate(\'MATCH (h:Hour)-[:HAS_MINUTE]->(m:Minut
e) RETURN h,m\',\' SET m.Node_Key = h.Node_Key\', {batchSize:1000, parallel:true, iterateLi
```

```
st:true})')
        session.run('CALL apoc.periodic.iterate(\'MATCH (m:Minute)-[:HAS_SECOND]->(s:Second) RETURN m,s\',\' SET s.Node_Key = m.Node_Key\', {batchSize:1000, parallel:true, iterateList:true})')

        # Create Unix Epoch Second value for every second node

        session.run('CALL apoc.periodic.iterate(\'MATCH (y:Year)-[:HAS_MONTH]->(m:Month)-[:HAS_DAY]->(d:Day)-[:HAS_HOUR]->(h:Hour)-[:HAS_MINUTE]->(x:Minute)-[:HAS_SECOND]->(s:Second) WITH toString(y.year+"-"+m.month+"-"+d.day+"T"+h.hour+":"+x.minute+":"+s.second) AS date,s RETURN date,s\',\'SET s.EpochSeconds = datetime(date).epochSeconds\', {batchSize:1000, parallel:true, iterateList:true})')


set_unix_epoch_seconds()


---


# Creates EVENTS_STARTS_AT, EVENT_ENDS_AT, LUNCH_STARTS_AT and LUNCH_ENDS_AT relationships between subjects and second event started

def set_start_end_time():

    with driver.session() as session:

        session.run('MATCH (d:day1),(s:Second) WHERE d.Node_Key = s.Node_Key AND d.StartTime = s.EpochSeconds MERGE (d)-[:EVENT_STARTS_AT]->(s)')
        session.run('MATCH (d:day2),(s:Second) WHERE d.Node_Key = s.Node_Key AND d.StartTime = s.EpochSeconds MERGE (d)-[:EVENT_STARTS_AT]->(s)')
        session.run('MATCH (d:day1),(s:Second) WHERE d.Node_Key = s.Node_Key AND d.EndTime = s.EpochSeconds MERGE (d)-[:EVENT_ENDS_AT]->(s)')
        session.run('MATCH (d:day2),(s:Second) WHERE d.Node_Key = s.Node_Key AND d.EndTime = s.EpochSeconds MERGE (d)-[:EVENT_ENDS_AT]->(s)')
        session.run('MATCH (d:day1),(s:Second) WHERE d.Node_Key = s.Node_Key AND d.LunchStartTime = s.EpochSeconds MERGE (d)-[:LUNCH_STARTS_AT]->(s)')
        session.run('MATCH (d:day2),(s:Second) WHERE d.Node_Key = s.Node_Key AND d.LunchStartTime = s.EpochSeconds MERGE (d)-[:LUNCH_STARTS_AT]->(s)')
        session.run('MATCH (d:day1),(s:Second) WHERE d.Node_Key = s.Node_Key AND d.LunchEndTime = s.EpochSeconds MERGE (d)-[:LUNCH_ENDS_AT]->(s)')
        session.run('MATCH (d:day2),(s:Second) WHERE d.Node_Key = s.Node_Key AND d.LunchEndTime = s.EpochSeconds MERGE (d)-[:LUNCH_ENDS_AT]->(s)')

        # Add Subject ID to all nodes to facilitate data queries
        session.run('CALL apoc.periodic.iterate(\'MATCH (n),(d:day1) WHERE n.Node_Key = d.Node_Key RETURN n,d\',\'SET n.SubjectID = d.SubjectID, n.Condition = d.Condition, n.Environment = d.Environment, n.Day = "1"\',{batchSize:1000, parallel:true, iterateList:true})')
        session.run('CALL apoc.periodic.iterate(\'MATCH (n),(d:day2) WHERE n.Node_Key = d.Node_Key RETURN n,d\',\'SET n.SubjectID = d.SubjectID, n.Condition = d.Condition, n.Environment = d.Environment, n.Day = "2"\',{batchSize:1000, parallel:true, iterateList:true})')

set_start_end_time()
```

```
# Incorporate Netflows
# Note: If an error is returned on unterminated quoted field, open the spreadsheet with the error
and replace all "\" with "/"
# To speed import, CREATE is used instead of MERGE. Running this script twice for the same da
ta will result in duplicate entries



def import_netflows_to_neo4j(subjectID, directory, environment, day):
   with driver.session() as session:
      neodir = session.run('CALL dbms.listConfig() YIELD name, value WHERE name = \'dbms
.directories.neo4j_home\' RETURN value').values()
      path = neodir [0][0] + '/import/' + directory + '/'
      os.chdir(path)
      new_path = path.replace("\\","/")
      file_listing = os.listdir(new_path)
      i = 0
      for filename in file_listing:
         x = ''
         i = i + 1
         x = 'CALL apoc.periodic.iterate("CALL apoc.load.csv(\'file:///' + new_path + filename +
'\',{header:true}) YIELD map AS row","CREATE (c:NetFlow {SubjectID: \'' + subjectID + '\',
Environment: toString(' + environment + '), UDP_SPort:  coalesce(row.`udp.srcport`,\'\'), IP_Fla
gs:  coalesce(row.`ip.flags`,\'\'), EthDst:  coalesce(row.`eth.dst`,\'\'), TCP_DPort:  coalesce(row.`
tcp.dstport`,\'\'), Protocol:  coalesce(row.`_ws.col.Protocol`,\'\'), File_Source: \'' + filename + '\'
, Frame_Number:  coalesce(row.`frame.number`,\'\'), SourceIP:  coalesce(row.`ip.src`,\'\'), DestI
P:  coalesce(row.`ip.dst`,\'\'), TCP_Sport:  coalesce(row.`tcp.srcport`,\'\'),TCP_Stream: coalesce(
row.`tcp.stream`,\'\'),IP_Protocol:  coalesce(row.`ip.proto`,\'\'), Frame_Length:  coalesce(row.`fr
ame.len`,\'\'), EthSrc:  coalesce(row.`eth.src`,\'\'), UDP_Dport:  coalesce(row.`udp.dstport`,\'\'),
UDP_Stream: coalesce(row.`udp.stream`,\'\'),  EpochSeconds:  coalesce(toInteger(row.`frame.ti
me_epoch`),\'\')})",{batchSize:1000, parallel:true, iterateList:true})'
         print(x)

         #This executes the Cypher syntax and returns any errors
         result = session.run(x).data()
         session.run('MATCH (n:NetFlow) WHERE n.DestIP STARTS WITH "\"" RETURN n.Fi
le_Source, COUNT(n.DestIP)')
         session.close()
         print("Failed Batches – " + str(result [0]['failedBatches']))
         print("Error Messages – " + str(result [0]['batch']['errors']))
         print("Nodes Created – " + str(result [0]['updateStatistics']['nodesCreated']))
         print("Properties Set – " + str(result [0]['updateStatistics']['propertiesSet']))

      session.run("CALL db.checkpoint()")

# Initates script to import all network activity for subject specified by the user at the start of the N
otebook.
```

```
# SubjectID, Filenames and Environment Numbers are automatically set

import_netflows_to_neo4j(subject, filename_day1, environment_day1, '1')
import_netflows_to_neo4j(subject, filename_day2, environment_day2, '2')



# Assigns all NetFlow nodes the same UUID as the respective day they were captured during the experiment.
# This allows faster queries, allowing selection of only NetFlow nodes captured in day of interest

def set_netflow_node_key():
    with driver.session() as session:
        session.run('CALL apoc.periodic.iterate(\'MATCH (n:NetFlow),(d:day1) WHERE n.SubjectID = d.SubjectID AND n.Environment = d.Environment AND d.Day = "1" RETURN n,d\',\'SET n.Node_Key = d.Node_Key\', {batchSize:1000, parallel:true, iterateList:true})')
        session.run('CALL apoc.periodic.iterate(\'MATCH (n:NetFlow),(d:day2) WHERE n.SubjectID = d.SubjectID AND n.Environment = d.Environment AND d.Day = "2" RETURN n,d\',\'SET n.Node_Key = d.Node_Key\', {batchSize:1000, parallel:true, iterateList:true})')


set_netflow_node_key()



# Creates a [PACKET_SENT_AT] relationship between NetFlow nodes and the respective Second node they were captured.

def match_netflow_and_seconds():
    with driver.session() as session:
        session.run('CALL apoc.periodic.iterate(\'MATCH (n:NetFlow),(s:Second) WHERE s.Node_Key = n.Node_Key AND n.EpochSeconds = s.EpochSeconds RETURN n,s\',\'MERGE (n)-[:PACKET_SENT_AT]->(s)\',{batchSize:1000, parallel:false, iterateList:true})')

match_netflow_and_seconds()


# Creates an Event property to assign each Second node a value depending on if they occured before, during or after the experiment trial duration
# or if it occured during the scheduled lunch period

def set_events_by_second():

    with driver.session() as session:
        #Day 1
        session.run('CALL{ WITH ["S01","S02","S03","S04","S05","S06","S07","S08","S09","S10","S11","S12,"  "S13","S14","S15","S16","S17","S18","S19","S20","S21","S22","S23","S24","S25","S26","S27","S28","S29","S30","S31","S32","S33","S34","S35","S36","S37","S38","S39","S40","S41","S42","S43","S44","S45","S46","S47","S48","S49","S50"] AS subj
```

ects UNWIND subjects AS x MATCH (d:day1) WHERE d.SubjectID = x MATCH (s:Second)<-[:EVENT_STARTS_AT]-(d) WHERE s.Node_Key = d.Node_Key RETURN d.StartTime AS StartTime, d.EndTime AS EndTime, d.LunchStartTime AS LunchStartTime, d.LunchEndTime AS LunchEndTime, d.SubjectID sub, s } MATCH (a:Second) WHERE s.Node_Key = a.Node_Key FOREACH(test IN CASE WHEN a.EpochSeconds < StartTime THEN [1] ELSE [] END | SET a.Event = "Before-Event") FOREACH(test IN CASE WHEN a.EpochSeconds >= StartTime AND a.EpochSeconds < LunchStartTime THEN [1] ELSE [] END | SET a.Event = "During-Event") FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchStartTime AND a.EpochSeconds < LunchEndTime THEN [1] ELSE [] END | SET a.Event = "Lunch") FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchEndTime AND a.EpochSeconds <= EndTime THEN [1] ELSE [] END | SET a.Event = "During-Event") FOREACH(test IN CASE WHEN a.EpochSeconds > EndTime THEN [1] ELSE [] END | SET a.Event = "Post-Event") RETURN COUNT (*)')

session.run('CALL{ WITH ["S51","S52","S53","S54","S55","S56","S57","S58","S59","S60","S61","S62","S63","S64","S65","S66","S67","S68","S69","S70","S71","S72","S73","S74","S75","S76","S77","S78","S79","S80","S81","S82","S83","S84","S85","S86","S87","S88","S89","S90","S91","S92","S93","S94","S95","S96","S97","S98","S99","S100"] AS subjects UNWIND subjects AS x MATCH (d:day1) WHERE d.SubjectID = x MATCH (s:Second)<-[:EVENT_STARTS_AT]-(d) WHERE s.Node_Key = d.Node_Key RETURN d.StartTime AS StartTime, d.EndTime AS EndTime, d.LunchStartTime AS LunchStartTime, d.LunchEndTime AS LunchEndTime, d.SubjectID AS sub, s } MATCH (a:Second) WHERE s.Node_Key = a.Node_Key FOREACH(test IN CASE WHEN a.EpochSeconds < StartTime THEN [1] ELSE [] END | SET a.Event = "Before-Event") FOREACH(test IN CASE WHEN a.EpochSeconds >= StartTime AND a.EpochSeconds < LunchStartTime THEN [1] ELSE [] END | SET a.Event = "During-Event") FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchStartTime AND a.EpochSeconds < LunchEndTime THEN [1] ELSE [] END | SET a.Event = "Lunch") FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchEndTime AND a.EpochSeconds <= EndTime THEN [1] ELSE [] END | SET a.Event = "During-Event") FOREACH(test IN CASE WHEN a.EpochSeconds > EndTime THEN [1] ELSE [] END | SET a.Event = "Post-Event") RETURN COUNT (*)')

session.run('CALL{ WITH ["S101","S102","S103","S104","S105","S106","S107","S108","S109","S110","S111","S112","S113","S114","S115","S116","S117","S118","S119","S120","S121","S122","S123","S124","S125","S126","S127","S128","S129","S130","S131","S132","S133","S134","S135","S136","S137","S138","S139"] AS subjects UNWIND subjects AS x MATCH (d:day1) WHERE d.SubjectID = x MATCH (s:Second)<-[:EVENT_STARTS_AT]-(d) WHERE s.Node_Key = d.Node_Key RETURN d.StartTime AS StartTime, d.EndTime AS EndTime, d.LunchStartTime AS LunchStartTime, d.LunchEndTime AS LunchEndTime, d.SubjectID AS sub, s } MATCH (a:Second) WHERE s.Node_Key = a.Node_Key FOREACH(test IN CASE WHEN a.EpochSeconds < StartTime THEN [1] ELSE [] END | SET a.Event = "Before-Event") FOREACH(test IN CASE WHEN a.EpochSeconds >= StartTime AND a.EpochSeconds < LunchStartTime THEN [1] ELSE [] END | SET a.Event = "During-Event") FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchStartTime AND a.EpochSeconds < LunchEndTime THEN [1] ELSE [] END | SET a.Event = "Lunch") FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchEndTime AND a.EpochSeconds <= EndTime THEN [1] ELSE [] END | SET a.Event = "During-Event") FOREACH(test IN CASE WHEN a.EpochSeconds > EndTime THEN [1] ELSE [] END | SET a.Event = "Post-Event") RETURN COUNT (*)')

#Day 2

session.run('CALL{ WITH ["S01","S02","S03","S04","S05","S06","S07","S08","S09","S10","S11","S12","S13","S14","S15","S16","S17","S18","S19","S20","S21","S22","S23","S24","S25","S26","S27","S28","S29","S30","S31","S32","S33","S34","S35","S36","S37","S38","S39","S40","S41","S42","S43","S44","S45","S46","S47","S48","S49","S50"] AS subjects UNWIND subjects AS x MATCH (d:day2) WHERE d.SubjectID = x MATCH (s:Second)<-[:EVENT_

STARTS_AT]-(d) WHERE s.Node_Key = d.Node_Key RETURN d.StartTime AS StartTime, d.EndTime AS EndTime, d.LunchStartTime AS LunchStartTime, d.LunchEndTime AS LunchEndTime, d.SubjectID AS sub, s } MATCH (a:Second) WHERE s.Node_Key = a.Node_Key FOREACH(test IN CASE WHEN a.EpochSeconds < StartTime THEN [1] ELSE [] END | SET a.Event = "Before-Event") FOREACH(test IN CASE WHEN a.EpochSeconds >= StartTime AND a.EpochSeconds < LunchStartTime THEN [1] ELSE [] END | SET a.Event = "During-Event") FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchStartTime AND a.EpochSeconds < LunchEndTime THEN [1] ELSE [] END | SET a.Event = "Lunch") FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchEndTime AND a.EpochSeconds <= EndTime THEN [1] ELSE [] END | SET a.Event = "During-Event") FOREACH(test IN CASE WHEN a.EpochSeconds > EndTime THEN [1] ELSE [] END | SET a.Event = "Post-Event") RETURN COUNT (*)')

session.run('CALL{ WITH ["S51","S52","S53","S54","S55","S56","S57","S58","S59","S60","S61","S62","S63","S64","S65","S66","S67","S68","S69","S70","S71","S72","S73","S74","S75","S76","S77","S78","S79","S80","S81","S82","S83","S84","S85","S86","S87","S88","S89","S90","S91","S92","S93","S94","S95","S96","S97","S98","S99","S100"] AS subjects UNWIND subjects AS x MATCH (d:day2) WHERE d.SubjectID = x MATCH (s:Second)<-[:EVENT_STARTS_AT]-(d) WHERE s.Node_Key = d.Node_Key RETURN d.StartTime AS StartTime, d.EndTime AS EndTime, d.LunchStartTime AS LunchStartTime, d.LunchEndTime AS LunchEndTime, d.SubjectID AS sub, s } MATCH (a:Second) WHERE s.Node_Key = a.Node_Key FOREACH(test IN CASE WHEN a.EpochSeconds < StartTime THEN [1] ELSE [] END | SET a.Event = "Before-Event") FOREACH(test IN CASE WHEN a.EpochSeconds >= StartTime AND a.EpochSeconds < LunchStartTime THEN [1] ELSE [] END | SET a.Event = "During-Event") FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchStartTime AND a.EpochSeconds < LunchEndTime THEN [1] ELSE [] END | SET a.Event = "Lunch") FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchEndTime AND a.EpochSeconds <= EndTime THEN [1] ELSE [] END | SET a.Event = "During-Event") FOREACH(test IN CASE WHEN a.EpochSeconds > EndTime THEN [1] ELSE [] END | SET a.Event = "Post-Event") RETURN COUNT (*)')

session.run('CALL{ WITH ["S101","S102","S103","S104","S105","S106","S107","S108","S109","S110","S111","S112","S113","S114","S115","S116","S117","S118","S119","S120","S121","S122","S123","S124","S125","S126","S127","S128","S129","S130","S131","S132","S133","S134","S135","S136","S137","S138","S139"] AS subjects UNWIND subjects AS x MATCH (d:day2) WHERE d.SubjectID = x MATCH (s:Second)<-[:EVENT_STARTS_AT]-(d) WHERE s.Node_Key = d.Node_Key RETURN d.StartTime AS StartTime, d.EndTime AS EndTime, d.LunchStartTime AS LunchStartTime, d.LunchEndTime AS LunchEndTime, d.SubjectID AS sub, s } MATCH (a:Second) WHERE s.Node_Key = a.Node_Key FOREACH(test IN CASE WHEN a.EpochSeconds < StartTime THEN [1] ELSE [] END | SET a.Event = "Before-Event") FOREACH(test IN CASE WHEN a.EpochSeconds >= StartTime AND a.EpochSeconds < LunchStartTime THEN [1] ELSE [] END | SET a.Event = "During-Event") FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchStartTime AND a.EpochSeconds < LunchEndTime THEN [1] ELSE [] END | SET a.Event = "Lunch") FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchEndTime AND a.EpochSeconds <= EndTime THEN [1] ELSE [] END | SET a.Event = "During-Event") FOREACH(test IN CASE WHEN a.EpochSeconds > EndTime THEN [1] ELSE [] END | SET a.Event = "Post-Event") RETURN COUNT (*)')

set_events_by_second()

 # Creates an AbsoluteTime property for each Second node with the second value of elasped time since the start of the experiment trial

```
def set_absolute_time_for_subjects():
  with driver.session() as session:
    #Day 1
    session.run('CALL{ WITH ["S01","S02","S03","S04","S05","S06","S07","S08","S09","S
10","S11","S12","S13","S14","S15","S16","S17","S18","S19","S20","S21","S22","S23","S24"
,"S25","S26","S27","S28","S29","S30","S31","S32","S33","S34","S35","S36","S37","S38","S
39","S40","S41","S42","S43","S44","S45","S46","S47","S48","S49","S50"] AS subjects UNW
IND subjects AS x MATCH (d:day1) WHERE d.SubjectID = x MATCH (s:Second)<-[:EVENT_
STARTS_AT]-(d) WHERE s.Node_Key = d.Node_Key RETURN d.StartTime AS StartTime, d.
EndTime AS EndTime, d.LunchStartTime AS LunchStartTime, d.LunchEndTime AS LunchEnd
Time, d.SubjectID AS sub, s } MATCH (a:Second) WHERE s.Node_Key = a.Node_Key FORE
ACH(test IN CASE WHEN a.EpochSeconds < StartTime THEN [1] ELSE [] END | SET a.Absol
uteTime = a.EpochSeconds-StartTime ) FOREACH(test IN CASE WHEN a.EpochSeconds >= St
artTime AND a.EpochSeconds < LunchStartTime THEN [1] ELSE [] END | SET a.AbsoluteTim
e = a.EpochSeconds-StartTime ) FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchSt
artTime AND a.EpochSeconds < LunchEndTime THEN [1] ELSE [] END | SET a.AbsoluteTime
= StartTime – LunchStartTime) FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchEn
dTime AND a.EpochSeconds <= EndTime THEN [1] ELSE [] END | SET a.AbsoluteTime = a.E
pochSeconds-(StartTime+(LunchEndTime-LunchStartTime))) FOREACH(test IN CASE WHEN
a.EpochSeconds > EndTime THEN [1] ELSE [] END | SET a.AbsoluteTime = StartTime – EndTi
me) RETURN COUNT (*)')
    session.run('CALL{ WITH ["S51","S52","S53","S54","S55","S56","S57","S58","S59","S
60","S61","S62","S63","S64","S65","S66","S67","S68","S69","S70","S71","S72","S73","S74"
,"S75","S76","S77","S78","S79","S80","S81","S82","S83","S84","S85","S86","S87","S88","S
89","S90","S91","S92","S93","S94","S95","S96","S97","S98","S99","S100"] AS subjects UN
WIND subjects AS x MATCH (d:day1) WHERE d.SubjectID = x MATCH (s:Second)<-[:EVEN
T_STARTS_AT]-(d) WHERE s.Node_Key = d.Node_Key RETURN d.StartTime AS StartTime,
d.EndTime AS EndTime, d.LunchStartTime AS LunchStartTime, d.LunchEndTime AS LunchEn
dTime, d.SubjectID AS sub, s } MATCH (a:Second) WHERE s.Node_Key = a.Node_Key FORE
ACH(test IN CASE WHEN a.EpochSeconds < StartTime THEN [1] ELSE [] END | SET a.Absol
uteTime = a.EpochSeconds-StartTime ) FOREACH(test IN CASE WHEN a.EpochSeconds >= St
artTime AND a.EpochSeconds < LunchStartTime THEN [1] ELSE [] END | SET a.AbsoluteTim
e = a.EpochSeconds-StartTime ) FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchSt
artTime AND a.EpochSeconds < LunchEndTime THEN [1] ELSE [] END | SET a.AbsoluteTime
= StartTime – LunchStartTime) FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchEn
dTime AND a.EpochSeconds <= EndTime THEN [1] ELSE [] END | SET a.AbsoluteTime = a.E
pochSeconds-(StartTime+(LunchEndTime-LunchStartTime))) FOREACH(test IN CASE WHEN
a.EpochSeconds > EndTime THEN [1] ELSE [] END | SET a.AbsoluteTime = StartTime – EndTi
me) RETURN COUNT (*)')
    session.run('CALL{ WITH ["S101","S102","S103","S104","S105","S106","S107","S108",
"S109","S110","S111","S112","S113","S114","S115","S116","S117","S118","S119","S120","
S121","S122","S123","S124","S125","S126","S127","S128","S129","S130","S131","S132","S
133","S134","S135","S136","S137","S138","S139"] AS subjects UNWIND subjects AS x MAT
CH (d:day1) WHERE d.SubjectID = x MATCH (s:Second)<-[:EVENT_STARTS_AT]-(d) WHE
RE s.Node_Key = d.Node_Key RETURN d.StartTime AS StartTime, d.EndTime AS EndTime, d
.LunchStartTime AS LunchStartTime, d.LunchEndTime AS LunchEndTime, d.SubjectID AS sub
, s } MATCH (a:Second) WHERE s.Node_Key = a.Node_Key FOREACH(test IN CASE WHEN
a.EpochSeconds < StartTime THEN [1] ELSE [] END | SET a.AbsoluteTime = a.EpochSeconds-
StartTime ) FOREACH(test IN CASE WHEN a.EpochSeconds >= StartTime AND a.EpochSeco
```

```
nds < LunchStartTime THEN [1] ELSE [] END | SET a.AbsoluteTime = a.EpochSeconds-StartTi
me ) FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchStartTime AND a.EpochSecon
ds < LunchEndTime THEN [1] ELSE [] END | SET a.AbsoluteTime = StartTime – LunchStartTi
me) FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchEndTime AND a.EpochSecond
s <= EndTime THEN [1] ELSE [] END | SET a.AbsoluteTime = a.EpochSeconds-(StartTime+(L
unchEndTime-LunchStartTime))) FOREACH(test IN CASE WHEN a.EpochSeconds > EndTime
THEN [1] ELSE [] END | SET a.AbsoluteTime = StartTime – EndTime)')
    #Day 2
    session.run('CALL{ WITH ["S01","S02","S03","S04","S05","S06","S07","S08","S09","S
10","S11","S12","S13","S14","S15","S16","S17","S18","S19","S20","S21","S22","S23","S24"
,"S25","S26","S27","S28","S29","S30","S31","S32","S33","S34","S35","S36","S37","S38","S
39","S40","S41","S42","S43","S44","S45","S46","S47","S48","S49","S50"] AS subjects UNW
IND subjects AS x MATCH (d:day2) WHERE d.SubjectID = x MATCH (s:Second)<-[:EVENT_
STARTS_AT]-(d) WHERE s.Node_Key = d.Node_Key RETURN d.StartTime AS StartTime, d.
EndTime AS EndTime, d.LunchStartTime AS LunchStartTime, d.LunchEndTime AS LunchEnd
Time, d.SubjectID AS sub, s } MATCH (a:Second) WHERE s.Node_Key = a.Node_Key FORE
ACH(test IN CASE WHEN a.EpochSeconds < StartTime THEN [1] ELSE [] END | SET a.Absol
uteTime = a.EpochSeconds-StartTime ) FOREACH(test IN CASE WHEN a.EpochSeconds >= St
artTime AND a.EpochSeconds < LunchStartTime THEN [1] ELSE [] END | SET a.AbsoluteTim
e = a.EpochSeconds-StartTime ) FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchSt
artTime AND a.EpochSeconds < LunchEndTime THEN [1] ELSE [] END | SET a.AbsoluteTime
= StartTime – LunchStartTime) FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchEn
dTime AND a.EpochSeconds <= EndTime THEN [1] ELSE [] END | SET a.AbsoluteTime = a.E
pochSeconds-(StartTime+(LunchEndTime-LunchStartTime))) FOREACH(test IN CASE WHEN
a.EpochSeconds > EndTime THEN [1] ELSE [] END | SET a.AbsoluteTime = StartTime – EndTi
me) RETURN COUNT (*)')
    session.run('CALL{ WITH ["S51","S52","S53","S54","S55","S56","S57","S58","S59","S
60","S61","S62","S63","S64","S65","S66","S67","S68","S69","S70","S71","S72","S73","S74"
,"S75","S76","S77","S78","S79","S80","S81","S82","S83","S84","S85","S86","S87","S88","S
89","S90","S91","S92","S93","S94","S95","S96","S97","S98","S99","S100"] AS subjects UN
WIND subjects AS x MATCH (d:day2) WHERE d.SubjectID = x MATCH (s:Second)<-[:EVEN
T_STARTS_AT]-(d) WHERE s.Node_Key = d.Node_Key RETURN d.StartTime AS StartTime,
d.EndTime AS EndTime, d.LunchStartTime AS LunchStartTime, d.LunchEndTime AS LunchEn
dTime, d.SubjectID AS sub, s } MATCH (a:Second) WHERE s.Node_Key = a.Node_Key FORE
ACH(test IN CASE WHEN a.EpochSeconds < StartTime THEN [1] ELSE [] END | SET a.Absol
uteTime = a.EpochSeconds-StartTime ) FOREACH(test IN CASE WHEN a.EpochSeconds >= St
artTime AND a.EpochSeconds < LunchStartTime THEN [1] ELSE [] END | SET a.AbsoluteTim
e = a.EpochSeconds-StartTime ) FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchSt
artTime AND a.EpochSeconds < LunchEndTime THEN [1] ELSE [] END | SET a.AbsoluteTime
= StartTime – LunchStartTime) FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchEn
dTime AND a.EpochSeconds <= EndTime THEN [1] ELSE [] END | SET a.AbsoluteTime = a.E
pochSeconds-(StartTime+(LunchEndTime-LunchStartTime))) FOREACH(test IN CASE WHEN
a.EpochSeconds > EndTime THEN [1] ELSE [] END | SET a.AbsoluteTime = StartTime – EndTi
me) RETURN COUNT (*)')
    session.run('CALL{ WITH ["S101","S102","S103","S104","S105","S106","S107","S108",
"S109","S110","S111","S112","S113","S114","S115","S116","S117","S118","S119","S120","
S121","S122","S123","S124","S125","S126","S127","S128","S129","S130","S131","S132","S
133","S134","S135","S136","S137","S138","S139"] AS subjects UNWIND subjects AS x MAT
CH (d:day2) WHERE d.SubjectID = x MATCH (s:Second)<-[:EVENT_STARTS_AT]-(d) WHE
RE s.Node_Key = d.Node_Key RETURN d.StartTime AS StartTime, d.EndTime AS EndTime, d
```

```
.LunchStartTime AS LunchStartTime, d.LunchEndTime AS LunchEndTime, d.SubjectID AS sub
, s } MATCH (a:Second) WHERE s.Node_Key = a.Node_Key FOREACH(test IN CASE WHEN
a.EpochSeconds < StartTime THEN [1] ELSE [] END | SET a.AbsoluteTime = a.EpochSeconds-
StartTime ) FOREACH(test IN CASE WHEN a.EpochSeconds >= StartTime AND a.EpochSeco
nds < LunchStartTime THEN [1] ELSE [] END | SET a.AbsoluteTime = a.EpochSeconds-StartTi
me ) FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchStartTime AND a.EpochSecon
ds < LunchEndTime THEN [1] ELSE [] END | SET a.AbsoluteTime = StartTime – LunchStartTi
me) FOREACH(test IN CASE WHEN a.EpochSeconds >= LunchEndTime AND a.EpochSecond
s <= EndTime THEN [1] ELSE [] END | SET a.AbsoluteTime = a.EpochSeconds-(StartTime+(L
unchEndTime-LunchStartTime))) FOREACH(test IN CASE WHEN a.EpochSeconds > EndTime
THEN [1] ELSE [] END | SET a.AbsoluteTime = StartTime – EndTime)')

set_absolute_time_for_subjects()
```

```
# Adds a SourceIP_Validity and DestIP_Validity property to each NetFlow node to identify type o
f system that sent and recieved each packet

def add_system_properties_to_netflow():
    with driver.session() as session:
        x = 'CALL apoc.periodic.iterate(\'MATCH (n:NetFlow),(s:system) WHERE n.SubjectID =
s.SubjectID AND n.Node_Key =s.UUID AND n.SourceIP = s.IP_Address AND n.Environment
= s.Environment RETURN n,s\',\'SET n.SourceIP_Validity = s.Validity \',{batchSize:1000, paral
lel:false, iterateList:true})'
        print(x)
        result = session.run(x).data()
        session.close()
        print("Failed Batches – " + str(result [0]['failedBatches']))
        print("Error Messages – " + str(result [0]['batch']['errors']))
        print("Nodes Created – " + str(result [0]['updateStatistics']['nodesCreated']))
        print("Properties Set – " + str(result [0]['updateStatistics']['propertiesSet']))

        y = 'CALL apoc.periodic.iterate(\'MATCH (n:NetFlow),(s:system) WHERE n.SubjectID =
s.SubjectID AND n.Node_Key =s.UUID AND n.DestIP = s.IP_Address AND n.Environment = s.
Environment RETURN n,s\',\'SET n.DestIP_Validity = s.Validity \',{batchSize:1000, parallel:fal
se, iterateList:true})'
        print(y)
        result = session.run(y).data()
        session.close()
        print("Failed Batches – " + str(result [0]['failedBatches']))
        print("Error Messages – " + str(result [0]['batch']['errors']))
        print("Nodes Created – " + str(result [0]['updateStatistics']['nodesCreated']))
        print("Properties Set – " + str(result [0]['updateStatistics']['propertiesSet']))


add_system_properties_to_netflow()
```

```
# Creates an AbsoluteTime property for each NetFlow node with the second value of elasped time since the start of the experiment trial

def set_abs_time_for_packets():
    with driver.session() as session:
        session.run("CALL apoc.periodic.iterate('MATCH (n:NetFlow)-[:PACKET_SENT_AT]->(s:Second) RETURN n,s','SET n.AbsoluteTime = s.AbsoluteTime, n.EventTime = s.Event',{batchSize:1000, parallel:false, iterateList:true})")

set_abs_time_for_packets()
```

APPENDIX B. FACTOR ANALYSIS

A. DAY 1 ANALYSIS

1. Absent-Informed

a. Principal Components

(1) Correlations

	Latency	Frequency	Duration	Confusion	Self-Doubt	Confidence	Surprise	Frustrated
Latency	1	-0.0497	-0.2701	0.0671	-0.1114	0.1448	0.0422	-0.1289
Frequency	-0.0497	1	0.3208	0.6416	0.3875	-0.4844	0.1558	0.3123
Duration	-0.2701	0.3208	1	0.1117	0.3766	-0.2322	-0.1864	0.3397
Confusion	0.0671	0.6416	0.1117	1	0.3199	-0.4865	0.4828	0.3421
Self-Doubt	-0.1114	0.3875	0.3766	0.3199	1	-0.6391	0.0121	0.6521
Confidence	0.1448	-0.4844	-0.2322	-0.4865	-0.6391	1	-0.027	-0.3414
Surprise	0.0422	0.1558	-0.1864	0.4828	0.0121	-0.027	1	0.2949
Frustrated	-0.1289	0.3123	0.3397	0.3421	0.6521	-0.3414	0.2949	1
Cyber security	-0.0514	0.2064	0.1851	0.0696	-0.1277	0.0784	0.139	0.1722
Network penetration	-0.1433	0.1272	0.0898	0.0061	-0.1102	0.1352	0.0587	-0.0084
Host penetration	-0.0723	0.0585	-0.0012	-0.0425	-0.2258	0.1916	0.0469	-0.1052
Network reconnaissance	-0.0671	-0.0092	-0.0107	-0.0491	-0.1682	0.2173	0.2945	-0.006
Incidence response	0.043	0.268	0.0981	0.0871	0.1033	0.0442	0.1682	0.2549
Generalized defense practice	0.0999	0.2234	0.0598	0.19	0.1009	-0.0773	0.2881	0.2873
Network protocol reverse engineering	0.3733	0.0361	-0.404	0.1903	-0.3411	0.2769	0.1972	-0.0556
Binary reverse engineering	0.1979	0.097	0.0555	0.2001	0.0386	0.1776	-0.0845	0.2151

	Cyber security	Network penetration	Host penetration	Network reconnaissance	Incidence response	Generalized defense practice	Network protocol reverse engineering	Binary reverse engineering
Latency	-0.0514	-0.1433	-0.0723	-0.0671	0.043	0.0999	0.3733	0.1979
Frequency	0.2064	0.1272	0.0585	-0.0092	0.268	0.2234	0.0361	0.097
Duration	0.1851	0.0898	-0.0012	-0.0107	0.0981	0.0598	-0.404	0.0555
Confusion	0.0696	0.0061	-0.0425	-0.0491	0.0871	0.19	0.1903	0.2001
Self-Doubt	-0.1277	-0.1102	-0.2258	-0.1682	0.1033	0.1009	-0.3411	0.0386
Confidence	0.0784	0.1352	0.1916	0.2173	0.0442	-0.0773	0.2769	0.1776
Surprise	0.139	0.0587	0.0469	0.2945	0.1682	0.2881	0.1972	-0.0845

Frustrated	0.1722	-0.0084	-0.1052	-0.006	0.2549	0.2873	-0.0556	0.2151
Cyber security	1	0.6928	0.6154	0.6321	0.7069	0.6816	0.4303	0.6076
Network penetration	0.6928	1	0.8217	0.8225	0.6965	0.6403	0.5768	0.5111
Host penetration	0.6154	0.8217	1	0.8987	0.4982	0.3949	0.534	0.4765
Network reconnaissance	0.6321	0.8225	0.8987	1	0.5674	0.4868	0.4209	0.3904
Incidence response	0.7069	0.6965	0.4982	0.5674	1	0.8467	0.377	0.5602
Generalized defense practice	0.6816	0.6403	0.3949	0.4868	0.8467	1	0.4861	0.4816
Network protocol reverse engineering	0.4303	0.5768	0.534	0.4209	0.377	0.4861	1	0.4824
Binary reverse engineering	0.6076	0.5111	0.4765	0.3904	0.5602	0.4816	0.4824	1

(2) Loading Matrix

	Prin1	**Prin2**	**Prin3**	**Prin4**	**Prin5**
Latency	0.05337	-0.19256	0.63066	0.53950	-0.04218
Frequency	0.21361	0.69903	0.13398	-0.06722	-0.47542
Duration	0.03224	0.51423	-0.59768	0.11970	-0.08429
Confusion	0.16249	0.65511	0.52406	-0.17386	-0.30441
Self-Doubt	-0.09458	0.81663	-0.13340	0.18363	0.16001
Confidence	0.14472	-0.76171	-0.01076	0.08027	0.25780
Surprise	0.24387	0.22993	0.55016	-0.59242	0.39355
Frustrated	0.16090	0.72223	0.00336	0.10625	0.48665
Cyber_security	0.84557	0.05095	-0.16273	0.05505	0.01966
Network_penetration	0.89287	-0.09842	-0.23119	-0.11705	-0.13023
Host_penetration	0.80715	-0.24075	-0.20661	-0.20844	-0.23962
Network_reconnaissance	0.82068	-0.19042	-0.16623	-0.33913	0.03951
Incidence_response	0.83171	0.17132	-0.07015	0.16582	0.19210
Generalized_defense_practice	0.79723	0.22119	0.10074	0.12725	0.23614
Network_protocol_reverse_engineering	0.66580	-0.28979	0.49525	0.09699	-0.14826
Binary_reverse_engineering	0.68962	0.03282	0.03128	0.49833	-0.04228

b. Exploratory Factor Analysis

(1) Unrotated Factor Loading

EFA was applied using "Principal Axis" factor method with "Principal Components" for prior communality so the Unrotated Factor Loading is the same as the PCA Loading Matrix above.

(2) Target Matrix

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	-0.00053	-0.00001	-0.00025	1.00000	-0.00000
Frequency	0.00622	1.00000	0.00407	-0.00003	-0.00012
Duration	0.00342	0.03262	0.16872	-0.16840	-0.19156
Confusion	0.00006	0.80678	0.00265	0.00819	0.04919
Self-Doubt	-0.00246	0.12138	0.72226	-0.00418	-0.00232
Confidence	0.00993	-0.71139	-0.06778	0.01307	0.00021
Surprise	0.00225	0.00552	0.00300	0.00000	1.00000
Frustrated	0.00177	0.01422	1.00000	-0.00012	0.01124
Cyber_security	1.00000	0.00006	0.00280	0.00005	-0.00003
Network_penetration	0.98566	0.00003	-0.00325	-0.00098	-0.00001
Host_penetration	0.81169	0.00000	-0.05752	-0.00220	-0.00000
Network_reconnaissance	0.80720	-0.00135	-0.00810	-0.00857	0.01741
Incidence_response	0.78566	0.00001	0.06239	0.00421	0.00007
Generalized_defense_practice	0.62929	0.00069	0.08046	0.01779	0.00956
Network_protocol_reverse_engineering	0.27046	0.00018	-0.04601	0.26724	0.01786
Binary_reverse_engineering	0.47596	0.00008	0.01283	0.14010	-0.02799

(3) Factor Structure

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	-0.023754	-0.056133	-0.065331	0.833722	0.081089
Frequency	0.171858	0.872087	0.345442	-0.015783	-0.010629
Duration	0.080218	0.310190	0.495616	-0.459321	-0.503160
Confusion	0.073879	0.847102	0.303702	0.212278	0.364845
Self-Doubt	-0.122915	0.514368	0.818503	-0.153105	-0.181871
Confidence	0.175085	-0.746771	-0.486516	0.213677	0.083757
Surprise	0.165435	0.215728	0.115691	0.098194	0.921608
Frustrated	0.112979	0.339469	0.851834	-0.005638	0.131987
Cyber_security	0.854111	0.076554	0.120609	0.120881	0.047214
Network_penetration	0.922440	0.044638	-0.116129	0.003044	0.068103
Host_penetration	0.845016	-0.005921	-0.318098	-0.032519	0.085975
Network_reconnaissance	0.847857	-0.079986	-0.202273	-0.077027	0.296815
Incidence_response	0.817962	0.090697	0.321485	0.235052	0.096381
Generalized_defense_practice	0.760188	0.149174	0.349357	0.323972	0.240545
Network_protocol_reverse_engineering	0.614050	0.012616	-0.299441	0.645740	0.362789
Binary_reverse_engineering	0.670232	0.062322	0.198321	0.486887	-0.156060

(4) Rotation Matrix

0.93664	-0.01781	0.00262	-0.03540	-0.02783
-0.03358	1.06836	-0.12005	0.02808	-0.03399
0.00202	-0.27928	0.79933	-0.00074	0.05133
-0.09434	0.01780	-0.01793	0.71684	-0.06547
-0.07428	-0.06270	0.07477	-0.07397	0.76467

(5) Final Communality Estimates

Latency	0.73049
Frequency	0.78276
Duration	0.64413
Confusion	0.85310
Self-Doubt	0.75294
Confidence	0.67416
Surprise	0.92085
Frustrated	0.79564
Cyber_security	0.74749
Network_penetration	0.89102
Host_penetration	0.85301
Network_reconnaissance	0.85397
Incidence_response	0.79041
Generalized_defense_practice	0.76660
Network_protocol_reverse_engineering	0.80394
Binary_reverse_engineering	0.72776

(6) Standard Score Coefficients

Latency	-0.037476	-0.002344	0.042240	0.535882	-0.053050
Frequency	0.030807	0.442427	-0.055136	-0.013531	-0.083274
Duration	0.039801	0.079400	0.161497	-0.221036	-0.314966
Confusion	-0.006904	0.405319	-0.045263	0.103289	0.175583
Self-Doubt	-0.026101	0.078998	0.331130	-0.010955	-0.094430
Confidence	0.032926	-0.325096	-0.055865	0.087047	0.058878
Surprise	0.005177	-0.003904	0.035228	-0.043413	0.656480
Frustrated	0.008609	-0.079473	0.421311	0.034911	0.124626
Cyber_security	0.168201	-0.007479	0.052553	-0.002684	-0.046980
Network_penetration	0.189292	0.043400	-0.091945	-0.107773	-0.035798
Host_penetration	0.176953	0.075810	-0.201897	-0.140863	-0.022162
Network_reconnaissance	0.172420	-0.034480	-0.106462	-0.181682	0.160364
Incidence_response	0.152573	-0.066465	0.181653	0.086794	-0.008253
Generalized_defense_practice	0.133256	-0.051199	0.192074	0.136947	0.088915
Network_protocol_reverse_engineering	0.096591	0.071312	-0.135367	0.298857	0.109800
Binary_reverse_engineering	0.120751	-0.006989	0.128735	0.283568	-0.229667

(7) Variance Explained by Each Factor

Factor	Variance	Percent	Cum Percent
Factor 1	5.2144	32.590	32.590
Factor 2	2.6103	16.314	48.905
Factor 3	2.6317	16.448	65.352
Factor 4	1.8660	11.663	77.015
Factor 5	1.6253	10.158	87.173

(8) Rotated Factor Loading

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	-0.155006	0.011379	-0.073118	0.874686	-0.081105
Frequency	0.138035	0.869117	0.000962	0.004211	-0.077693
Duration	0.193795	0.165680	0.375780	-0.400998	-0.410921
Confusion	-0.038276	0.830105	0.023045	0.186725	0.290824
Self-Doubt	-0.107737	0.244910	0.714611	-0.108940	-0.066718
Confidence	0.178413	-0.666783	-0.223707	0.155117	0.035932
Surprise	0.045330	0.078274	0.208058	-0.094782	0.956704
Frustrated	0.091105	-0.024351	0.892762	-0.063069	0.248284
Cyber_security	0.863441	-0.009752	0.118124	-0.007610	-0.054408
Network_penetration	0.950092	0.050540	-0.143064	-0.139406	-0.056237
Host_penetration	0.874189	0.085770	-0.358417	-0.164049	-0.052687
Network_reconnaissance	0.865199	-0.089205	-0.141344	-0.264184	0.216748
Incidence_response	0.803871	-0.085461	0.358100	0.101818	0.016293
Generalized_defense_practice	0.711601	-0.038816	0.386943	0.179069	0.159012
Network_protocol_reverse_engineering	0.501320	0.120891	-0.323199	0.538038	0.138826
Binary_reverse_engineering	0.641101	-0.002062	0.163251	0.446072	-0.311665

2. Absent-Uninformed

a. Principal Components

(1) Correlations

	Latency	Frequency	Duration	Confusion	Self-Doubt	Confidence	Surprise	Frustrated
Latency	1.0000	-0.3024	-0.4723	-0.1040	-0.1202	0.0305	-0.1322	-0.0740
Frequency	-0.3024	1.0000	0.2913	-0.2112	0.2519	-0.0893	-0.1388	0.1446
Duration	-0.4723	0.2913	1.0000	-0.2355	-0.1187	-0.1485	-0.0831	-0.2598
Confusion	-0.1040	-0.2112	-0.2355	1.0000	0.2557	-0.0687	0.3010	0.6010
Self-Doubt	-0.1202	0.2519	-0.1187	0.2557	1.0000	-0.2568	0.4680	0.1168
Confidence	0.0305	-0.0893	-0.1485	-0.0687	-0.2568	1.0000	-0.1310	-0.0755
Surprise	-0.1322	-0.1388	-0.0831	0.3010	0.4680	-0.1310	1.0000	0.2342
Frustrated	-0.0740	0.1446	-0.2598	0.6010	0.1168	-0.0755	0.2342	1.0000
Cyber security	-0.0184	0.3248	0.0712	-0.3693	-0.1772	0.3973	0.0303	-0.1609
Network penetration	-0.2047	0.2404	0.0359	-0.1205	-0.1918	0.5092	-0.0503	-0.0400
Host penetration	-0.0668	0.1558	-0.0429	-0.2271	-0.3573	0.5838	-0.1256	-0.1016
Network reconnaissance	-0.0742	0.0737	0.0525	-0.2041	-0.4564	0.3612	-0.0313	-0.0736
Incidence response	0.3239	-0.2239	-0.2430	-0.0377	-0.4084	0.1844	-0.1956	-0.2361
Generalized defense practice	0.2588	-0.1610	-0.2455	-0.2606	-0.3436	0.3492	-0.1787	-0.3704
Network protocol reverse engineering	0.3733	0.0361	-0.404	0.1903	-0.3411	0.2769	0.1972	-0.0556

Binary reverse engineering	0.1979	0.097	0.0555	0.2001	0.0386	0.1776	-0.0845	0.2151

	Cyber security	Network penetration	Host penetration	Network reconnaissance	Incidence response	Generalized defense practice	Network protocol reverse engineering	Binary reverse engineering
Latency	-0.0184	-0.2047	-0.0668	-0.0742	0.3239	0.2588	0.0384	-0.0033
Frequency	0.3248	0.2404	0.1558	0.0737	-0.2239	-0.161	0.2139	0.3655
Duration	0.0712	0.0359	-0.0429	0.0525	-0.243	-0.2455	-0.0633	-0.0316
Confusion	-0.3693	-0.1205	-0.2271	-0.2041	-0.0377	-0.2606	-0.1511	-0.1265
Self-Doubt	-0.1772	-0.1918	-0.3573	-0.4564	-0.4084	-0.3436	-0.3047	-0.3213
Confidence	0.3973	0.5092	0.5838	0.3612	0.1844	0.3492	0.4951	0.4688
Surprise	0.0303	-0.0503	-0.1256	-0.0313	-0.1956	-0.1787	-0.0346	-0.1769
Frustrated	-0.1609	-0.04	-0.1016	-0.0736	-0.2361	-0.3704	0.0137	0.1295
Cyber security	1	0.6532	0.6379	0.7513	0.2628	0.6025	0.5139	0.6186
Network penetration	0.6532	1	0.874	0.7928	0.1366	0.3129	0.4565	0.698
Host penetration	0.6379	0.874	1	0.7872	0.2412	0.4124	0.5732	0.7461
Network reconnaissance	0.7513	0.7928	0.7872	1	0.3935	0.5541	0.5753	0.6139
Incidence response	0.2628	0.1366	0.2412	0.3935	1	0.6955	0.272	0.1543
Generalized defense practice	0.6025	0.3129	0.4124	0.5541	0.6955	1	0.3341	0.2447
Network protocol reverse engineering	0.5139	0.4565	0.5732	0.5753	0.272	0.3341	1	0.7384
Binary reverse engineering	0.6186	0.698	0.7461	0.6139	0.1543	0.2447	0.7384	1

There was 1 missing value in the data. The correlations are estimated by Robust method.

(2) Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5
Latency	0.03348	-0.66588	0.22775	-0.04356	0.49682
Frequency	0.17307	0.65637	-0.26961	-0.04451	0.53212
Duration	-0.02054	0.41412	-0.69215	-0.05317	-0.35414
Confusion	-0.34621	0.14678	0.72378	-0.14057	-0.30866
Self-Doubt	-0.49453	0.37779	0.19006	0.54476	0.33767
Confidence	0.61876	-0.03503	0.21740	-0.11425	-0.09646
Surprise	-0.22475	0.28318	0.44022	0.66217	-0.19271
Frustrated	-0.19675	0.39655	0.68219	-0.34948	0.05441
Cyber_security	0.81185	0.15195	-0.06591	0.35852	0.13793
Network_penetration	0.80924	0.34534	0.09831	0.07558	-0.14208
Host_penetration	0.88407	0.17207	0.08871	-0.01984	-0.08713
Network_reconnaissance	0.87340	0.07180	0.05656	0.11971	-0.23347
Incidence_response	0.45370	-0.63755	0.10820	0.06402	-0.07948
Generalized_defense_practice	0.64678	-0.53878	-0.00109	0.31022	0.00292

Network_protocol_reverse_engineering	0.72624	0.12586	0.18218	-0.11984	0.16669
Binary_reverse_engineering	0.79720	0.30966	0.16741	-0.25382	0.21105

There was 1 missing value in the data. The correlations are estimated by Robust method.

b. Exploratory Factor Analysis

(1) Unrotated Factor Loading

EFA was applied using "Principal Axis" factor method with "Principal Components" for prior communality so the Unrotated Factor Loading is the same as the PCA Loading Matrix above.

(2) Target Matrix

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**
Latency	-0.00193	-0.00324	1.00000	-0.00109	-0.00751
Frequency	0.01615	-0.00098	-0.00754	1.00000	0.00000
Duration	-0.00027	-0.04045	-0.85253	0.00477	-0.00700
Confusion	-0.00239	0.82449	0.00001	-0.02883	0.02430
Self-Doubt	-0.05172	0.00100	0.00030	0.11406	0.57892
Confidence	0.83008	0.00085	0.00438	-0.01644	-0.01130
Surprise	-0.00000	0.01313	-0.00084	-0.00260	1.00000
Frustrated	0.00015	1.00000	0.00086	0.01208	0.00087
Cyber_security	0.67202	-0.09209	0.00015	0.00326	0.00835
Network_penetration	0.96172	-0.00004	-0.00999	0.00006	0.00020
Host_penetration	1.00000	-0.00072	-0.00017	-0.00002	-0.00150
Network_reconnaissance	0.88891	-0.00803	-0.00111	-0.01303	-0.00000
Incidence_response	0.05964	-0.04373	0.17207	-0.30789	-0.01018
Generalized_defense_practice	0.17078	-0.19081	0.09524	-0.10449	0.00000
Network_protocol_reverse_engineering	0.86858	0.00003	0.01119	0.00620	-0.00698
Binary_reverse_engineering	0.76377	0.00198	0.00069	0.03812	-0.01688

(3) Factor Structure

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**
Latency	-0.049249	-0.152416	0.831194	-0.216796	-0.177374
Frequency	0.201513	-0.054312	-0.211961	0.854355	-0.028253
Duration	-0.049491	-0.256618	-0.798092	0.233455	-0.162494
Confusion	-0.205647	0.806077	-0.003734	-0.183743	0.315303
Self-Doubt	-0.398483	0.224494	-0.023001	0.439360	0.756512
Confidence	0.631745	-0.058664	0.161438	-0.192941	-0.194731
Surprise	-0.087782	0.268939	-0.110444	-0.064230	0.872432
Frustrated	-0.050574	0.852346	0.045006	0.243243	0.135823
Cyber_security	0.811961	-0.486332	0.092677	0.069876	0.076324

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Network_penetration	0.862484	-0.146146	-0.130420	0.033485	-0.025075
Host_penetration	0.904593	-0.218941	0.018825	-0.052792	-0.182094
Network_reconnaissance	0.885292	-0.314996	-0.013881	-0.226924	-0.096432
Incidence_response	0.374006	-0.374713	0.491398	-0.598230	-0.214972
Generalized_defense_practice	0.565599	-0.603474	0.463812	-0.504655	-0.083335
Network_protocol_reverse_engineering	0.744295	-0.095223	0.211831	0.085206	-0.211257
Binary_reverse_engineering	0.831718	-0.015050	0.118988	0.257657	-0.297903

(4) Rotation Matrix

1.00254	0.08433	-0.02863	0.01504	0.04719
0.14185	0.86490	0.01654	-0.04634	-0.06599
-0.05927	0.01033	0.95677	0.04779	0.02727
0.00839	-0.06117	0.14248	0.70353	-0.01254
0.07636	-0.05690	0.01773	0.01307	0.87888

(5) Final Communality Estimates

Latency	0.74511
Frequency	0.81860
Duration	0.77923
Confusion	0.78030
Self-Doubt	0.83419
Confidence	0.45370
Surprise	0.80011
Frustrated	0.78645
Cyber_security	0.83410
Network_penetration	0.80969
Host_penetration	0.82704
Network_reconnaissance	0.84001
Incidence_response	0.63443
Generalized_defense_practice	0.80486
Network_protocol_reverse_engineering	0.61860
Binary_reverse_engineering	0.86841

(6) Standard Score Coefficients

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	-0.034712	-0.066600	0.496098	0.067130	-0.090236
Frequency	0.029286	-0.049205	0.050637	0.532586	-0.039396
Duration	-0.019282	-0.124880	-0.462886	-0.000162	-0.102027
Confusion	0.007577	0.389711	-0.049146	-0.191087	0.094391
Self-Doubt	-0.049663	-0.051628	0.123327	0.243079	0.459593
Confidence	0.124157	0.074332	0.023498	-0.091674	-0.067175
Surprise	0.032720	0.022254	-0.066010	-0.135877	0.587894
Frustrated	0.027471	0.432474	0.067821	0.115030	-0.051084
Cyber_security	0.152354	-0.194274	0.050384	0.078420	0.199286

Network_penetration	0.177826	0.019905	-0.119779	-0.014877	0.075771
Host_penetration	0.175921	0.008817	-0.046649	-0.025855	-0.017757
Network_reconnaissance	0.174543	-0.047956	-0.102475	-0.150448	0.066436
Incidence_response	0.055715	-0.109028	0.167443	-0.263930	-0.030819
Generalized_defense_practice	0.091097	-0.237562	0.171456	-0.203977	0.112589
Network_protocol_reverse_engineering	0.140614	0.052150	0.116076	0.110152	-0.077124
Binary_reverse_engineering	0.157834	0.110151	0.086073	0.204019	-0.154883

(7) Variance Explained by Each Factor

Factor	Variance	Percent	Cum Percent
Factor 1	5.3666	33.541	33.541
Factor 2	2.5134	15.709	49.250
Factor 3	1.9554	12.221	61.471
Factor 4	1.9052	11.908	73.379
Factor 5	1.7833	11.146	84.525

(8) Rotated Factor Loading

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	-0.191415	-0.112094	0.829201	-0.025646	-0.134267
Frequency	0.238395	-0.133291	-0.051995	0.880698	0.004431
Duration	-0.063629	-0.320838	-0.803534	0.104138	-0.173226
Confusion	0.012585	0.826336	-0.000367	-0.315038	0.186828
Self-Doubt	-0.276448	-0.022452	0.148636	0.436310	0.714851
Confidence	0.634949	0.151451	0.062431	-0.152664	-0.112610
Surprise	0.087235	0.160119	-0.079849	-0.123272	0.857339
Frustrated	0.175630	0.880971	0.126942	0.148336	0.019452
Cyber_security	0.769441	-0.356829	0.038766	0.183087	0.254383
Network_penetration	0.916530	0.050372	-0.210751	0.043675	0.095208
Host_penetration	0.910244	0.017253	-0.083573	-0.005849	-0.047682
Network_reconnaissance	0.875170	-0.081813	-0.153644	-0.185513	0.049449
Incidence_response	0.234308	-0.197644	0.338677	-0.472335	-0.110473
Generalized_defense_practice	0.407107	-0.436786	0.316138	-0.339502	0.083394
Network_protocol_reverse_engineering	0.747777	0.101734	0.167788	0.164639	-0.104843
Binary_reverse_engineering	0.864365	0.198651	0.097571	0.317649	-0.198173

3. Present-Informed

a. Principal Components

(1) Correlations

	Latency	Frequency	Duration	Confusion	Self-Doubt	Confidence	Surprise	Frustrated
Latency	1	0.0591	-0.345	0.0072	0.2633	0.2031	0.0522	0.081
Frequency	0.0591	1	0.4036	0.4382	0.0896	0.0825	0.2548	0.0944
Duration	-0.345	0.4036	1	0.0569	-0.1998	-0.0975	0.0149	-0.1686

Confusion	0.0072	0.4382	0.0569	1	0.4003	-0.2154	0.2998	0.4469
Self-Doubt	0.2633	0.0896	-0.1998	0.4003	1	-0.5206	0.3005	0.5183
Confidence	0.2031	0.0825	-0.0975	-0.2154	-0.5206	1	-0.1203	-0.3401
Surprise	0.0522	0.2548	0.0149	0.2998	0.3005	-0.1203	1	0.692
Frustrated	0.081	0.0944	-0.1686	0.4469	0.5183	-0.3401	0.692	1
Cyber security	0.089	0.3154	-0.1737	-0.0267	-0.0888	0.3204	0.3111	0.3445
Network penetration	0.1566	-0.1712	-0.2462	-0.1596	-0.1762	0.2293	0.0785	0.2294
Host penetration	0.029	-0.181	-0.1766	-0.0789	-0.2296	0.2769	0.1868	0.2277
Network reconnaissance	0.1251	-0.1641	-0.2857	-0.3296	-0.2658	0.2971	0.1984	0.3148
Incidence response	0.134	0.293	-0.211	-0.1211	-0.0136	0.0467	0.1695	0.218
Generalized defense practice	0.2476	0.0482	-0.2873	-0.0726	0.0579	0.0705	0.2172	0.1031
Network protocol reverse engineering	0.1079	-0.0143	-0.1547	0.0909	-0.2817	0.2437	-0.0195	0.097
Binary reverse engineering	-0.1125	0.1279	-0.0456	-0.0433	-0.3017	0.1879	0.1296	0.1082

	Cyber security	**Network penetration**	**Host penetration**	**Network reconnaissance**	**Incidence response**	**Generalized defense practice**	**Network protocol reverse engineering**	**Binary reverse engineering**
Latency	0.089	0.1566	0.029	0.1251	0.134	0.2476	0.1079	-0.1125
Frequency	0.3154	-0.1712	-0.181	-0.1641	0.293	0.0482	-0.0143	0.1279
Duration	-0.1737	-0.2462	-0.1766	-0.2857	-0.211	-0.2873	-0.1547	-0.0456
Confusion	-0.0267	-0.1596	-0.0789	-0.3296	-0.1211	-0.0726	0.0909	-0.0433
Self-Doubt	-0.0888	-0.1762	-0.2296	-0.2658	-0.0136	0.0579	-0.2817	-0.3017
Confidence	0.3204	0.2293	0.2769	0.2971	0.0467	0.0705	0.2437	0.1879
Surprise	0.3111	0.0785	0.1868	0.1984	0.1695	0.2172	-0.0195	0.1296
Frustrated	0.3445	0.2294	0.2277	0.3148	0.218	0.1031	0.097	0.1082
Cyber security	1	0.5178	0.4288	0.6872	0.662	0.2991	0.3058	0.3853
Network penetration	0.5178	1	0.8914	0.6835	0.4215	0.323	0.7423	0.5871
Host penetration	0.4288	0.8914	1	0.5711	0.2502	0.2235	0.8021	0.7354
Network reconnaissance	0.6872	0.6835	0.5711	1	0.539	0.4299	0.4817	0.42
Incidence response	0.662	0.4215	0.2502	0.539	1	0.659	0.3381	0.3796
Generalized defense practice	0.2991	0.323	0.2235	0.4299	0.659	1	0.2809	0.1748
Network protocol reverse engineering	0.3058	0.7423	0.8021	0.4817	0.3381	0.2809	1	0.8296
Binary reverse engineering	0.3853	0.5871	0.7354	0.42	0.3796	0.1748	0.8296	1

(2) Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5
Latency	0.17486	0.18976	-0.50128	0.22003	0.63524
Frequency	0.00238	0.40446	0.60650	0.57327	0.24886
Duration	-0.31666	-0.08660	0.72019	0.14398	-0.19678
Confusion	-0.13897	0.63596	0.37666	-0.17514	0.43511
Self-Doubt	-0.23891	0.76521	-0.30608	-0.17874	0.11738
Confidence	0.34948	-0.47709	0.05778	0.38206	0.36941
Surprise	0.24282	0.69860	0.14912	-0.02735	-0.19736
Frustrated	0.29579	0.80968	-0.00818	-0.29394	-0.17013
Cyber_security	0.71992	0.22315	0.09201	0.37950	-0.15817
Network_penetration	0.87112	-0.10171	-0.04414	-0.25635	0.03711
Host_penetration	0.83749	-0.13030	0.13850	-0.40939	0.07799
Network_reconnaissance	0.82568	-0.04648	-0.19679	0.09006	-0.29153
Incidence_response	0.65512	0.23895	-0.08719	0.49041	-0.20145
Generalized_defense_practice	0.51543	0.20488	-0.33456	0.36464	-0.05058
Network_protocol_reverse_engineering	0.78749	-0.15844	0.23186	-0.29810	0.31475
Binary_reverse_engineering	0.73741	-0.12840	0.41386	-0.21305	0.06888

b. Exploratory Factor Analysis

(1) Unrotated Factor Loading

EFA was applied using "Principal Axis" factor method with "Principal Components" for prior communality so the Unrotated Factor Loading is the same as the PCA Loading Matrix above.

(2) Target Matrix

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	-0.00000	0.00357	-0.00010	0.00192	1.00000
Frequency	-0.00109	0.01494	0.00011	1.00000	-0.00060
Duration	-0.00365	-0.01151	-0.00381	0.18415	-0.50338
Confusion	0.00091	-0.04697	0.25769	0.36588	0.00698
Self-Doubt	-0.03906	-0.00136	0.68617	0.00015	0.08270
Confidence	0.04126	0.01856	-0.67129	0.01631	0.02279
Surprise	0.00146	0.07579	0.73362	0.02720	-0.00121
Frustrated	0.00987	0.01612	1.00000	0.00016	0.00011
Cyber_security	0.06124	0.77805	0.00060	0.01247	-0.00000
Network_penetration	0.73872	0.06104	0.00000	-0.01411	0.00268
Host_penetration	0.97574	0.00696	0.00008	-0.00392	0.00001
Network_reconnaissance	0.15940	0.52588	-0.00000	-0.03887	0.00000
Incidence_response	0.00675	1.00000	0.00023	0.00210	0.00066
Generalized_defense_practice	0.00126	0.79695	0.00037	-0.00034	0.07544
Network_protocol_reverse_engineering	1.00000	0.00131	-0.00086	0.00049	0.00270

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Binary_reverse_engineering	0.90535	0.01089	-0.00021	0.00380	-0.00642

(3) Factor Structure

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	0.042836	0.148083	0.030278	0.070395	0.859299
Frequency	-0.058951	0.160476	0.076782	0.926763	-0.054044
Duration	-0.182063	-0.243281	-0.151185	0.471610	-0.657213
Confusion	0.013122	-0.299527	0.559895	0.642262	0.205974
Self-Doubt	-0.301878	-0.163582	0.768335	0.106289	0.443277
Confidence	0.316030	0.273656	-0.636849	0.117028	0.167944
Surprise	0.140088	0.313738	0.666027	0.272418	-0.001345
Frustrated	0.230090	0.259701	0.874579	0.103901	0.135800
Cyber_security	0.481015	0.820670	0.075923	0.171901	0.032166
Network_penetration	0.869150	0.556450	0.001628	-0.259518	0.134136
Host_penetration	0.943479	0.413254	0.016125	-0.190095	0.020234
Network_reconnaissance	0.617005	0.816173	-0.022557	-0.325964	0.040398
Incidence_response	0.335301	0.863708	0.065915	0.091501	0.120530
Generalized_defense_practice	0.223906	0.678114	0.076788	-0.067760	0.342951
Network_protocol_reverse_engineering	0.925265	0.335104	-0.091673	0.012051	0.118613
Binary_reverse_engineering	0.851421	0.388253	-0.072880	0.089150	-0.160651

(4) Rotation Matrix

1.02168	-0.20236	0.00917	0.00152	-0.02113
-0.26992	0.96957	0.00824	0.02625	-0.02209
0.02294	-0.01202	0.99993	-0.02672	-0.06205
0.04758	0.00000	-0.08314	0.72024	0.03261
0.01808	-0.04740	-0.09449	-0.00597	0.82791

(5) Final Communality Estimates

Latency	0.76981
Frequency	0.92201
Duration	0.68590
Confusion	0.78562
Self-Doubt	0.78203
Confidence	0.63552
Surprise	0.60894
Frustrated	0.85848
Cyber_security	0.74559
Network_penetration	0.83823
Host_penetration	0.91123
Network_reconnaissance	0.81574
Incidence_response	0.77497
Generalized_defense_practice	0.55510
Network_protocol_reverse_engineering	0.88694

Binary_reverse_engineering	0.78167

(6) Standard Score Coefficients

Latency	-0.001666	-0.022801	-0.050437	0.115688	0.587812
Frequency	-0.026503	0.075734	-0.056590	0.528694	0.003204
Duration	-0.016783	-0.019473	-0.064089	0.235684	-0.381297
Confusion	0.091736	-0.189380	0.183502	0.345972	0.175327
Self-Doubt	-0.062883	-0.060425	0.289460	-0.003180	0.238598
Confidence	0.061198	0.045247	-0.294530	0.174137	0.179423
Surprise	0.009629	0.103933	0.253790	0.078522	-0.092643
Frustrated	0.053746	0.043917	0.356087	-0.026795	-0.021834
Cyber_security	0.028160	0.273379	-0.001807	0.097552	-0.062471
Network_penetration	0.213237	0.051837	0.024998	-0.105724	0.040055
Host_penetration	0.270063	-0.026451	0.040831	-0.066422	-0.011987
Network_reconnaissance	0.064742	0.249707	0.005685	-0.177538	-0.078232
Incidence_response	-0.036599	0.323055	-0.011569	0.047776	-0.019439
Generalized_defense_practice	-0.044660	0.242202	-0.002309	-0.028323	0.142959
Network_protocol_reverse_engineering	0.283427	-0.072499	-0.027159	0.078265	0.096241
Binary_reverse_engineering	0.245017	-0.004021	-0.014090	0.089076	-0.107326

(7) Variance Explained by Each Factor

Factor	Variance	Percent	Cum Percent
Factor 1	4.3034	26.896	26.896
Factor 2	3.7530	23.456	50.353
Factor 3	2.5780	16.113	66.466
Factor 4	1.8691	11.682	78.148
Factor 5	1.6519	10.324	88.472

(8) Rotated Factor Loading

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	-0.026491	0.099283	-0.129007	0.104787	0.876028
Frequency	-0.124027	0.277716	-0.024880	0.937988	-0.050890
Duration	-0.075679	-0.134233	-0.102781	0.459756	-0.619432
Confusion	0.226603	-0.376086	0.470775	0.581962	0.158486
Self-Doubt	-0.263065	-0.067430	0.703074	0.010452	0.339028
Confidence	0.231116	0.160134	-0.701529	0.226869	0.269987
Surprise	0.028056	0.327232	0.665930	0.215250	-0.135101
Frustrated	0.170059	0.190469	0.882343	0.024142	-0.031644
Cyber_security	0.151241	0.769227	0.063934	0.220634	-0.038081
Network_penetration	0.768265	0.191492	0.031344	-0.201485	0.086456
Host_penetration	0.945795	-0.019576	0.060484	-0.139043	-0.020828
Network_reconnaissance	0.301480	0.666072	0.023824	-0.269315	-0.028490
Incidence_response	-0.059898	0.895618	0.043654	0.138537	0.049351
Generalized_defense_practice	-0.101811	0.700438	0.032201	-0.029930	0.286968
Network_protocol_reverse_engineering	0.972457	-0.107124	-0.094605	0.079579	0.113145

Binary_reverse_engineering	0.853535	0.026276	-0.036685	0.145632	-0.181070

4. Present-Uninformed

a. Principal Components

(1) Correlations

	Latency	Frequency	Duration	Confusion	Self-Doubt	Confidence	Surprise	Frustrated
Latency	1	-0.0609	-0.4337	0.1022	0.018	-0.3846	-0.1724	-0.2483
Frequency	-0.0609	1	0.2459	0.0127	-0.0052	0.2114	-0.2427	0.0048
Duration	-0.4337	0.2459	1	-0.0247	-0.0623	0.5191	-0.1999	-0.0545
Confusion	0.1022	0.0127	-0.0247	1	0.5051	-0.0591	0.1148	0.2475
Self-Doubt	0.018	-0.0052	-0.0623	0.5051	1	-0.0494	0.1679	0.2575
Confidence	-0.3846	0.2114	0.5191	-0.0591	-0.0494	1	-0.4569	-0.6178
Surprise	-0.1724	-0.2427	-0.1999	0.1148	0.1679	-0.4569	1	0.4542
Frustrated	-0.2483	0.0048	-0.0545	0.2475	0.2575	-0.6178	0.4542	1
Cyber security	-0.094	0.24	0.4007	0.2362	0.1312	0.346	-0.3005	-0.188
Network penetration	-0.2989	-0.0112	0.5545	0.1457	0.1177	0.5824	-0.1307	-0.2224
Host penetration	-0.2915	0.1494	0.6377	0.055	0.0209	0.5746	-0.2935	-0.1869
Network reconnaissance	-0.405	0.2009	0.5834	0.0676	0.1413	0.6287	-0.2238	-0.2198
Incidence response	-0.2884	0.1827	0.252	-0.1238	0.0607	0.566	-0.1407	-0.3231
Generalized defense practice	-0.2607	0.1809	0.3016	-0.0533	0.142	0.3918	-0.1014	-0.271
Network protocol reverse engineering	-0.1829	-0.1746	0.2464	-0.0636	0.0787	0.2954	-0.2494	-0.1025
Binary reverse engineering	-0.2172	-0.0103	0.3645	0.0327	0.1153	0.3607	-0.5709	-0.0468

	Cyber security	Network penetration	Host penetration	Network reconnaissance	Incidence response	Generalized defense practice	Network protocol reverse engineering	Binary reverse engineering
Latency	-0.094	-0.2989	-0.2915	-0.405	-0.2884	-0.2607	-0.1829	-0.2172
Frequency	0.24	-0.0112	0.1494	0.2009	0.1827	0.1809	-0.1746	-0.0103
Duration	0.4007	0.5545	0.6377	0.5834	0.252	0.3016	0.2464	0.3645
Confusion	0.2362	0.1457	0.055	0.0676	-0.1238	-0.0533	-0.0636	0.0327
Self-Doubt	0.1312	0.1177	0.0209	0.1413	0.0607	0.142	0.0787	0.1153
Confidence	0.346	0.5824	0.5746	0.6287	0.566	0.3918	0.2954	0.3607
Surprise	-0.3005	-0.1307	-0.2935	-0.2238	-0.1407	-0.1014	-0.2494	-0.5709
Frustrated	-0.188	-0.2224	-0.1869	-0.2198	-0.3231	-0.271	-0.1025	-0.0468
Cyber security	1	0.4803	0.5273	0.331	0.382	0.6056	0.2441	0.4778

Network penetration	0.4803	1	0.9194	0.8127	0.4012	0.396	0.4908	0.5643
Host penetration	0.5273	0.9194	1	0.7586	0.4951	0.3931	0.5385	0.6402
Network reconnaissance	0.331	0.8127	0.7586	1	0.5132	0.4979	0.6207	0.6002
Incidence response	0.382	0.4012	0.4951	0.5132	1	0.7286	0.5399	0.3519
Generalized defense practice	0.6056	0.396	0.3931	0.4979	0.7286	1	0.3781	0.3194
Network protocol reverse engineering	0.2441	0.4908	0.5385	0.6207	0.5399	0.3781	1	0.7804
Binary reverse engineering	0.4778	0.5643	0.6402	0.6002	0.3519	0.3194	0.7804	1

There were 3 missing values in the data. The correlations are estimated by Robust method.

(2) Loading Matrix

	Prin1	**Prin2**	**Prin3**	**Prin4**	**Prin5**
Latency	-0.39866	-0.26601	0.71921	-0.05559	0.02154
Frequency	0.20259	-0.12263	-0.00052	0.72851	-0.24130
Duration	0.65861	0.06699	-0.31765	0.18623	-0.41031
Confusion	0.00774	0.64835	0.48102	0.23721	-0.13955
Self-Doubt	0.06993	0.68148	0.37787	0.14410	0.21515
Confidence	0.75946	-0.32265	-0.08135	0.18515	-0.04536
Surprise	-0.42028	0.52910	-0.47456	0.06230	0.34727
Frustrated	-0.34707	0.72713	-0.25438	-0.06190	-0.19148
Cyber_security	0.62707	0.10477	0.33861	0.33277	0.03856
Network_penetration	0.83052	0.21186	-0.03154	-0.11307	-0.14676
Host_penetration	0.87453	0.10879	-0.02845	-0.07469	-0.22379
Network_reconnaissance	0.86519	0.14628	-0.11622	-0.05957	-0.06392
Incidence_response	0.69311	-0.10538	-0.09034	0.11211	0.56832
Generalized_defense_practice	0.65162	0.01041	-0.00061	0.27917	0.57461
Network_protocol_reverse_engineering	0.66662	0.10283	0.08065	-0.57828	0.20073
Binary_reverse_engineering	0.73644	0.10338	0.25593	-0.42401	-0.17076

b. Exploratory Factor Analysis

(1) Unrotated Factor Loading

EFA was applied using "Principal Axis" factor method with "Principal Components" for prior communality so the Unrotated Factor Loading is the same as the PCA Loading Matrix above.

(2) Target Matrix

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	-0.19202	-0.00868	0.01619	-0.60359	-0.00386
Frequency	0.00150	0.00442	0.00039	-0.00213	1.00000
Duration	0.72793	0.00014	-0.00226	0.00515	0.10379
Confusion	0.00049	-0.00467	1.00000	-0.00000	0.00286
Self-Doubt	0.00000	0.00755	0.96101	0.00676	-0.00135
Confidence	0.34887	0.15930	-0.02609	-0.03030	0.04267
Surprise	-0.07069	0.00002	0.00436	1.00000	-0.00259
Frustrated	-0.00007	-0.15213	0.08622	0.56844	-0.00083
Cyber_security	0.16174	0.17128	0.10826	-0.05452	0.04339
Network_penetration	0.97150	0.02268	0.00299	-0.00000	-0.00001
Host_penetration	1.00000	0.01635	0.00019	-0.00125	0.00032
Network_reconnaissance	0.87492	0.06807	0.00009	0.00005	0.00001
Incidence_response	0.04800	0.94689	-0.00052	-0.00000	-0.00011
Generalized_defense_practice	0.02398	1.00000	0.00177	0.00001	0.00065
Network_protocol_reverse_engineering	0.35031	0.04968	0.00001	-0.00248	-0.27868
Binary_reverse_engineering	0.70055	0.00130	0.00417	-0.05320	-0.04642

(3) Factor Structure

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	-0.470704	-0.252152	0.234771	-0.589585	-0.190007
Frequency	0.153059	0.167275	-0.003847	-0.095240	0.778820
Duration	0.741014	0.242389	-0.146472	0.068051	0.435587
Confusion	0.047957	-0.128032	0.824947	0.027769	0.055114
Self-Doubt	0.043181	0.120870	0.789849	0.141383	-0.135855
Confidence	0.681536	0.603506	-0.287323	-0.324135	0.338199
Surprise	-0.390514	-0.136248	0.151294	0.825795	-0.122228
Frustrated	-0.169469	-0.487892	0.394305	0.687005	-0.105772
Cyber_security	0.531149	0.531020	0.329110	-0.326002	0.268139
Network_penetration	0.866632	0.447425	0.107479	-0.098783	0.013197
Host_penetration	0.910798	0.452457	0.030718	-0.179917	0.097969
Network_reconnaissance	0.872606	0.547404	0.016002	-0.069902	0.062198
Incidence_response	0.501534	0.904511	-0.104490	-0.105197	0.005965
Generalized_defense_practice	0.446748	0.898281	0.065622	-0.077828	0.109681
Network_protocol_reverse_engineering	0.656992	0.466277	0.043435	-0.200920	-0.545899
Binary_reverse_engineering	0.782039	0.309495	0.154821	-0.380510	-0.299569

(4) Rotation Matrix

0.99362	-0.22571	-0.02268	0.04226	-0.03072
-0.27115	0.98213	0.02423	0.05760	-0.00900
-0.00390	0.03489	0.94228	0.00291	0.05068
0.06295	0.06711	-0.02405	0.92802	-0.01741
-0.03361	-0.02422	0.06906	-0.02832	0.72196

(5) Final Communality Estimates

Latency	0.75051
Frequency	0.64503
Duration	0.74220
Confusion	0.72754
Self-Doubt	0.67914
Confidence	0.72384
Surprise	0.80627
Frustrated	0.75438
Cyber_security	0.63107
Network_penetration	0.76997
Host_penetration	0.83311
Network_reconnaissance	0.79109
Incidence_response	0.83523
Generalized_defense_practice	0.83283
Network_protocol_reverse_engineering	0.83616
Binary_reverse_engineering	0.82748

(6) Standard Score Coefficients

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**
Latency	-0.119489	-0.029871	0.188108	-0.419664	-0.115415
Frequency	0.012968	0.020863	0.017340	-0.025736	0.538577
Duration	0.192080	-0.106902	-0.102777	0.120953	0.281392
Confusion	0.019136	-0.082610	0.465641	-0.047177	0.069962
Self-Doubt	-0.025442	0.091262	0.440179	0.038763	-0.079715
Confidence	0.100205	0.119311	-0.142473	-0.082383	0.181534
Surprise	-0.082384	0.102916	0.030598	0.446932	-0.056853
Frustrated	0.044339	-0.199880	0.163268	0.335593	-0.022007
Cyber_security	0.051182	0.137104	0.212876	-0.142655	0.155635
Network_penetration	0.183856	-0.008359	0.049198	0.021244	-0.027371
Host_penetration	0.196740	-0.029555	0.011578	-0.017691	0.028479
Network_reconnaissance	0.169721	0.051729	-0.002140	0.053168	-0.000908
Incidence_response	-0.012192	0.400725	-0.046012	0.042120	-0.061430
Generalized_defense_practice	-0.032147	0.416083	0.052826	0.043742	0.018420
Network_protocol_reverse_engineering	0.114764	0.086468	0.019291	-0.053825	-0.425500
Binary_reverse_engineering	0.176377	-0.073563	0.092513	-0.169603	-0.242516

(7) Variance Explained by Each Factor

Factor	**Variance**	**Percent**	**Cum Percent**
Factor 1	5.5625	34.765	34.765
Factor 2	3.7269	23.293	58.059
Factor 3	1.8041	11.275	69.334
Factor 4	1.9966	12.479	81.813
Factor 5	1.4798	9.248	91.061

(8) Rotated Factor Loading

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	-0.522620	-0.103539	0.229944	-0.694151	-0.079768
Frequency	0.024072	0.091093	0.118941	-0.106112	0.791525
Duration	0.806807	-0.149055	-0.102564	0.160027	0.354824
Confusion	0.104600	-0.138592	0.842745	-0.007325	0.177199
Self-Doubt	-0.025781	0.214004	0.793648	0.157184	-0.038993
Confidence	0.478740	0.306346	-0.223382	-0.196639	0.250620
Surprise	-0.341919	0.188021	0.124005	0.805842	-0.115664
Frustrated	0.140523	-0.417846	0.344464	0.628668	-0.062476
Cyber_security	0.303459	0.348665	0.400237	-0.235482	0.284374
Network_penetration	0.853180	0.053813	0.108824	0.045844	-0.050683
Host_penetration	0.896196	0.014051	0.043043	-0.036623	0.026433
Network_reconnaissance	0.802334	0.178958	0.028375	0.089775	-0.018821
Incidence_response	0.098013	0.870575	-0.060990	0.069561	-0.072306
Generalized_defense_practice	0.016206	0.908893	0.130188	0.081211	0.063147
Network_protocol_reverse_engineering	0.603442	0.206307	-0.026601	-0.045631	-0.614567
Binary_reverse_engineering	0.821074	-0.103342	0.111940	-0.258877	-0.338869

5. Combined Day 1

a. Principal Components

(1) Correlations

	Latency	Frequency	Duration	Confusion	Self-Doubt	Confidence	Surprise	Frustrated
Latency	1	-0.1249	-0.3948	0.0364	0.0727	-0.007	-0.1137	-0.0685
Frequency	-0.1249	1	0.3231	0.1655	0.1208	-0.0304	0.0041	0.1047
Duration	-0.3948	0.3231	1	-0.0748	-0.0917	0.015	-0.0595	-0.0795
Confusion	0.0364	0.1655	-0.0748	1	0.3728	-0.2579	0.2735	0.4682
Self-Doubt	0.0727	0.1208	-0.0917	0.3728	1	-0.4401	0.2587	0.3858
Confidence	-0.007	-0.0304	0.015	-0.2579	-0.4401	1	-0.1738	-0.2841
Surprise	-0.1137	0.0041	-0.0595	0.2735	0.2587	-0.1738	1	0.4156
Frustrated	-0.0685	0.1047	-0.0795	0.4682	0.3858	-0.2841	0.4156	1
Cyber security	-0.0543	0.2521	0.1192	-0.0885	-0.1035	0.2764	0.1218	0.0559
Network penetration	-0.1535	0.0505	0.1071	-0.1053	-0.1741	0.3852	0.0422	-0.0238
Host penetration	-0.129	0.0676	0.1232	-0.1577	-0.2782	0.4476	-0.004	-0.0717
Network reconnaissance	-0.1474	0.0237	0.0824	-0.1747	-0.2759	0.3856	0.1415	-0.0053
Incidence response	0.023	0.1432	-0.0579	-0.0797	-0.1118	0.15	0.0919	0.0064
Generalized defense practice	0.0706	0.0653	-0.0607	-0.0794	-0.0849	0.1956	0.1127	-0.0698
Network protocol reverse engineering	-0.0036	0.0362	-0.038	-0.0343	-0.2361	0.3857	0.0448	-0.0128

Binary reverse engineering	-0.0845	0.1768	0.0977	-0.0149	-0.1905	0.3394	-0.0847	0.0971

	Cyber security	Network penetration	Host penetration	Network reconnaissance	Incidence response	Generalized defense practice	Network protocol reverse engineering	Binary reverse engineering
Latency	-0.0543	-0.1535	-0.129	-0.1474	0.023	0.0706	-0.0036	-0.0845
Frequency	0.2521	0.0505	0.0676	0.0237	0.1432	0.0653	0.0362	0.1768
Duration	0.1192	0.1071	0.1232	0.0824	-0.0579	-0.0607	-0.038	0.0977
Confusion	-0.0885	-0.1053	-0.1577	-0.1747	-0.0797	-0.0794	-0.0343	-0.0149
Self-Doubt	-0.1035	-0.1741	-0.2782	-0.2759	-0.1118	-0.0849	-0.2361	-0.1905
Confidence	0.2764	0.3852	0.4476	0.3856	0.15	0.1956	0.3857	0.3394
Surprise	0.1218	0.0422	-0.004	0.1415	0.0919	0.1127	0.0448	-0.0847
Frustrated	0.0559	-0.0238	-0.0717	-0.0053	0.0064	-0.0698	-0.0128	0.0971
Cyber security	1	0.6161	0.5705	0.6372	0.5025	0.5661	0.4236	0.5389
Network penetration	0.6161	1	0.8832	0.7901	0.3921	0.4122	0.5905	0.6102
Host penetration	0.5705	0.8832	1	0.7473	0.3208	0.354	0.6413	0.6775
Network reconnaissance	0.6372	0.7901	0.7473	1	0.4956	0.5044	0.5525	0.515
Incidence response	0.5025	0.3921	0.3208	0.4956	1	0.7229	0.3376	0.2924
Generalized defense practice	0.5661	0.4122	0.354	0.5044	0.7229	1	0.3586	0.2725
Network protocol reverse engineering	0.4236	0.5905	0.6413	0.5525	0.3376	0.3586	1	0.7229
Binary reverse engineering	0.5389	0.6102	0.6775	0.515	0.2924	0.2725	0.7229	1

There were 8 missing values in the data. The correlations are estimated by Robust method.

(2) Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5
Latency	-0.12336	-0.07342	0.68807	-0.04828	0.47832
Frequency	0.13097	0.30850	-0.47561	0.37726	0.54912
Duration	0.11651	-0.04716	-0.79707	0.27017	-0.01691
Confusion	-0.21012	0.66871	-0.01381	-0.24418	0.24753
Self-Doubt	-0.35689	0.63498	0.05471	0.02330	0.11763
Confidence	0.53664	-0.43655	0.01841	-0.17422	0.05290
Surprise	0.01607	0.63979	0.07527	-0.03966	-0.51943
Frustrated	-0.09297	0.77116	-0.03445	-0.26540	-0.04957
Cyber_security	0.76279	0.23319	-0.00085	0.23340	0.06360
Network_penetration	0.86652	0.07553	-0.07263	-0.17023	-0.08947
Host_penetration	0.87144	-0.02603	-0.11258	-0.24976	-0.02123
Network_reconnaissance	0.85854	0.07228	0.02015	-0.00947	-0.24005
Incidence_response	0.59405	0.20295	0.32365	0.52782	-0.01517
Generalized_defense_practice	0.61622	0.17303	0.38087	0.50976	-0.04639

Network_protocol_reverse_engineering	0.74668	0.04617	0.08226	-0.33999	0.15572
Binary_reverse_engineering	0.74894	0.09945	-0.13568	-0.31902	0.31154

b. Exploratory Factor Analysis

(1) Unrotated Factor Loading

EFA was applied using "Principal Axis" factor method with "Principal Components" for prior communality so the Unrotated Factor Loading is the same as the PCA Loading Matrix above.

(2) Target Matrix

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**
Latency	-0.00053	-0.00001	-0.00025	1.00000	-0.00000
Frequency	0.00622	1.00000	0.00407	-0.00003	-0.00012
Duration	0.00342	0.03262	0.16872	-0.16840	-0.19156
Confusion	0.00006	0.80678	0.00265	0.00819	0.04919
Self-Doubt	-0.00246	0.12138	0.72226	-0.00418	-0.00232
Confidence	0.00993	-0.71139	-0.06778	0.01307	0.00021
Surprise	0.00225	0.00552	0.00300	0.00000	1.00000
Frustrated	0.00177	0.01422	1.00000	-0.00012	0.01124
Cyber_security	1.00000	0.00006	0.00280	0.00005	-0.00003
Network_penetration	0.98566	0.00003	-0.00325	-0.00098	-0.00001
Host_penetration	0.81169	0.00000	-0.05752	-0.00220	-0.00000
Network_reconnaissance	0.80720	-0.00135	-0.00810	-0.00857	0.01741
Incidence_response	0.78566	0.00001	0.06239	0.00421	0.00007
Generalized_defense_practice	0.62929	0.00069	0.08046	0.01779	0.00956
Network_protocol_reverse_engineering	0.27046	0.00018	-0.04601	0.26724	0.01786
Binary_reverse_engineering	0.47596	0.00008	0.01283	0.14010	-0.02799

(3) Factor Structure

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**
Latency	-0.112561	-0.010968	-0.045312	0.836362	0.038272
Frequency	0.086777	0.129528	0.168593	-0.136449	0.819678
Duration	0.076819	-0.028396	-0.131780	-0.655378	0.445617
Confusion	-0.083121	-0.148031	0.717739	0.166730	0.207287
Self-Doubt	-0.323148	-0.061803	0.674791	0.134927	0.150762
Confidence	0.553746	0.145617	-0.500136	-0.021177	-0.078057
Surprise	-0.004549	0.292910	0.643883	-0.327310	-0.406538
Frustrated	0.013858	0.001299	0.809207	-0.067527	-0.023790
Cyber_security	0.653036	0.707233	0.024121	-0.224878	0.126559
Network_penetration	0.869471	0.516796	-0.071115	-0.314524	-0.105579
Host_penetration	0.905634	0.421506	-0.159448	-0.279295	-0.065524
Network_reconnaissance	0.791601	0.659012	-0.092638	-0.363541	-0.223920

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Incidence_response	0.369739	0.864496	-0.013191	-0.039840	-0.002738
Generalized_defense_practice	0.389153	0.885412	-0.040738	-0.017357	-0.064940
Network_protocol_reverse_engineering	0.819303	0.331515	-0.052799	0.009360	-0.032399
Binary_reverse_engineering	0.843846	0.260617	-0.017033	-0.062312	0.211407

(4) Rotation Matrix

1.08794	-0.20761	0.09068	0.04376	-0.02870
-0.27674	1.02108	-0.04526	0.03940	0.04489
0.06712	-0.04266	1.02377	0.01944	0.00528
0.11094	0.07929	0.02036	0.85740	-0.04723
0.00382	0.07177	-0.01838	-0.01597	0.77982

(5) Final Communality Estimates

Latency	0.72518
Frequency	0.78239
Duration	0.72440
Confusion	0.61241
Self-Doubt	0.54793
Confidence	0.51205
Surprise	0.68663
Frustrated	0.67741
Cyber_security	0.69475
Network_penetration	0.79881
Host_penetration	0.83559
Network_reconnaissance	0.80044
Incidence_response	0.77765
Generalized_defense_practice	0.81674
Network_protocol_reverse_engineering	0.70626
Binary_reverse_engineering	0.78805

(6) Standard Score Coefficients

Latency	-0.004998	-0.002462	-0.031312	0.590409	0.145477
Frequency	-0.006213	0.027531	0.042776	0.033597	0.658949
Duration	-0.013889	-0.021798	-0.070494	-0.398870	0.283705
Confusion	0.048856	-0.101311	0.311465	0.146127	0.164079
Self-Doubt	-0.057743	0.028718	0.272217	0.077120	0.108536
Confidence	0.139653	-0.063974	-0.181591	0.053210	-0.029034
Surprise	-0.022825	0.172031	0.295488	-0.286784	-0.396787
Frustrated	0.055119	-0.032083	0.365001	-0.039618	-0.055997
Cyber_security	0.086743	0.219190	0.032516	-0.043350	0.110415
Network_penetration	0.196781	0.052198	0.024182	-0.108216	-0.088124
Host_penetration	0.223412	-0.014282	-0.010816	-0.072464	-0.048152
Network_reconnaissance	0.139769	0.166204	0.008237	-0.173004	-0.191292
Incidence_response	-0.037795	0.398040	-0.007310	0.026870	0.024882

Generalized_defense_practice	-0.034357	0.406147	-0.016202	0.036875	-0.021371
Network_protocol_reverse_engineering	0.228937	-0.056602	0.031656	0.128443	0.010819
Binary_reverse_engineering	0.245671	-0.110960	0.041818	0.115817	0.200394

(7) Variance Explained by Each Factor

Factor	Variance	Percent	Cum Percent
Factor 1	4.7450	29.656	29.656
Factor 2	3.2388	20.242	49.899
Factor 3	2.3824	14.890	64.789
Factor 4	1.6713	10.446	75.235
Factor 5	1.2411	7.757	82.991

(8) Rotated Factor Loading

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	-0.015324	0.163741	-0.026894	0.866733	-0.022813
Frequency	0.001125	0.201631	0.152333	-0.173822	0.862254
Duration	-0.040793	-0.081272	-0.157419	-0.732211	0.504026
Confusion	0.117515	-0.176877	0.743738	0.162641	0.169168
Self-Doubt	-0.248692	0.057916	0.645917	0.103654	0.141318
Confidence	0.545805	-0.071280	-0.430535	0.067519	-0.080714
Surprise	-0.078271	0.204613	0.619128	-0.249244	-0.359418
Frustrated	0.153599	-0.111397	0.831200	-0.029529	-0.034517
Cyber_security	0.416807	0.540588	0.047834	-0.054687	0.211678
Network_penetration	0.782181	0.143294	0.011060	-0.125139	-0.057819
Host_penetration	0.868235	0.021399	-0.061268	-0.101229	-0.033952
Network_reconnaissance	0.599318	0.347937	-0.041651	-0.163332	-0.149531
Incidence_response	-0.003865	0.906574	-0.048643	0.125645	0.104351
Generalized_defense_practice	0.012082	0.919872	-0.073659	0.159124	0.041131
Network_protocol_reverse_engineering	0.869648	-0.018324	0.056156	0.183543	-0.033374
Binary_reverse_engineering	0.920529	-0.100727	0.097641	0.084013	0.210069

B. DAY 2 ANALYSIS

1. Absent-Informed / Absent-Uninformed

a. Principal Components

(1) Correlations

	Latency	Frequency	Duration	Confusion	Self-Doubt	Confidence	Surprise	Frustrated
Latency	1	0.0363	-0.2781	0.1009	0.066	-0.231	0.0937	0.0406
Frequency	0.0363	1	-0.163	-0.0358	0.2261	-0.3565	-0.1141	-0.1059
Duration	-0.2781	-0.163	1	-0.2238	-0.0946	0.1932	-0.145	-0.0861
Confusion	0.1009	-0.0358	-0.2238	1	0.7728	-0.2198	0.6059	0.5377

Self-Doubt	0.066	0.2261	-0.0946	0.7728	1	-0.4264	0.3035	0.4885
Confidence	-0.231	-0.3565	0.1932	-0.2198	-0.4264	1	0.0867	-0.071
Surprise	0.0937	-0.1141	-0.145	0.6059	0.3035	0.0867	1	0.27
Frustrated	0.0406	-0.1059	-0.0861	0.5377	0.4885	-0.071	0.27	1
Cyber security	-0.2973	0.1718	0.1539	0.0499	0.0909	-0.3074	0.2803	-0.0443
Network penetration	-0.2845	-0.1073	0.0225	0.1227	-0.0772	-0.1619	0.2086	-0.3116
Host penetration	0.0553	-0.1596	-0.0561	0.0593	-0.1235	-0.1097	0.3592	-0.2033
Network reconnaissance	0.08	-0.0591	0.0383	0.132	-0.1132	-0.1392	0.3812	-0.3012
Incidence response	-0.2514	0.3041	0.0151	-0.0971	-0.1089	-0.134	-0.1351	-0.3562
Generalized defense practice	-0.3567	0.3503	0.0697	-0.1323	-0.0056	-0.2661	-0.2985	-0.1702
Network protocol reverse engineering	-0.1991	-0.2499	-0.1666	0.0055	-0.0115	0.1986	0.0692	0.0443
Binary reverse engineering	-0.2182	0.1487	0.1487	-0.0171	-0.0532	0.2165	0.2287	0.0151

	Cyber security	**Network penetration**	**Host penetration**	**Network reconnaissance**	**Incidence response**	**Generalized defense practice**	**Network protocol reverse engineering**	**Binary reverse engineering**
Latency	-0.2973	-0.2845	0.0553	0.08	-0.2514	-0.3567	-0.1991	-0.2182
Frequency	0.1718	-0.1073	-0.1596	-0.0591	0.3041	0.3503	-0.2499	0.1487
Duration	0.1539	0.0225	-0.0561	0.0383	0.0151	0.0697	-0.1666	0.1487
Confusion	0.0499	0.1227	0.0593	0.132	-0.0971	-0.1323	0.0055	-0.0171
Self-Doubt	0.0909	-0.0772	-0.1235	-0.1132	-0.1089	-0.0056	-0.0115	-0.0532
Confidence	-0.3074	-0.1619	-0.1097	-0.1392	-0.134	-0.2661	0.1986	0.2165
Surprise	0.2803	0.2086	0.3592	0.3812	-0.1351	-0.2985	0.0692	0.2287
Frustrated	-0.0443	-0.3116	-0.2033	-0.3012	-0.3562	-0.1702	0.0443	0.0151
Cyber security	1	0.6702	0.59	0.5889	0.6595	0.6141	0.3822	0.5724
Network penetration	0.6702	1	0.8121	0.8104	0.6735	0.6147	0.5528	0.48
Host penetration	0.59	0.8121	1	0.8945	0.4608	0.3398	0.51	0.4466
Network reconnaissance	0.5889	0.8104	0.8945	1	0.5187	0.4158	0.3816	0.3432
Incidence response	0.6595	0.6735	0.4608	0.5187	1	0.8223	0.3259	0.5208
Generalized defense practice	0.6141	0.6147	0.3398	0.4158	0.8223	1	0.4426	0.4267
Network protocol reverse engineering	0.3822	0.5528	0.51	0.3816	0.3259	0.4426	1	0.4519
Binary reverse engineering	0.5724	0.48	0.4466	0.3432	0.5208	0.4267	0.4519	1

(2) Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5
Latency	-0.27385	0.26685	0.02464	-0.70171	-0.06017
Frequency	0.08372	0.02202	-0.78817	-0.05056	0.01213
Duration	0.05864	-0.32456	0.10502	0.41038	0.73849
Confusion	0.00037	0.91067	0.01154	0.12850	0.03629
Self-Doubt	-0.08612	0.79091	-0.36000	0.24417	0.04332
Confidence	-0.13184	-0.36740	0.66837	0.36641	-0.13078
Surprise	0.17815	0.69854	0.43597	-0.02194	0.22823
Frustrated	-0.27560	0.62577	0.02529	0.46353	-0.17135
Cyber_security	0.82789	0.14928	-0.16117	0.15300	0.20579
Network_penetration	0.91954	0.05924	0.12297	-0.08984	0.02569
Host_penetration	0.81042	0.16218	0.32630	-0.33395	0.03664
Network_reconnaissance	0.81007	0.15674	0.22801	-0.38618	0.19797
Incidence_response	0.81247	-0.19821	-0.33707	0.02742	-0.05804
Generalized_defense_practice	0.75102	-0.19267	-0.47697	0.18494	-0.16675
Network_protocol_reverse_engineering	-0.27385	0.26685	0.02464	-0.70171	-0.06017
Binary_reverse_engineering	0.08372	0.02202	-0.78817	-0.05056	0.01213

b. Exploratory Factor Analysis

(1) Unrotated Factor Loading

EFA was applied using "Principal Axis" factor method with "Principal Components" for prior communality so the Unrotated Factor Loading is the same as the PCA Loading Matrix above.

(2) Target Matrix

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	0.00295	-1.00000	0.00006	0.00349	-0.07335
Frequency	-0.00388	0.00485	-0.00003	1.00000	-0.00003
Duration	0.00000	0.01153	-0.00269	-0.00323	1.00000
Confusion	0.00352	-0.00174	1.00000	0.00047	-0.00098
Self-Doubt	-0.00156	0.00001	0.80067	0.07288	-0.00002
Confidence	-0.00811	0.01973	-0.01279	-0.83256	0.00084
Surprise	0.15215	-0.01378	0.45845	-0.03288	0.00077
Frustrated	-0.05586	0.00315	0.77401	-0.00300	-0.00234
Cyber_security	0.36917	0.26187	0.00345	0.03460	0.00891
Network_penetration	0.72531	0.11472	-0.00012	0.00018	-0.00020
Host_penetration	0.94372	0.00208	-0.00001	-0.00089	-0.00319
Network_reconnaissance	1.00000	0.00011	-0.00004	0.00005	-0.00000
Incidence_response	0.17396	0.36690	-0.02409	0.09669	-0.00016
Generalized_defense_practice	0.04167	0.60128	-0.01043	0.14155	-0.00120
Network_protocol_reverse_engineering	0.06834	0.37781	0.00008	-0.04626	-0.18047
Binary_reverse_engineering	0.14447	0.82012	0.00104	-0.00186	0.00043

(3) Factor Structure

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	0.054075	-0.668519	0.094023	0.083312	-0.289339
Frequency	-0.087143	0.151872	-0.057375	0.772311	-0.023176
Duration	-0.021789	0.174637	-0.153333	-0.115316	0.861615
Confusion	0.159154	-0.162517	0.902233	0.042942	-0.117896
Self-Doubt	-0.066942	-0.064031	0.806087	0.336741	-0.035905
Confidence	-0.182216	0.137688	-0.192696	-0.783497	0.095453
Surprise	0.448340	-0.190144	0.677315	-0.276718	0.025592
Frustrated	-0.284859	-0.038354	0.747999	-0.155124	-0.096843
Cyber_security	0.695323	0.615026	0.078619	0.344082	0.071983
Network_penetration	0.877884	0.552763	-0.075805	0.122941	-0.164645
Host_penetration	0.937052	0.267824	-0.016461	-0.029183	-0.236118
Network_reconnaissance	0.954886	0.215818	-0.033436	0.092850	-0.104517
Incidence_response	0.569494	0.708651	-0.310864	0.472221	-0.132589
Generalized_defense_practice	0.414250	0.807142	-0.270473	0.542488	-0.173097
Network_protocol_reverse_engineering	0.446792	0.619404	0.009802	-0.266275	-0.564450
Binary_reverse_engineering	0.461677	0.675193	0.024831	-0.035028	-0.027076

(4) Rotation Matrix

0.81046	-0.10053	-0.03843	-0.01829	0.03485
-0.14566	0.92406	0.10885	-0.05028	0.00503
-0.01451	0.07733	0.93472	0.01921	0.01788
-0.03889	-0.04356	0.02642	0.77860	-0.00291
0.07595	0.04820	0.02567	0.00282	0.82594

(5) Final Communality Estimates

Latency	0.64282
Frequency	0.63140
Duration	0.83359
Confusion	0.84729
Self-Doubt	0.82405
Confidence	0.75043
Surprise	0.76234
Frustrated	0.71241
Cyber_security	0.79942
Network_penetration	0.87291
Host_penetration	0.90241
Network_reconnaissance	0.92109
Incidence_response	0.81713
Generalized_defense_practice	0.89066
Network_protocol_reverse_engineering	0.83390
Binary_reverse_engineering	0.57890

(6) Standard Score Coefficients

Latency	0.117579	-0.313090	-0.025470	0.074111	-0.200343
Frequency	-0.061194	0.052282	-0.029241	0.373072	-0.004361
Duration	0.011350	0.022348	0.001080	-0.032253	0.715542
Confusion	0.048358	-0.042442	0.332810	0.008701	-0.012873
Self-Doubt	-0.041667	0.028208	0.302123	0.155967	0.027568
Confidence	-0.070809	0.125401	-0.042550	-0.386922	-0.003848
Surprise	0.169621	-0.130946	0.254966	-0.142621	0.126650
Frustrated	-0.119355	0.113705	0.290386	-0.086830	-0.088325
Cyber_security	0.131056	0.127378	0.061301	0.116707	0.163427
Network_penetration	0.199349	0.073297	-0.013007	0.002124	-0.029984
Host_penetration	0.265203	-0.061195	-0.009872	-0.058051	-0.073570
Network_reconnaissance	0.286897	-0.107145	-0.015569	0.011159	0.050103
Incidence_response	0.072379	0.180861	-0.097259	0.174799	-0.053420
Generalized_defense_practice	-0.002288	0.265557	-0.077182	0.202942	-0.110722
Network_protocol_reverse_engineering	0.018395	0.255766	0.015550	-0.205122	-0.458541
Binary_reverse_engineering	0.039066	0.230571	0.049203	-0.074035	0.016719

(7) Variance Explained by Each Factor

Factor	Variance	Percent	Cum Percent
Factor 1	4.3089	26.931	26.931
Factor 2	3.3829	21.143	48.074
Factor 3	2.7387	17.117	65.190
Factor 4	2.1788	13.618	78.808
Factor 5	1.3268	8.293	87.101

(8) Rotated Factor Loading

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	0.213967	-0.791988	-0.063512	0.163171	-0.325474
Frequency	-0.191353	0.099364	0.016174	0.776705	-0.030496
Duration	0.062949	0.238564	-0.083487	-0.144748	0.886482
Confusion	0.134243	-0.049448	0.893587	0.095728	-0.064054
Self-Doubt	-0.147744	0.084089	0.851202	0.395013	-0.010672
Confidence	-0.167768	0.267808	-0.189337	-0.816401	0.071832
Surprise	0.524707	-0.175363	0.619511	-0.257594	0.111882
Frustrated	-0.372108	0.221061	0.787501	-0.100981	-0.105074
Cyber_security	0.568000	0.474139	0.180462	0.242331	0.212538
Network_penetration	0.784479	0.323000	-0.032562	0.004251	-0.016374
Host_penetration	0.933497	0.003045	-0.048793	-0.119843	-0.095553
Network_reconnaissance	0.981876	-0.072158	-0.067219	0.009120	0.038340
Incidence_response	0.393731	0.508157	-0.201014	0.352386	-0.029456
Generalized_defense_practice	0.176312	0.668067	-0.122751	0.424797	-0.085215
Network_protocol_reverse_engineering	0.243166	0.572903	0.069277	-0.370787	-0.480081
Binary_reverse_engineering	0.302570	0.641438	0.133665	-0.140857	0.078875

2. Absent-Uninformed/Present-Informed

a. Principal Components

(1) Correlations

	Latency	Frequency	Duration	Confusion	Self-Doubt	Confidence	Surprise	Frustrated
Latency	1	-0.1145	-0.3837	-0.1439	-0.3656	-0.1695	-0.0001	-0.247
Frequency	-0.1145	1	0.1481	0.0977	-0.0841	-0.009	-0.0362	0.0118
Duration	-0.3837	0.1481	1	0.2635	0.2081	-0.0645	-0.0585	0.2513
Confusion	-0.1439	0.0977	0.2635	1	0.3639	-0.3184	-0.0088	0.5001
Self-Doubt	-0.3656	-0.0841	0.2081	0.3639	1	-0.1577	0.2225	0.6198
Confidence	-0.1695	-0.009	-0.0645	-0.3184	-0.1577	1	0.2722	-0.3081
Surprise	-0.0001	-0.0362	-0.0585	-0.0088	0.2225	0.2722	1	0.4261
Frustrated	-0.247	0.0118	0.2513	0.5001	0.6198	-0.3081	0.4261	1
Cyber security	-0.0149	0.0354	0.3529	0.0537	-0.1639	-0.0355	-0.2921	-0.0136
Network penetration	-0.2724	-0.0692	0.108	-0.009	-0.0827	0.1243	-0.2264	-0.2097
Host penetration	-0.2891	-0.0056	0.1154	-0.0189	-0.1852	0.0982	-0.3883	-0.3339
Network reconnaissance	-0.1624	-0.159	0.2089	-0.1877	-0.2283	0.167	-0.1814	-0.2235
Incidence response	0.1817	0.0087	0.1225	0.0592	-0.5913	0.0199	0.0208	0.031
Generalized defense practice	0.1329	0.071	0.2108	0.0239	-0.458	-0.0704	-0.2846	-0.0694
Network protocol reverse engineering	0.0487	-0.0355	0.2919	0.0781	0.0094	-0.0054	-0.5144	-0.2965
Binary reverse engineering	-0.0703	-0.1538	0.2132	0.011	-0.0941	0.0536	-0.3562	-0.3039

	Cyber security	Network penetration	Host penetration	Network reconnaissance	Incidence response	Generalized defense practice	Network protocol reverse engineering	Binary reverse engineering
Latency	-0.0149	-0.2724	-0.2891	-0.1624	0.1817	0.1329	0.0487	-0.0703
Frequency	0.0354	-0.0692	-0.0056	-0.159	0.0087	0.071	-0.0355	-0.1538
Duration	0.3529	0.108	0.1154	0.2089	0.1225	0.2108	0.2919	0.2132
Confusion	0.0537	-0.009	-0.0189	-0.1877	0.0592	0.0239	0.0781	0.011
Self-Doubt	-0.1639	-0.0827	-0.1852	-0.2283	-0.5913	-0.458	0.0094	-0.0941
Confidence	-0.0355	0.1243	0.0982	0.167	0.0199	-0.0704	-0.0054	0.0536
Surprise	-0.2921	-0.2264	-0.3883	-0.1814	0.0208	-0.2846	-0.5144	-0.3562
Frustrated	-0.0136	-0.2097	-0.3339	-0.2235	0.031	-0.0694	-0.2965	-0.3039
Cyber security	1	0.6358	0.6032	0.7222	0.4858	0.756	0.5994	0.693
Network penetration	0.6358	1	0.8971	0.7226	0.1487	0.3146	0.5076	0.7681
Host penetration	0.6032	0.8971	1	0.6951	0.1516	0.4189	0.566	0.7784

Network reconnaissance	0.7222	0.7226	0.6951	1	0.4898	0.6641	0.5766	0.6449
Incidence response	0.4858	0.1487	0.1516	0.4898	1	0.7435	0.1808	0.1199
Generalized defense practice	0.756	0.3146	0.4189	0.6641	0.7435	1	0.3037	0.3312
Network protocol reverse engineering	0.5994	0.5076	0.566	0.5766	0.1808	0.3037	1	0.7714
Binary reverse engineering	0.693	0.7681	0.7784	0.6449	0.1199	0.3312	0.7714	1

(2) Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5
Latency	-0.08038	-0.58863	0.29935	-0.41549	0.34153
Frequency	-0.04329	0.07360	0.18881	-0.01652	-0.84798
Duration	0.26056	0.58054	0.21523	0.10334	-0.30442
Confusion	-0.06374	0.64612	0.32836	-0.25619	0.02050
Self-Doubt	-0.34402	0.78965	-0.27212	-0.02683	0.21221
Confidence	0.08160	-0.28163	-0.34806	0.68938	-0.13936
Surprise	-0.46063	0.07853	0.14422	0.67970	0.27481
Frustrated	-0.35554	0.70107	0.43288	0.17324	0.23081
Cyber_security	0.85199	0.17326	0.27317	0.02504	0.05547
Network_penetration	0.80602	0.18170	-0.30968	0.15547	0.08739
Host_penetration	0.84925	0.11088	-0.30521	0.04100	-0.05884
Network_reconnaissance	0.86948	-0.00039	0.03146	0.28441	0.15475
Incidence_response	0.46731	-0.24524	0.73721	0.23474	0.06058
Generalized_defense_practice	0.69481	-0.13160	0.60728	0.05391	-0.03131
Network_protocol_reverse_engineering	0.74593	0.15850	-0.16467	-0.32830	0.03442
Binary_reverse_engineering	0.84115	0.14953	-0.28852	-0.11793	0.15441

b. Exploratory Factor Analysis

(1) Unrotated Factor Loading

EFA was applied using "Principal Axis" factor method with "Principal Components" for prior communality so the Unrotated Factor Loading is the same as the PCA Loading Matrix above.

(2) Target Matrix

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	-0.04147	0.04850	-0.11061	-0.23788	-0.13178
Frequency	-0.00550	0.00061	-0.00002	-0.00025	1.00000
Duration	0.05030	0.00579	0.41345	0.00030	0.21435
Confusion	-0.00000	0.00000	0.88716	-0.09668	0.00334

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Self-Doubt	0.00000	-0.26432	0.52621	0.00028	-0.00185
Confidence	0.00187	-0.00009	-0.10192	1.00000	0.00073
Surprise	-0.13456	0.00024	0.03593	0.59839	-0.02836
Frustrated	-0.03120	0.00021	1.00000	0.00166	-0.00084
Cyber_security	0.47862	0.22263	0.00655	-0.00105	0.00021
Network_penetration	0.97783	0.00055	-0.00000	0.01004	-0.00002
Host_penetration	0.97947	0.00082	-0.00251	0.00036	0.00178
Network_reconnaissance	0.58708	0.12804	-0.00016	0.01919	-0.00094
Incidence_response	0.00062	1.00000	0.00000	0.00014	-0.00000
Generalized_defense_practice	0.05822	0.76241	0.00000	-0.00285	0.00134
Network_protocol_reverse_engineering	0.80885	0.00060	-0.00001	-0.08314	0.00003
Binary_reverse_engineering	1.00000	0.00030	-0.00014	-0.00194	-0.00117

(3) Factor Structure

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	-0.291458	0.303697	-0.434586	-0.510695	-0.458166
Frequency	-0.080731	0.043218	-0.007722	-0.061189	0.820488
Duration	0.312559	0.141683	0.515324	0.079038	0.480578
Confusion	-0.005951	-0.004369	0.667801	-0.249278	0.130599
Self-Doubt	-0.064683	-0.607155	0.727186	0.148354	-0.100101
Confidence	0.110765	-0.036046	-0.303167	0.703899	0.060273
Surprise	-0.450013	-0.021828	0.334254	0.682101	-0.346903
Frustrated	-0.293388	-0.007276	0.894007	0.179995	-0.121964
Cyber_security	0.757167	0.628612	0.107201	-0.099979	0.170321
Network_penetration	0.880530	0.175470	-0.031272	0.159425	0.111382
Host_penetration	0.902708	0.192457	-0.141639	0.034109	0.245511
Network_reconnaissance	0.803060	0.551270	-0.075458	0.186465	0.027043
Incidence_response	0.173571	0.934776	-0.036940	-0.002859	0.001243
Generalized_defense_practice	0.447279	0.887243	-0.052129	-0.161563	0.160834
Network_protocol_reverse_engineering	0.777428	0.174368	-0.074119	-0.336964	0.158203
Binary_reverse_engineering	0.899927	0.176710	-0.087549	-0.114601	0.053583

(4) Rotation Matrix

0.93124	-0.08250	0.00336	-0.01297	-0.06316
-0.11555	0.78579	0.02302	0.02432	0.02076
0.02867	0.02762	0.89945	-0.07236	-0.01744
-0.00244	0.03552	-0.08211	0.85390	0.01017
-0.15101	0.01344	-0.01736	0.03033	0.82345

(5) Final Communality Estimates

Latency	0.73183
Frequency	0.76227
Duration	0.55459
Confusion	0.59541

Self-Doubt	0.86170
Confidence	0.70178
Surprise	0.77665
Frustrated	0.88858
Cyber_security	0.83423
Network_penetration	0.81039
Host_penetration	0.83181
Network_reconnaissance	0.86182
Incidence_response	0.88077
Generalized_defense_practice	0.87276
Network_protocol_reverse_engineering	0.71761
Binary_reverse_engineering	0.85089

(6) Standard Score Coefficients

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	-0.108978	0.146276	-0.157344	-0.316456	-0.319416
Frequency	-0.031120	0.012534	-0.059730	-0.040798	0.685913
Duration	0.065915	0.040763	0.204223	0.055865	0.308313
Confusion	0.000136	0.021923	0.266848	-0.174186	0.048742
Self-Doubt	0.051065	-0.220839	0.282664	0.053387	-0.114796
Confidence	0.038682	-0.025213	-0.113829	0.466660	0.074177
Surprise	-0.091841	0.095384	0.166202	0.435931	-0.236685
Frustrated	-0.059316	0.088814	0.380889	0.100521	-0.129434
Cyber_security	0.125051	0.179257	0.087812	-0.022008	0.008236
Network_penetration	0.201307	-0.038306	0.018735	0.128140	-0.029359
Host_penetration	0.201113	-0.045641	-0.038281	0.048371	0.084261
Network_reconnaissance	0.147598	0.140953	0.024337	0.165761	-0.093551
Incidence_response	-0.045637	0.394125	0.039798	0.051683	-0.044037
Generalized_defense_practice	0.022193	0.327248	0.024448	-0.051031	0.048757
Network_protocol_reverse_engineering	0.167335	-0.043568	-0.017513	-0.199207	0.017213
Binary_reverse_engineering	0.201457	-0.050704	-0.007087	-0.050624	-0.080582

(7) Variance Explained by Each Factor

Factor	Variance	Percent	Cum Percent
Factor 1	4.9471	30.919	30.919
Factor 2	2.9742	18.589	49.508
Factor 3	2.4876	15.547	65.055
Factor 4	1.5722	9.826	74.882
Factor 5	1.4362	8.976	83.858

(8) Rotated Factor Loading

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	-0.280926	0.306332	-0.332294	-0.444965	-0.404371
Frequency	-0.310360	0.103963	-0.032625	-0.012368	0.891805
Duration	0.206186	0.135670	0.520147	0.015588	0.419590

Confusion	-0.007125	0.023594	0.733098	-0.373542	0.097950
Self-Doubt	0.123997	-0.582166	0.695468	-0.026922	-0.145015
Confidence	0.083738	-0.026515	-0.443345	0.784275	0.084443
Surprise	-0.415942	0.139595	0.220921	0.642701	-0.231440
Frustrated	-0.266850	0.123119	0.890546	0.027173	-0.083105
Cyber_security	0.653362	0.491541	0.186300	-0.090261	0.009030
Network_penetration	0.903699	-0.012380	-0.023936	0.163337	-0.089409
Host_penetration	0.889589	-0.008552	-0.119876	0.060791	0.046280
Network_reconnaissance	0.744758	0.401546	-0.052125	0.225848	-0.135853
Incidence_response	-0.036658	0.950127	0.022695	0.069235	0.007375
Generalized_defense_practice	0.243302	0.828134	0.036038	-0.096672	0.096002
Network_protocol_reverse_engineering	0.790911	-0.025989	0.015720	-0.339626	-0.037988
Binary_reverse_engineering	0.946409	-0.042894	-0.029465	-0.115138	-0.168052

3. Present-Informed/Absent-Informed

a. Principal Components

(1) Correlations

	Latency	Frequency	Duration	Confusion	Self-Doubt	Confidence	Surprise	Frustrated
Latency	1	-0.2589	-0.3468	-0.1681	-0.0977	-0.0931	-0.3556	-0.23
Frequency	-0.2589	1	0.3362	0.3348	0.264	0.0404	0.2587	0.1847
Duration	-0.3468	0.3362	1	0.1971	0.1012	-0.0014	0.2926	0.1768
Confusion	-0.1681	0.3348	0.1971	1	0.5926	0.1042	0.7739	0.5214
Self-Doubt	-0.0977	0.264	0.1012	0.5926	1	-0.3835	0.5227	0.5785
Confidence	-0.0931	0.0404	-0.0014	0.1042	-0.3835	1	0.0085	-0.5208
Surprise	-0.3556	0.2587	0.2926	0.7739	0.5227	0.0085	1	0.4301
Frustrated	-0.23	0.1847	0.1768	0.5214	0.5785	-0.5208	0.4301	1
Cyber security	0.194	0.1708	0.0148	0.437	0.0586	0.4611	0.413	-0.3582
Network penetration	0.4402	-0.1519	-0.5838	-0.1273	-0.3881	0.3429	-0.2347	-0.5452
Host penetration	0.4585	-0.2592	-0.5543	-0.1762	-0.5026	0.295	-0.1537	-0.4564
Network reconnaissance	0.4133	-0.0804	-0.3963	-0.1082	-0.2673	0.2231	-0.1255	-0.6136
Incidence response	0.0513	0.2263	0.0491	0.5028	0.0025	0.3331	0.2717	-0.0699
Generalized defense practice	0.171	-0.159	-0.0836	0.1501	-0.1289	0.2343	-0.153	-0.1473
Network protocol reverse engineering	0.4244	-0.2954	-0.4502	-0.0667	-0.4034	0.122	-0.116	-0.2178
Binary reverse engineering	0.3161	0.0113	-0.2377	0.301	-0.0988	0.0428	0.2797	0.0809

	Cyber security	Network penetration	Host penetration	Network reconnaissance	Incidence response	Generalized defense practice	Network protocol reverse engineering	Binary reverse engineering
Latency	0.194	0.4402	0.4585	0.4133	0.0513	0.171	0.4244	0.3161
Frequency	0.1708	-0.1519	-0.2592	-0.0804	0.2263	-0.159	-0.2954	0.0113
Duration	0.0148	-0.5838	-0.5543	-0.3963	0.0491	-0.0836	-0.4502	-0.2377
Confusion	0.437	-0.1273	-0.1762	-0.1082	0.5028	0.1501	-0.0667	0.301
Self-Doubt	0.0586	-0.3881	-0.5026	-0.2673	0.0025	-0.1289	-0.4034	-0.0988
Confidence	0.4611	0.3429	0.295	0.2231	0.3331	0.2343	0.122	0.0428
Surprise	0.413	-0.2347	-0.1537	-0.1255	0.2717	-0.153	-0.116	0.2797
Frustrated	-0.3582	-0.5452	-0.4564	-0.6136	-0.0699	-0.1473	-0.2178	0.0809
Cyber security	1	0.4662	0.3854	0.5821	0.6635	0.3929	0.2246	0.3339
Network penetration	0.4662	1	0.8994	0.7791	0.2696	0.3001	0.772	0.4956
Host penetration	0.3854	0.8994	1	0.6209	0.1351	0.193	0.7828	0.6212
Network reconnaissance	0.5821	0.7791	0.6209	1	0.4173	0.4258	0.5091	0.3664
Incidence response	0.6635	0.2696	0.1351	0.4173	1	0.682	0.2517	0.3594
Generalized defense practice	0.3929	0.3001	0.193	0.4258	0.682	1	0.237	0.1043
Network protocol reverse engineering	0.2246	0.772	0.7828	0.5091	0.2517	0.237	1	0.6518
Binary reverse engineering	0.3339	0.4956	0.6212	0.3664	0.3594	0.1043	0.6518	1

(2) Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5
Latency	0.53995	-0.14302	0.33642	0.33560	0.32686
Frequency	-0.28528	0.43980	-0.20249	-0.27900	0.37025
Duration	-0.54409	0.23723	-0.42093	-0.09637	-0.15108
Confusion	-0.20455	0.89862	0.16939	0.00694	-0.06611
Self-Doubt	-0.53820	0.47388	0.34771	0.23835	0.41152
Confidence	0.40219	0.20285	-0.62690	-0.32239	-0.17157
Surprise	-0.29015	0.78330	0.17954	-0.31060	-0.02792
Frustrated	-0.63670	0.30773	0.56554	0.14029	-0.26189
Cyber_security	0.50319	0.68825	-0.27522	-0.02957	0.25642
Network_penetration	0.93197	0.03122	0.09732	-0.11980	0.12117
Host_penetration	0.89551	-0.03584	0.22550	-0.25970	-0.04803
Network_reconnaissance	0.81804	0.14931	-0.10013	0.11897	0.32845
Incidence_response	0.37248	0.72691	-0.26435	0.31606	-0.17472
Generalized_defense_practice	0.43617	0.30882	-0.28627	0.68768	-0.28990
Network_protocol_reverse_engineering	0.78397	0.03019	0.38332	-0.10299	-0.27992
Binary_reverse_engineering	0.52609	0.41373	0.50986	-0.24677	-0.20455

b. Exploratory Factor Analysis

(1) Unrotated Factor Loading

EFA was applied using "Principal Axis" factor method with "Principal Components" for prior communality so the Unrotated Factor Loading is the same as the PCA Loading Matrix above.

(2) Target Matrix

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	0.04133	-0.01558	0.00253	1.00000	0.00229
Frequency	-0.06745	0.96549	-0.00974	-0.00074	-0.00652
Duration	-0.25858	0.05132	-0.00092	-0.50971	0.00084
Confusion	0.00279	0.80893	0.03848	-0.01728	0.05850
Self-Doubt	-0.05194	0.37173	0.38934	0.00488	-0.00000
Confidence	0.00339	0.00293	-1.00000	-0.01763	0.02353
Surprise	0.00561	1.00000	0.01188	-0.06452	0.00009
Frustrated	0.00000	0.02684	0.88999	-0.08636	-0.00000
Cyber_security	0.00872	0.36316	-0.14298	0.04244	0.12136
Network_penetration	0.37210	-0.00016	-0.12971	0.24001	0.00279
Host_penetration	0.68818	-0.00222	-0.07908	0.08134	0.00000
Network_reconnaissance	0.05052	0.00064	-0.15003	0.52262	0.04080
Incidence_response	0.00645	0.09354	-0.00977	0.00002	0.73296
Generalized_defense_practice	0.00013	-0.00540	-0.00080	0.00459	1.00000
Network_protocol_reverse_engineering	0.98459	-0.00824	-0.00228	0.02037	0.00529
Binary_reverse_engineering	1.00000	0.04430	0.00045	0.00120	0.00173

(3) Factor Structure

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	0.322340	-0.183577	-0.081387	0.744156	0.133868
Frequency	-0.280379	0.600164	-0.038942	-0.145077	-0.066990
Duration	-0.470277	0.270469	0.090239	-0.597702	0.021243
Confusion	0.134423	0.850214	0.292754	-0.285117	0.445505
Self-Doubt	-0.346457	0.579883	0.650313	-0.091021	-0.011889
Confidence	0.186651	0.148143	-0.761318	-0.013092	0.317118
Surprise	0.118709	0.826781	0.252890	-0.395301	0.150146
Frustrated	-0.079952	0.281962	0.883787	-0.546489	-0.053800
Cyber_security	0.294989	0.646255	-0.530520	0.380869	0.596101
Network_penetration	0.723705	-0.040006	-0.620993	0.720417	0.269596
Host_penetration	0.840396	-0.113524	-0.542589	0.578394	0.156683
Network_reconnaissance	0.435439	0.091357	-0.612846	0.788703	0.412778
Incidence_response	0.282203	0.518419	-0.288073	0.117962	0.888557
Generalized_defense_practice	0.182141	0.015224	-0.237580	0.251438	0.907275
Network_protocol_reverse_engineering	0.891097	-0.127304	-0.291821	0.412824	0.270694
Binary_reverse_engineering	0.847554	0.315291	-0.068800	0.197071	0.264665

(4) Rotation Matrix

0.97439	0.00067	0.03914	-0.06006	-0.10673
-0.00832	0.97024	-0.02053	0.04355	-0.17322
0.09784	-0.05627	0.85011	0.12728	0.10815
-0.14304	0.09182	0.15595	0.82719	-0.05404
-0.13031	-0.15343	0.05830	-0.02236	0.92076

(5) Final Communality Estimates

Latency	0.64465
Frequency	0.53074
Duration	0.56161
Confusion	0.88247
Self-Doubt	0.86129
Confidence	0.72928
Surprise	0.82723
Frustrated	0.90819
Cyber_security	0.86926
Network_penetration	0.90804
Host_penetration	0.92382
Network_reconnaissance	0.82354
Incidence_response	0.86743
Generalized_defense_practice	0.92451
Network_protocol_reverse_engineering	0.85141
Binary_reverse_engineering	0.81064

(6) Standard Score Coefficients

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**
Latency	-0.047259	-0.016883	0.099008	0.381886	0.019013
Frequency	-0.143347	0.275171	-0.132274	0.099568	-0.160655
Duration	-0.084089	0.052096	-0.082019	-0.226769	0.061428
Confusion	0.091938	0.237709	0.100974	-0.091826	0.150612
Self-Doubt	-0.164060	0.220948	0.212601	0.246360	-0.021833
Confidence	0.045602	0.049223	-0.325770	-0.179070	0.055724
Surprise	0.127040	0.269532	0.030355	-0.146486	-0.034865
Frustrated	0.140662	0.010147	0.334262	-0.190608	0.083315
Cyber_security	-0.048229	0.253309	-0.179606	0.174239	0.102535
Network_penetration	0.129945	0.038672	-0.124828	0.176942	-0.030915
Host_penetration	0.245078	-0.002051	-0.097232	0.045195	-0.071941
Network_reconnaissance	-0.054019	0.091400	-0.135082	0.328653	0.049986
Incidence_response	0.030907	0.093740	-0.032303	-0.035497	0.393792
Generalized_defense_practice	-0.024515	-0.116376	0.051784	0.001080	0.531454
Network_protocol_reverse_engineering	0.327833	-0.068777	0.031789	-0.071296	0.056972
Binary_reverse_engineering	0.348425	0.089187	0.068124	-0.097865	0.001604

(7) Variance Explained by Each Factor

Factor	Variance	Percent	Cum Percent
Factor 1	3.7290	23.306	23.306
Factor 2	3.1363	19.602	42.908
Factor 3	3.5159	21.975	64.882
Factor 4	3.3838	21.149	86.031
Factor 5	2.7264	17.040	103.071

(8) Rotated Factor Loading

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	0.15894	-0.13963	0.27830	0.77865	0.08220
Frequency	-0.28499	0.70302	-0.20667	-0.02616	-0.25479
Duration	-0.40553	0.20069	-0.20381	-0.55673	0.10619
Confusion	0.13774	0.72602	0.28277	-0.16704	0.26098
Self-Doubt	-0.30154	0.60645	0.66754	0.32100	-0.02178
Confidence	0.04040	0.10317	-0.86567	-0.36225	0.13453
Surprise	0.21728	0.80601	0.12926	-0.29588	-0.09731
Frustrated	0.15783	0.18173	0.82581	-0.26201	0.06435
Cyber_security	0.02797	0.60853	-0.39793	0.25393	0.26826
Network_penetration	0.52772	0.01048	-0.32862	0.45505	-0.01984
Host_penetration	0.73678	-0.05912	-0.29041	0.27912	-0.13466
Network_reconnaissance	0.13835	0.11601	-0.30998	0.62027	0.17812
Incidence_response	0.02945	0.27574	-0.13365	-0.02018	0.76535
Generalized_defense_practice	-0.11318	-0.29545	0.01607	0.10457	1.02191
Network_protocol_reverse_engineering	0.83651	-0.17155	-0.01612	0.14842	0.06719
Binary_reverse_engineering	0.86256	0.28802	0.12787	0.05155	-0.04761

4. Present-Uninformed/Absent-Uninformed

a. Principal Components

(1) Correlations

	Latency	Frequency	Duration	Confusion	Self-Doubt	Confidence	Surprise	Frustrated
Latency	1	-0.2338	0.0438	-0.0468	-0.0881	-0.12	0.1084	0.0991
Frequency	-0.2338	1	0.2939	0.4228	0.5377	-0.3673	0.4403	0.4784
Duration	0.0438	0.2939	1	-0.1638	-0.148	0.293	0.2416	-0.0761
Confusion	-0.0468	0.4228	-0.1638	1	0.6712	-0.7013	0.2335	0.5003
Self-Doubt	-0.0881	0.5377	-0.148	0.6712	1	-0.5298	0.4849	0.5998
Confidence	-0.12	-0.3673	0.293	-0.7013	-0.5298	1	-0.1697	-0.4872
Surprise	0.1084	0.4403	0.2416	0.2335	0.4849	-0.1697	1	0.4839
Frustrated	0.0991	0.4784	-0.0761	0.5003	0.5998	-0.4872	0.4839	1
Cyber security	-0.0233	-0.0036	0.148	-0.001	-0.1385	0.1087	-0.2617	-0.0395
Network penetration	0.0046	-0.0517	0.2009	-0.1319	-0.3865	0.266	-0.2039	-0.1846

Host penetration	0.0241	-0.0844	0.2183	-0.3233	-0.495	0.3636	-0.1436	-0.2241
Network reconnaissance	0.0726	-0.152	0.0972	-0.1194	-0.2895	0.3654	-0.1338	-0.2051
Incidence response	0.2719	0.2394	-0.0592	-0.0252	0.1982	0.0662	0.3982	-0.0498
Generalized defense practice	0.043	0.0486	-0.3628	0.1696	0.2929	-0.0071	0.0917	-0.0022
Network protocol reverse engineering	0.2216	-0.1421	0.0753	-0.2033	-0.2119	0.4416	0.1847	0.1796
Binary reverse engineering	-0.0578	-0.1894	0.212	-0.2751	-0.3217	0.5462	-0.0733	0.0724

	Cyber security	Network penetration	Host penetration	Network reconnaissance	Incidence response	Generalized defense practice	Network protocol reverse engineering	Binary reverse engineering
Latency	-0.0233	0.0046	0.0241	0.0726	0.2719	0.043	0.2216	-0.0578
Frequency	-0.0036	-0.0517	-0.0844	-0.152	0.2394	0.0486	-0.1421	-0.1894
Duration	0.148	0.2009	0.2183	0.0972	-0.0592	-0.3628	0.0753	0.212
Confusion	-0.001	-0.1319	-0.3233	-0.1194	-0.0252	0.1696	-0.2033	-0.2751
Self-Doubt	-0.1385	-0.3865	-0.495	-0.2895	0.1982	0.2929	-0.2119	-0.3217
Confidence	0.1087	0.266	0.3636	0.3654	0.0662	-0.0071	0.4416	0.5462
Surprise	-0.2617	-0.2039	-0.1436	-0.1338	0.3982	0.0917	0.1847	-0.0733
Frustrated	-0.0395	-0.1846	-0.2241	-0.2051	-0.0498	-0.0022	0.1796	0.0724
Cyber security	1	0.4404	0.5197	0.3032	0.1199	0.4794	0.2501	0.4719
Network penetration	0.4404	1	0.928	0.7708	0.0996	0.202	0.4423	0.5121
Host penetration	0.5197	0.928	1	0.7208	0.2002	0.223	0.525	0.6073
Network reconnaissance	0.3032	0.7708	0.7208	1	0.2151	0.3331	0.6372	0.6064
Incidence response	0.1199	0.0996	0.2002	0.2151	1	0.6557	0.241	-0.0248
Generalized defense practice	0.4794	0.202	0.223	0.3331	0.6557	1	0.2399	0.1248
Network protocol reverse engineering	0.2501	0.4423	0.525	0.6372	0.241	0.2399	1	0.8093
Binary reverse engineering	0.4719	0.5121	0.6073	0.6064	-0.0248	0.1248	0.8093	1

(2) Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5
Latency	0.05927	0.12064	-0.11295	0.53789	-0.35462
Frequency	-0.40095	0.55243	0.38271	-0.19947	0.42560
Duration	0.21647	0.01521	0.75432	0.02752	0.40008
Confusion	-0.56160	0.50473	-0.08873	-0.40858	-0.15780
Self-Doubt	-0.68688	0.57228	-0.02878	-0.04418	0.00831
Confidence	0.69993	-0.28044	0.17957	0.31640	0.16113

Surprise	-0.32503	0.55785	0.41304	0.47204	0.08621
Frustrated	-0.43627	0.58152	0.32908	-0.09820	-0.47237
Cyber_security	0.48843	0.34060	-0.20089	-0.40889	0.12022
Network_penetration	0.76895	0.27740	0.02274	-0.32813	0.08452
Host_penetration	0.85206	0.26033	0.04073	-0.18734	0.12385
Network_reconnaissance	0.76827	0.33872	-0.06310	-0.08054	-0.07277
Incidence_response	0.11423	0.57615	-0.32934	0.55382	0.40471
Generalized_defense_practice	0.18489	0.62397	-0.65519	0.08683	0.17270
Network_protocol_reverse_engineering	0.65360	0.43536	0.19085	0.26041	-0.42152
Binary_reverse_engineering	0.75973	0.25199	0.27354	-0.08487	-0.32206

b. Exploratory Factor Analysis

(1) Unrotated Factor Loading

EFA was applied using "Principal Axis" factor method with "Principal Components" for prior communality so the Unrotated Factor Loading is the same as the PCA Loading Matrix above.

(2) Target Matrix

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**
Latency	-0.00086	-0.00079	0.01827	-0.00699	1.00000
Frequency	-0.00010	0.24816	0.01106	0.42450	-0.04068
Duration	0.00794	-0.01581	-0.01255	1.00000	-0.00275
Confusion	-0.00065	1.00000	0.00020	-0.00110	-0.00646
Self-Doubt	-0.04018	0.72672	0.03499	0.00530	0.00000
Confidence	0.05956	-0.73121	-0.00001	0.01504	0.00776
Surprise	-0.00892	0.06971	0.03591	0.44720	0.11552
Frustrated	0.00000	0.74174	-0.00757	0.01120	0.06545
Cyber_security	0.75023	0.00052	0.01989	-0.00538	-0.05552
Network_penetration	1.00000	-0.00583	0.00037	0.00008	-0.00548
Host_penetration	0.90604	-0.03021	0.00267	0.00125	-0.00042
Network_reconnaissance	0.94171	-0.00809	0.00805	-0.00015	0.00456
Incidence_response	0.00040	-0.00003	1.00000	0.00915	0.01436
Generalized_defense_practice	0.04586	0.01007	0.75143	-0.03398	0.00000
Network_protocol_reverse_engineering	0.41902	-0.00022	0.00004	0.00165	0.37177
Binary_reverse_engineering	0.79063	-0.00325	-0.01216	0.00140	0.05020

(3) Factor Structure

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**
Latency	-0.027373	-0.041809	0.136055	-0.055940	0.590216
Frequency	-0.130106	0.603787	0.280219	0.640411	-0.211761
Duration	0.192484	-0.217863	-0.222271	0.769635	0.032588
Confusion	-0.207276	0.853136	0.182891	-0.025811	-0.179615

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Self-Doubt	-0.405331	0.848537	0.385083	0.206604	-0.016658
Confidence	0.437315	-0.777518	-0.126303	0.148666	0.222145
Surprise	-0.213572	0.438003	0.313212	0.699746	0.473988
Frustrated	-0.108697	0.799055	-0.055916	0.293641	0.378646
Cyber_security	0.646412	-0.028294	0.234155	-0.129998	-0.233342
Network_penetration	0.866101	-0.273502	0.069320	0.021729	-0.061297
Host_penetration	0.893894	-0.388107	0.107317	0.075341	0.027100
Network_reconnaissance	0.828531	-0.271483	0.158038	-0.025917	0.193980
Incidence_response	0.104023	0.053675	0.892510	0.212661	0.216452
Generalized_defense_practice	0.314518	0.208513	0.848810	-0.244900	-0.031346
Network_protocol_reverse_engineering	0.706760	-0.152039	0.024534	0.174449	0.710710
Binary_reverse_engineering	0.827272	-0.266072	-0.198608	0.125169	0.409074

(4) Rotation Matrix

1.01446	0.14854	-0.03736	-0.00302	-0.02130
0.16143	0.92961	-0.08401	-0.04034	0.00016
-0.06353	-0.15154	0.83922	0.00871	0.01980
-0.02109	-0.08080	-0.00438	0.82229	-0.06768
-0.09337	0.00017	0.03596	-0.12649	0.65906

(5) Final Communality Estimates

Latency	0.44591
Frequency	0.83333
Duration	0.77691
Confusion	0.76986
Self-Doubt	0.80215
Confidence	0.72686
Surprise	0.81770
Frustrated	0.86956
Cyber_security	0.57658
Network_penetration	0.78356
Host_penetration	0.84588
Network_reconnaissance	0.72073
Incidence_response	0.92397
Generalized_defense_practice	0.89017
Network_protocol_reverse_engineering	0.89865
Binary_reverse_engineering	0.82644

(6) Standard Score Coefficients

Latency	-0.050398	-0.023336	0.038230	-0.059619	0.401313
Frequency	0.010070	0.145086	0.111978	0.344213	-0.194434
Duration	0.029460	-0.090115	-0.085978	0.478115	-0.046636
Confusion	0.037843	0.284356	-0.002514	-0.089183	-0.098312
Self-Doubt	-0.050119	0.224380	0.127123	0.056384	-0.002912

Confidence	0.031403	-0.238222	0.001980	0.142632	0.094482
Surprise	-0.071602	0.069019	0.112976	0.364355	0.272763
Frustrated	0.045348	0.280660	-0.173304	0.070499	0.265837
Cyber_security	0.186432	0.059700	0.093156	-0.090085	-0.224087
Network_penetration	0.223161	0.000139	0.010045	0.006776	-0.131142
Host_penetration	0.209197	-0.046497	0.042761	0.046613	-0.082502
Network_reconnaissance	0.194373	-0.006685	0.041975	-0.033815	0.044898
Incidence_response	-0.044539	-0.057804	0.482385	0.120566	0.058239
Generalized_defense_practice	0.062533	0.061809	0.416884	-0.172169	-0.086245
Network_protocol_reverse_engineering	0.153750	0.025713	-0.077299	0.048558	0.405843
Binary_reverse_engineering	0.208104	0.018500	-0.175053	0.039298	0.206071

(7) Variance Explained by Each Factor

Factor	Variance	Percent	Cum Percent
Factor 1	4.4578	27.861	27.861
Factor 2	3.7367	23.354	51.215
Factor 3	2.0986	13.116	64.332
Factor 4	1.8219	11.387	75.719
Factor 5	1.6610	10.381	86.100

(8) Rotated Factor Loading

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	-0.143058	-0.092450	0.210227	-0.214944	0.673479
Frequency	0.051883	0.496050	0.127387	0.667107	-0.366682
Duration	0.148053	-0.241196	-0.196199	0.856737	-0.218590
Confusion	0.078649	0.896277	-0.041233	-0.103224	-0.146945
Self-Doubt	-0.201139	0.722815	0.219697	0.110594	0.009956
Confidence	0.193225	-0.756570	0.052446	0.209402	0.133271
Surprise	-0.201180	0.247402	0.278769	0.570939	0.378789
Frustrated	0.121423	0.889567	-0.250671	0.096829	0.341999
Cyber_security	0.722260	0.153994	0.153788	-0.085395	-0.286901
Network_penetration	0.876903	-0.033515	0.032683	0.064654	-0.183364
Host_penetration	0.848081	-0.180111	0.111193	0.115754	-0.102248
Network_reconnaissance	0.789715	-0.064153	0.153356	-0.056233	0.120178
Incidence_response	-0.018084	-0.194197	0.952502	0.165588	0.235351
Generalized_defense_practice	0.325211	0.151126	0.808908	-0.292356	0.058471
Network_protocol_reverse_engineering	0.640721	0.043074	0.032613	0.004200	0.634335
Binary_reverse_engineering	0.807090	0.024450	-0.215585	0.047050	0.284937

5. Combined Day 2

a. Principal Components

(1) Correlations

	Latency	Frequency	Duration	Confusion	Self-Doubt	Confidence	Surprise	Frustrated
Latency	1	-0.1645	-0.2364	-0.0211	-0.0908	-0.1706	-0.0392	-0.0523
Frequency	-0.1645	1	0.1278	0.1739	0.1771	-0.1076	0.1076	0.1196
Duration	-0.2364	0.1278	1	-0.0499	0.0224	0.1174	0.1314	0.0692
Confusion	-0.0211	0.1739	-0.0499	1	0.5442	-0.2356	0.3661	0.4823
Self-Doubt	-0.0908	0.1771	0.0224	0.5442	1	-0.3652	0.2967	0.5567
Confidence	-0.1706	-0.1076	0.1174	-0.2356	-0.3652	1	0.0664	-0.3485
Surprise	-0.0392	0.1076	0.1314	0.3661	0.2967	0.0664	1	0.4054
Frustrated	-0.0523	0.1196	0.0692	0.4823	0.5567	-0.3485	0.4054	1
Cyber security	-0.0368	0.0917	0.1185	0.1074	-0.0641	0.0649	0.0372	-0.1527
Network penetration	-0.0351	-0.0911	-0.0666	-0.0801	-0.2298	0.1537	-0.1254	-0.3293
Host penetration	0.0523	-0.1541	-0.0591	-0.1794	-0.3205	0.1775	-0.105	-0.3154
Network reconnaissance	0.0663	-0.0761	-0.0223	-0.098	-0.2348	0.1598	-0.0148	-0.3519
Incidence response	0.0988	0.164	0.0049	0.1109	-0.1661	0.0866	0.1242	-0.1177
Generalized defense practice	0.0564	0.0184	-0.0574	0.042	-0.1117	-0.0217	-0.1501	-0.0829
Network protocol reverse engineering	0.0949	-0.1833	-0.0244	-0.0611	-0.1269	0.1769	-0.0918	-0.0724
Binary reverse engineering	0.0278	-0.0508	0.0646	-0.0218	-0.132	0.1833	0.035	-0.0222

	Cyber security	Network penetration	Host penetration	Network reconnaissance	Incidence response	Generalized defense practice	Network protocol reverse engineering	Binary reverse engineering
Latency	-0.0368	-0.0351	0.0523	0.0663	0.0988	0.0564	0.0949	0.0278
Frequency	0.0917	-0.0911	-0.1541	-0.0761	0.164	0.0184	-0.1833	-0.0508
Duration	0.1185	-0.0666	-0.0591	-0.0223	0.0049	-0.0574	-0.0244	0.0646
Confusion	0.1074	-0.0801	-0.1794	-0.098	0.1109	0.042	-0.0611	-0.0218
Self-Doubt	-0.0641	-0.2298	-0.3205	-0.2348	-0.1661	-0.1117	-0.1269	-0.132
Confidence	0.0649	0.1537	0.1775	0.1598	0.0866	-0.0217	0.1769	0.1833
Surprise	0.0372	-0.1254	-0.105	-0.0148	0.1242	-0.1501	-0.0918	0.035
Frustrated	-0.1527	-0.3293	-0.3154	-0.3519	-0.1177	-0.0829	-0.0724	-0.0222
Cyber security	1	0.5742	0.5436	0.5769	0.539	0.5791	0.373	0.4963
Network penetration	0.5742	1	0.8886	0.777	0.3465	0.3781	0.5657	0.5637
Host penetration	0.5436	0.8886	1	0.7359	0.2991	0.3404	0.6139	0.631

Network reconnaissance	0.5769	0.777	0.7359	1	0.4764	0.5019	0.5498	0.5086
Incidence response	0.539	0.3465	0.2991	0.4764	1	0.7331	0.2794	0.2618
Generalized defense practice	0.5791	0.3781	0.3404	0.5019	0.7331	1	0.3063	0.2409
Network protocol reverse engineering	0.373	0.5657	0.6139	0.5498	0.2794	0.3063	1	0.7209
Binary reverse engineering	0.4963	0.5637	0.631	0.5086	0.2618	0.2409	0.7209	1

(2) Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5
Latency	0.06906	-0.06814	-0.67028	0.08278	0.42078
Frequency	-0.12464	0.35580	0.26302	-0.45531	-0.28378
Duration	-0.02441	0.10366	0.68142	-0.09429	-0.05662
Confusion	-0.19554	0.76058	-0.09825	0.08218	0.06058
Self-Doubt	-0.40340	0.67173	-0.06273	0.21115	-0.21762
Confidence	0.27714	-0.35734	0.54540	0.02346	0.46878
Surprise	-0.15468	0.54295	0.30888	0.17248	0.57339
Frustrated	-0.42020	0.66873	-0.02749	0.28317	0.01740
Cyber_security	0.71183	0.36832	0.08485	-0.21571	-0.07674
Network_penetration	0.86268	0.03967	0.01155	0.16106	-0.20824
Host_penetration	0.87137	-0.03586	0.00012	0.24906	-0.12484
Network_reconnaissance	0.85668	0.08711	-0.02193	-0.00009	-0.00116
Incidence_response	0.57489	0.35931	-0.08569	-0.54591	0.30664
Generalized_defense_practice	0.60153	0.29396	-0.24498	-0.50797	0.02667
Network_protocol_reverse_engineering	0.71115	0.10395	-0.02752	0.42682	-0.01091
Binary_reverse_engineering	0.70003	0.19743	0.14396	0.41310	-0.02312

b. Exploratory Factor Analysis

(1) Unrotated Factor Loading

EFA was applied using "Principal Axis" factor method with "Principal Components" for prior communality so the Unrotated Factor Loading is the same as the PCA Loading Matrix above.

(2) Target Matrix

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	-0.00005	0.00000	0.00459	1.00000	0.00015
Frequency	-0.03172	0.02965	0.12102	-0.41525	-0.01992
Duration	0.00000	-0.00000	-0.00003	-0.79690	0.12553
Confusion	-0.00010	1.00000	0.01146	0.00000	0.00171

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Self-Doubt	-0.00205	0.96940	-0.00266	-0.00272	-0.00764
Confidence	0.00597	-0.22176	-0.00008	-0.00856	0.83762
Surprise	-0.00015	0.21270	0.00045	-0.00013	1.00000
Frustrated	-0.00317	0.99554	-0.00508	-0.00002	0.00262
Cyber_security	0.29954	0.00009	0.42606	-0.01070	-0.00000
Network_penetration	0.85259	-0.00896	0.02267	-0.00006	-0.00452
Host_penetration	0.89775	-0.01756	0.00719	0.00013	-0.00027
Network_reconnaissance	0.62236	-0.01193	0.13836	0.00013	0.00002
Incidence_response	0.00907	-0.00001	1.00000	0.00021	0.01334
Generalized_defense_practice	0.02756	-0.00011	0.93445	0.00078	-0.00741
Network_protocol_reverse_engineering	1.00000	-0.00000	0.00013	0.00450	0.00117
Binary_reverse_engineering	0.99249	0.00028	0.00027	-0.00005	0.01076

(3) Factor Structure

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	0.011871	-0.021009	0.103437	0.785728	-0.020676
Frequency	-0.207461	0.256008	0.291019	-0.542485	-0.138787
Duration	-0.008946	0.010644	-0.013162	-0.629704	0.321863
Confusion	-0.109052	0.769394	0.167453	-0.018239	0.090525
Self-Doubt	-0.232468	0.820069	-0.118149	-0.140740	-0.105436
Confidence	0.218006	-0.508745	-0.014733	-0.137466	0.650819
Surprise	-0.087949	0.512082	0.046123	-0.043427	0.720062
Frustrated	-0.245494	0.824164	-0.153436	-0.039874	0.115578
Cyber_security	0.622589	-0.043602	0.715714	-0.186775	0.014336
Network_penetration	0.880866	-0.303633	0.428252	-0.019894	-0.086341
Host_penetration	0.905207	-0.357852	0.357332	0.066485	-0.016400
Network_reconnaissance	0.799254	-0.299305	0.574743	0.053352	0.033435
Incidence_response	0.331435	-0.058860	0.902696	0.039046	0.153309
Generalized_defense_practice	0.385760	-0.088507	0.857894	0.060743	-0.157175
Network_protocol_reverse_engineering	0.815920	-0.132613	0.220783	0.155406	0.104939
Binary_reverse_engineering	0.817171	-0.067412	0.236700	-0.011907	0.197167

(4) Rotation Matrix

1.11065	0.14845	-0.15811	-0.02997	-0.02935
0.16071	1.04703	-0.01376	0.03251	-0.00652
-0.26069	-0.04770	0.97531	0.02721	0.02815
-0.02608	0.02525	0.02864	1.09110	-0.02560
-0.07478	-0.03767	0.04361	0.06999	1.12092

(5) Final Communality Estimates

Latency	0.64260
Frequency	0.49915
Duration	0.48778
Confusion	0.63680

Self-Doubt	0.70983
Confidence	0.72227
Surprise	0.77266
Frustrated	0.70502
Cyber_security	0.70198
Network_penetration	0.81522
Host_penetration	0.83819
Network_reconnaissance	0.74198
Incidence_response	0.85899
Generalized_defense_practice	0.76701
Network_protocol_reverse_engineering	0.69960
Binary_reverse_engineering	0.72093

(6) Standard Score Coefficients

Latency	-0.033302	0.009899	0.070267	0.584871	0.090586
Frequency	-0.087436	0.066486	0.190500	-0.388798	-0.173976
Duration	0.003527	-0.021356	-0.013627	-0.426809	0.199387
Confusion	-0.003748	0.302037	0.085689	0.039052	0.056867
Self-Doubt	0.015483	0.318445	-0.047505	-0.081066	-0.135671
Confidence	0.018132	-0.209800	-0.035231	-0.049330	0.558409
Surprise	-0.018559	0.180121	0.032858	0.085502	0.612592
Frustrated	0.007359	0.315264	-0.058974	0.021695	0.071532
Cyber_security	0.102152	0.033333	0.232428	-0.140899	-0.035886
Network_penetration	0.224210	-0.030833	0.020373	-0.064939	-0.126754
Host_penetration	0.236438	-0.048934	-0.020041	0.002031	-0.056601
Network_reconnaissance	0.162923	-0.044082	0.125440	0.015590	0.002433
Incidence_response	-0.049190	-0.010834	0.418596	0.067577	0.152905
Generalized_defense_practice	-0.013792	-0.005822	0.377455	0.038001	-0.123012
Network_protocol_reverse_engineering	0.241613	0.036611	-0.079593	0.090890	0.052910
Binary_reverse_engineering	0.245782	0.056816	-0.076262	-0.018733	0.109235

(7) Variance Explained by Each Factor

Factor	Variance	Percent	Cum Percent
Factor 1	4.4386	27.741	27.741
Factor 2	2.8760	17.975	45.717
Factor 3	2.9727	18.580	64.296
Factor 4	1.4228	8.892	73.188
Factor 5	1.2052	7.532	80.721

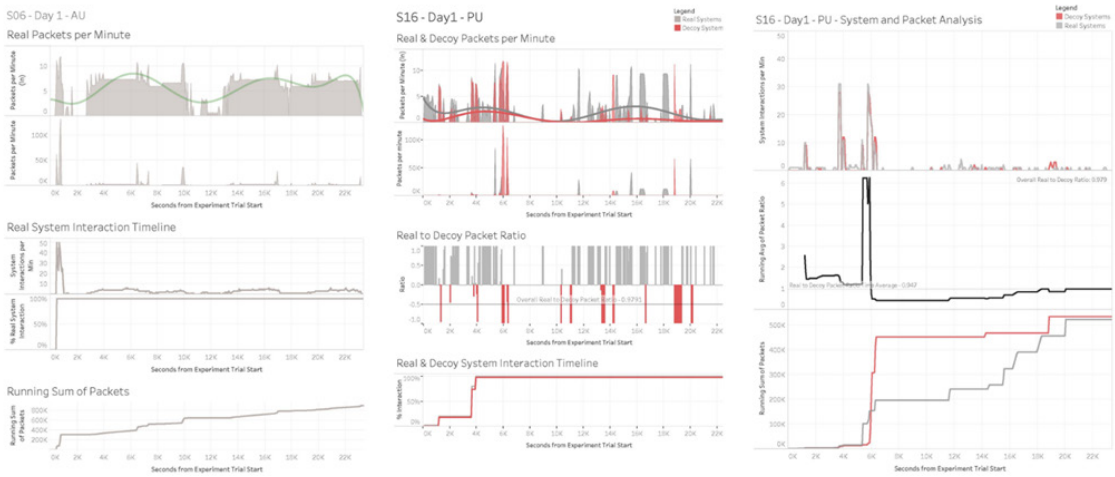
(8) Rotated Factor Loading

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Latency	-0.086536	0.010367	0.167803	0.797876	0.020586
Frequency	-0.265762	0.164210	0.368773	-0.509712	-0.130758
Duration	0.002202	-0.036300	-0.028121	-0.622142	0.298407
Confusion	0.032256	0.784220	0.182682	0.040266	0.078243

Self-Doubt	0.054263	0.828294	-0.123493	-0.100508	-0.137039
Confidence	0.051659	-0.520519	-0.029815	-0.149993	0.653237
Surprise	-0.037487	0.488950	0.102029	0.020095	0.715418
Frustrated	0.025228	0.825683	-0.135200	0.009532	0.089520
Cyber_security	0.454506	0.085459	0.543266	-0.181749	-0.002596
Network_penetration	0.831290	-0.072799	0.111797	-0.064812	-0.135459
Host_penetration	0.864288	-0.116658	0.033268	0.016672	-0.066525
Network_reconnaissance	0.642762	-0.111307	0.335924	0.029155	0.009203
Incidence_response	-0.038058	-0.042101	0.926158	0.081827	0.194997
Generalized_defense_practice	0.074601	-0.035893	0.827781	0.082737	-0.126485
Network_protocol_reverse_engineering	0.871148	0.110691	-0.090883	0.119125	0.048046
Binary_reverse_engineering	0.887488	0.168368	-0.082460	-0.041835	0.132206

SUPPLEMENTAL

Timelines of individual subject activity were generated to highlight the high degree of variability of subject activity over time and between subjects. For subjects where cyber deception was not present (Absent-Informed, Absent-Uninformed) charts only include packets sent per minute to real systems, number of real and decoy systems interactions per minute, and running sum of packets sent to real systems. For subjects where cyber deception was present (Present-Informed, Present-Uninformed) charts include packets sent per minute to real and decoy systems (actual packets per minute and natural log of packets per minute), the ratio of packets sent to real and decoy systems over time, timeline of interaction percentages with real and decoy systems, number of real and decoy system interactions per minute over time, running average of real to decoy packet ratio and a running sum of packets sent to real and decoy systems. Examples of both charts are depicted in the figure depicted on this page.



This supplemental material is available through the Dudley Knox Library at the Naval Postgraduate School.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Aerts, D. (2009). Quantum structure in cognition. *Journal of Mathematical Psychology, 53(5).* 314–348

Aggarwal, P., Du, Y., Singh, K., & Gonzalez, C. (2021). Decoys in cybersecurity: An exploratory study to test the effectiveness of 2-sided deception. *Proceedings of the 1st International Workshop on Adaptive Cyber Defense.* https://arxiv.org/abs/2108.11037v1

Alavi, M., Visentin, D. C., Thapa, D. K., Hunt, G. E., Watson, R., & Cleary, M. (2020). Exploratory factor analysis and principal component analysis in clinical studies: Which one should you use? *Journal of Advanced Nursing*, *76*(8), 1886–1889. https://doi.org/10.1111/jan.14377

Al-Shamisi, A. (2014). *Active offensive cyber situational awareness: theory and practice.* [Doctoral thesis, Brunel University] https://bura.brunel.ac.uk/bitstream/2438/13427/1/FulltextThesis.pdf

Al-Shaer, E., Wei, J., Hamlin, K.W., & Wang, C. (2019). *Autonomous cyber deception.* Springer.

Ashenden, D., Black, R., Reid, I., & Henderson, S. (2021). Design thinking for cyber deception. *In Proceedings of the 54th Hawaii International Conference on System Sciences. The University of Hawai'i at Mānoa*. http://hdl.handle.net/10125/70853

Bell, J. B. (2003). Toward a theory of deception. *International Journal of Intelligence and Counterintelligence*, *16*(2), 244–279.

Bennett, M., & Waltz, E. (2007). Counterdeception principles and applications for national security. Artech House, Inc.

Boltz, M., Dyer, R., & Miller, A. (2010). Jo Are You Lying to Me? Temporal Cues for Deception. *Journal of Language and Social Psychology*, 29(4), 458–466. https://doi.org/10.1177/0261927X10385976

Borko, H. (1968). Information science: what is it?. *American documentation*, 19(1), 3–5.

Boyd, J. (1995). *The Essence of Winning and Losing.* [Presentation slides]. https://web.archive.org/web/20110514132512/http://www.danford.net/boyd/essence4.htm

Brault S., Bideau B., Kulpa R., & Craig C. (2012) Detecting Deception in Movement: The Case of the Side-Step in Rugby. *PLoS ONE* 7(6): e37494. https://doi.org/10.1371/journal.pone.0037494

Burgoon, J.K., Buller, D.B., Floyd, K., & Grandpre, J. (1996). Deceptive realities: Sender, receiver, and observer perspectives in deceptive conversations. *Communication Research*, 23, 724–748.

Burgoon, J.K., Bliar, J.P., Qin, T., Nunamaker, J.F. (2003). Detecting deception through linguistic analysis. *Intelligence and Security Informatics*, 2665, 91–101.

Busemeyer J. R. (1985). Decision making under uncertainty: a comparison of simple scalability, fixed-sample, and sequential-sampling models. *Journal of experimental psychology. Learning, memory, and cognition*, *11*(3), 538–564. https://doi.org/10.1037//0278-7393.11.3.538

Busemeyer, J. R., Wang, Z., & Lambert-Mogiliansky, A. (2009). Empirical comparison of Markov and quantum models of decision making. *Special Issue: Quantum Cognition*, *53*(5), 423–433.

Busemeyer, J. R., & Wang, Z. (2015). What Is Quantum Cognition, and How Is It Applied to Psychology? *Current Directions in Psychological Science*, *24*(3), 163–169. https://doi.org/10.1177/0963721414568663

Busemeyer, J. R., Kvam, P. D., & Pleskac, T. J. (2019). Markov versus quantum dynamic models of belief change during evidence monitoring. *Scientific Reports*, 9(1), 18025. https://doi.org/10.1038/s41598-019-54383-9

Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin.

Cohen, F. (1998). "A Note on the Role of Deception in Information Protection," Available: http://all.net/journal/deception/deception.htm

Cohen, F., Marin, I., Sappington J., Stewart C., & Thomas, E. (2001). Red teaming experiments with deception technologies. *IA Newsletter*. http://all.net/journal/deception/experiments/experiments.html

Coltman, T., Devinney, T., Midgley, D., & Venaik, S. (2008). Formative Versus Reflective Measurement Models: Two Applications of Formative Measurement. *Journal of Business Research*, 1250–1262.

Cranford, E. A., Gonzalez, C., Aggarwal, P., Tambe, M., & Lebiere, C. (2020). What Attackers Know and What They Have to Lose: Framing Effects on Cyber-attacker Decision Making. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 64(1), 456–460. https://doi.org/10.1177/1071181320641102

Cranford, E. A., Gonzalez, C., Aggarwal, P., Tambe, M., Cooney, S., & Lebiere, C. (2021). Towards a Cognitive Theory of Cyber Deception. *Cognitive Science*, 45(7).

Custy, J., & Rowe, N. C. (2008). Conditional Entropy for Deception Analysis, *13th ICCRTS: C2 for Complex Endeavors.*

Daniel, D. C., & Herbig, K. L. (1982). Propositions on Military Deception. *Journal of Strategic Studies*, 5(1), 155–177.

Du, Y., Prébot, B., Xi, X., & González, C. (2022). Towards Autonomous Cyber Defense: Predictions from a cognitive model. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 66, 1121 – 1125.

Dunn, O. J. (1964). Multiple Comparisons Using Rank Sums. *Technometrics*, *6*(3), 241–252. https://doi.org/10.1080/00401706.1964.10490181

Dunteman, G. H. (1989). *Principal component analysis*. Sage Publications.

European Union Agency for Cybersecurity [ENISA]. (2022). *ENISA Threat Landscape 2022.* https://www.enisa.europa.eu/publications/enisa-threat-landscape-2022

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the Use of Exploratory Factor Analysis in Psychological Research.

Ferguson-Walter, K., Shade, T., Rogers, A., Trumbo, M., Nauer, K., Divis, K., Jones, A., Combs, A., Abbott, R., & Niedbala, E. (2019). The Tularosa Study: An Experimental Design and Implementation to Quantify the Effectiveness of Cyber Deception. *In Proceedings of the 52th Hawaii International Conference on System Sciences.* https://doi.org/10.24251/HICSS.2019.874

Ferguson-Walter, K. (2020). *An Empirical Assessment of the Effectiveness of Deception for Cyber Defense*. Doctoral Dissertation. https://doi.org/10.7275/z0rb-ek46

Ferguson-Walter, K. (2022). Imposing a Cyber Penalty Against Attackers with Cyber Deception. USENIX. https://www.usenix.org/publications/loginonline/imposing-cyber-penalty-against-attackers-cyber-deception

Ferguson-Walter, K., Major, M., Johnson, C., & Muhleman, D. (2021). Examining the Efficacy of Decoy-based and Psychological Cyber Deception. *In Proceedings of the 30th USENIX Security Symposium.* https://www.usenix.org/conference/usenixsecurity21/presentation/ferguson-walter

Ferguson-Walter, K., Major, M., Johnson, C., & Muhleman, D. (2021). *Examining the Efficacy of Decoy-based and Psychological Cyber Deception*. [PDF slides]. USENIX. https://www.usenix.org/system/files/sec21_slides_ferguson-walter.pdf

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). SAGE Publications.

Gartzke, E., & Lindsay, J.R. (2015). Weaving Tangled Webs: Offense, Defense, and Deception in Cyberspace. *Security Studies,* 24, 316 – 348.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance based learning in dynamic decision making. *Cognitive Science*, 27(4), 591–635

González, C., Aggarwal, P., Lebiere, C., & Cranford, E.A. (2020). Design of Dynamic and Personalized Deception: A Research Framework and New Insights. *Hawaii International Conference on System Sciences.* http://hdl.handle.net/10125/63965

Haig, B. D. (2005). Exploratory Factor Analysis, Theory Generation, and Scientific Method. Multivariate Behavioral Research, 40(3), 303–329.

Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. T. (2004a). Lies in conversation: An examination of deception using automated linguistic analysis. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 26, No. 26).

Hancock, J.T., Thom-Santelli, J., & Ritchie, T. (2004b). Deception and design: Effects of communication technology on lying behavior. *In Proceedings of CHI 2004*.

Hayatle, O., Youssef, A., & Otrok, H. (2012). Dempster-Shafer Evidence Combining for (Anti)-Honeypot Technologies. *Information Security Journal: A Global Perspective*, *21*(6), 306–316. https://doi.org/10.1080/19393555.2012.738375

Heckman, K., Walsh, M., Stech, F., O'Boyle, T. & Di-Cato, S. (2013). Active cyber defense with denial and deception: A cyber-wargame experiment. *Computers & Security*, 37:72-77

Henson, R. K., & Roberts, J. K. (2006). Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comment on Improved Practice. *Educational and Psychological* Measurement, 66(3), 393–416. https://doi.org/10.1177/0013164405282485

Heuer, R.J. (1981). Cognitive Factors in Deception and Counterdeception. In Daniel, D., Herbig, K., Reese, W., Heuer, R., Sarbin, T., Moose, P. & Sherwin, R. (Eds). *Multidisciplinary Perspectives on Military Deception* (p. 45-101). Naval Postgraduate School Technical Report. https://calhoun.nps.edu/handle/10945/31806

Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, *110*(3), 457–506. https://doi.org/10.1007/s10994-021-05946-3

Hutchins, E., Cloppert, M., & Amin, R. (2011). Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains. *Leading Issues in Information Warfare & Security Research*, *1*.

Jajodia, S., Subrahmanian, V.S., Swarup, V., & Wang, C.X. (2016). Cyber Deception: Building the Scientific Foundation. *Springer International Publishing*.

Joint Chiefs of Staff. (2018). *Cyberspace Operations* (Joint Publication 3-12). https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_12.pdf?ver=2018-07-16-134954-150

Jolliffe, I. T. (1986). *Principal Component Analysis*. Springer New York.

Kaufmann, W. J., & Smarr, L. L. (1993). *Supercomputing and the Transformation of Science.* Scientific American Library: Distributed by W.H. Freeman.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research*. Belmont, CA: Wadsworth.

Kiureghian, A.D., & Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? *Structural Safety,* 31, 105–112.

Kozyreva, A., Pleskac, T. J., Pachur, T., & Hertwig, R. (2019). Interpreting Uncertainty: A Brief History of Not Knowing. In Hertwig, R., Pleskac, T. J., & Pachur, T. (Eds.). *Taming Uncertainty* (p. 343–362). The MIT Press. https://doi.org/10.7551/mitpress/11114.003.0026

Kvam, P. D., Busemeyer, J. R., & Pleskac, T. J. (2021). Temporal oscillations in preference strength provide evidence for an open system model of constructed preference. *Scientific Reports*, *11*(1), 8169. https://doi.org/10.1038/s41598-021-87659-0

Kvam, P. D., Pleskac, T. J., Yu, S., & Busemeyer, J. R. (2015). Interference effects of choice on confidence: Quantum characteristics of evidence accumulation. *Proceedings of the National Academy of Sciences*, 112(34), 10645–10650. https://doi.org/10.1073/pnas.1500688112

Levine, T. R. (2014). Truth-default theory (TDT): A theory of human deception and deception detection. *Journal of Language and Social Psychology*, 33(4) 378–392.

Levine, T. R., & McCornack, S. A. (2014). Theorizing About Deception. *Journal of Language and Social Psychology*, 33(4), 431–440. https://doi.org/10.1177/0261927X14536397

Lockheed Martin (2022). *Cyber Kill Chain®.* https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html

McCornack, S. A., Morrison, K., Paik, J. E., Wisner, A., & Zhu, X. (2014). Information Manipulation Theory 2 (IMT2): A propositional theory of deceptive discourse production. *Journal of Language and Social Psychology*, 33(4) 348–377.

Meilán, J., Martínez-Sánchez, F., Carro, J., López, D., Millian-Morell, L. & Arana, J. (2014). Speech in Alzheimer's Disease: Can Temporal and Acoustic Parameters Discriminate Dementia? *Dementia and Geriatric Cognitive Disorders.* https://doi.org/10.1159/000356726

MITRE. (2022). *MITRE ATT&CK®.* https://attack.mitre.org/

Multiple entries. 2022. In Merriam-Webster.com.

Mun, J.C. (2017). Quantitative Research Methods Using Risk Simulator and ROV BizStats Software. Dublin, CA: IIPER Press.

National Institute of Standards and Technology (NIST). (2012). *NIST/SEMATECH e-Handbook of Statistical Methods.* https://doi.org/10.18434/M32189

Neo4j. (2021) *The Neo4j Graph Data Platform.* https://neo4j.com/product/

Newman, M.L., Pennebaker, J.W., Berry, D.S., & Richards, J.M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin,* 29, 665–675.

Nunes, E., Kulkarni, N., Shakarian, P., Ruef, A., & Little, J. (2016). Cyber-Deception and Attribution in Capture-the-Flag Exercises. In S. Jajodia, V.S. Subrahmanian, V. Swarup & C. Wang (Ed.). *Cyber Deception: Building the Scientific Foundation*, pp.149-167. Springer International Publishing.

Osborne, J. (2013). Osborne, J. W. (2013). Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data. SAGE Publications, Inc., https://dx.doi.org/10.4135/9781452269948

Pawlick, J., Colbert, E., & Zhu, Q. (2019). A game-theoretic taxonomy and survey of defensive deception for cybersecurity and privacy. *ACM Computing Surveys (CSUR)*, *52*(4), 1–28.

Pothos, E. M., & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling?. *The Behavioral and Brain Sciences*, *36*(3), 255–274. https://doi.org/10.1017/S0140525X12001525

Pothos, E. M., & Busemeyer, J. R. (2022). Quantum Cognition. *Annual review of psychology*, *73*, 749–778. https://doi.org/10.1146/annurev-psych-033020-123501

Reifman, A., & Keyton, K. (2022). *Encyclopedia of Research Design* (By pages 1637–1638; Vol. 1–0). SAGE Publications, Inc. https://doi.org/10.4135/9781412961288

Rowe N., Duong, B., and Custy, E. (2006). Fake Honeypots: A Defensive Tactic for Cyberspace," *2006 IEEE Information Assurance Workshop*, pp. 223–230. https://doi.org/10.1109/IAW.2006.1652099

Rush, G., Tauritz, D. R., & Kent, A. D. (2015). Coevolutionary Agent-based Network Defense Lightweight Event System (CANDLES). In *Proceedings of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation* (pp. 859–866). Association for Computing Machinery. https://doi.org/10.1145/2739482.2768429

Schuh, D.L. (2020). *The Cyberspace Advantage: Inviting Them In!* MITRE. https://www.mitre.org/news-insights/publication/cyberspace-advantage-inviting-them

Schreiber, J. B. (2021). Issues and recommendations for exploratory factor analysis and principal component analysis. *Research in Social and Administrative Pharmacy*, *17*(5), 1004–1011. https://doi.org/10.1016/j.sapharm.2020.07.027

Senft, M. (2019). *Lessons Not Learned: Why Our Post-9/11 Counterterrorism Experiences Should Inform Our Cybersecurity Strategy.* Modern War Institute. https://mwi.usma.edu/lessons-not-learned-post-9-11-counterterrorism-experiences-inform-cybersecurity-strategy/

Senft, M. & Corrado, E. (2022). A Network-Graph Based IT Artifact Aiding the Theory Building Process. *In Proceedings of the 55th Hawaii International Conference on System Sciences.* http://hdl.handle.net/10125/80136

Sentz, K., & Ferson, S. (2002). *Combination of Evidence in Dempster-Shafer Theory*. SANDIA Report No. SAND2002-0835. https://doi.org/10.2172/800792

Shade, T., Rogers, A., Ferguson-Walter, K.J., Elsen, S.B., Fayette, D., & Heckman, K.E. (2020). The Moonraker Study: An Experimental Evaluation of Host-Based Deception. *Hawaii International Conference on System Sciences.* http://hdl.handle.net/10125/63970

Stech, F., Heckman, K., & Strom, B. (2016). Integrating Cyber-D&D into Adversary Modeling for Active Cyber Defense. In S. Jajodia, V.S. Subrahmanian, V. Swarup & C. Wang (Eds). *Cyber Deception: Building the Scientific Foundation* (pp. 1–22)

Stoll, C. (1989). The Cuckoo's Egg: Tracking a Spy Through the Maze of Computer Espionage. New York: Pocket Books.

Sun, L., Srivastava, R. P., & Mock, T. J. (2006). An Information Systems Security Risk Assessment Model Under the Dempster-Shafer Theory of Belief Functions. *Journal of Management Information Systems,* 22(4), 109–142. https://doi.org/10.2753/MIS0742-1222220405

Swiler, L. P., & Giunta, A. A. (2007). Aleatory and Epistemic Uncertainty Quantification for Engineering Applications. *Sandia Technical Report No. SAND2007-4655C*. Sandia National Laboratories, Washington, DC.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics (6th ed.).* Pearson Education.

Tirenin, W. & Faatz, D. (1999). A concept for strategic cyber defense. *MILCOM 1999. IEEE Military Communications. Conference Proceedings (Cat. No.99CH36341)*, *1*, 458–463 vol.1. https://doi.org/10.1109/MILCOM.1999.822725

Trueblood, J. & Busemeyer J. (2011). A quantum probability account of order effects in inference. *Cognitive Science.* 35, 1518–52

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Graphics Press.

Tukey, J. (1969). Analyzing data: Sanctification or detective work? *American Psychologist, 24*(2), 83–91. https://doi.org/10.1037/h0027108

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley Pub. Co.

Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty* **5**, 297–323. https://doi.org/10.1007/BF00122574

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Walker, W. E., Harremoës, P., Rotmans, J., van der Sluijs, J. P., van Asselt, M. B. A., Janssen, P., & Krayer von Krauss, M. P. (2003). Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. *Integrated Assessment*, *4*(1), 5–17. https://doi.org/10.1076/iaij.4.1.5.16466

Wang, C., & Lu, Z. (2018). Cyber Deception: Overview and the Road Ahead. *In IEEE Security & Privacy Vol. 16*, pp. 80–85. https://ieeexplore.ieee.org/document/8328971

Whaley, B. (1982). Toward a general theory of deception. *Journal of Strategic Studies*, 5(1), 178–192.

Whaley, B. (2007). *Stratagem.* Artech House Publishers.

Whaley, B. & Aykroyd, S. S. (2007). Textbook of Political-Military Counterdeception: Basic Principles & Methods. National Defense Intelligence College.

Whaley, B., (2016). Turnabout and Deception: Crafting the Double-cross and the Theory of Outs. Aykroyd, S. S. (Ed). Naval Institute Press.

Wittek, P., Lim, I. S., & Rubio-Campillo, X. (2013). Quantum probabilistic description of dealing with risk and ambiguity in foraging decisions. In *International Symposium on Quantum Interaction* (pp. 296–307)

Yuill, J. (2006). *Defensive Computer-Security Deception Operations: Processes, Principles and Techniques*. [Doctoral dissertations, North Carolina State University]. https://repository.lib.ncsu.edu/handle/1840.16/5648?show=full

Yuill, J., Freer, F., & Denning, D. (2006). Using Deception to Hide Things from Hackers: Processes Principles and Techniques. In *Journal of Information Warfare Vol. 5, No. 3,* pp. 26–40. http://hdl.handle.net/10945/37169

Yuill, J., Denning, D., & Freer, F. (2007). Psychological Vulnerabilities to Deception for Use in Computer Security. In *DOD Cyber Crime Conference 2007.* https://faculty.nps.edu/dedennin/publications/Psychological-Vulnerabilities-to-Deception--Yuill-et-al.pdf

Zhao, Y., Hemberg, E., Derbinsky, N., Mata, G., and O'Reilly, U. (2021). Simulating a Logistics Enterprise Using an Asymmetrical Wargame Simulation with Soar Reinforcement Learning and Coevolutionary Algorithms. In *2021 Genetic and Evolutionary Computation Conference Companion (GECCO '21 Companion)*, July 10–14, 2021, Lille, France. ACM, New York, NY, USA, 9 pages. https://dl.acm.org/doi/10.1145/3449726.3463172

Zheng, R., Busemeyer, J.R. and Nosofsky, R.M. (2023), Integrating Categorization and Decision-Making. Cognitive Science, 47: e13235. https://doi.org/10.1111/cogs.13235

Zolton, M. (2022). *15 Biggest Cybersecurity Attacks in 2022*. PrivacyAffairs.com. https://www.privacyaffairs.com/cybersecurity-attacks-in-2022/

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

DUDLEY KNOX LIBRARY

NAVAL POSTGRADUATE SCHOOL

WWW.NPS.EDU

WHERE SCIENCE MEETS THE ART OF WARFARE