RULE-BASED APPROACH FOR DETECTING ADVANCED PERSISTENT THREAT USING BEHAVIORAL FEATURES OF CREDENTIAL DUMPING TECHNIQUE

NACHAAT ABDELATIF ALI MOHAMED

UNIVERSITI SAINS MALAYSIA

2022

RULE-BASED APPROACH FOR DETECTING ADVANCED PERSISTENT THREAT USING BEHAVIORAL FEATURES OF CREDENTIAL DUMPING TECHNIQUE

by

NACHAAT ABDELATIF ALI MOHAMED

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

April 2022

ACKNOWLEDGEMENT

I would like to give all thanks to the most compassionate merciful and graceful Allah, the almighty, who has provided me with many blessings that cannot be counted. I wish to convey my wholehearted gratitude to my supervisor Professor Dato' Dr. Bahari Belaton for his time, helpful comments, guidance, help and support. And thanks to Associate Professor Dr. Aman Jantan. I am also grateful to the insightful guidance, feedback, and support of the committee members, namely Professor. Dr. Mohamad Shanudin Zakaria, Dr. Mohammed FR Anbar, Dr. Manmeet (Mandy) Mahinderjit Singh, Prof. Rosni Abdullah, Dr. Wan Tat Chee, and Dr Shankar Karuppayah. I am very grateful to my best friend Mr. Othman Alothman (Bu Jarrah). I would like to sincerely thank my parents who never stop praying for me and supporting me in my education journey. I am so grateful to my dear wife ‘s Manar endless support and care. I would also like to thank my daughters MennatAllah, and Habiba, my son Mohamed for their patience and love. I wish to thank my brother, my sisters, and their families. I would like to thank all my Ph.D. classmates for every single recommendation and motivation. I am also grateful to the Dean, faculty members, lecturers, and all staff at school of computer science, Universiti Sains Malaysia (USM).

TABLE OF CONTENTS

ACKNOWLEDGEMENT ..... **ii**

TABLE OF CONTENTS ..... **iii**

LIST OF TABLES ..... **x**

LIST OF FIGURES ..... **xii**

LIST OF ABBREVIATIONS ..... **xv**

ABSTRAK ..... **xviii**

ABSTRACT ..... **xx**

CHAPTER 1 ..... **1**

1.1 Background 1

1.2 Introduction 2

1.3 Research Motivation 6

1.4 Research Problem 13

1.5 Research Questions 15

1.6 Research objectives 16

1.7 Research Scope 17

1.8 Research Methodology 19

1.9 Organization of Chapters 21

CHAPTER 2 LITERATURE REVIEW **23**

2.0 Introduction 23

2.1 The Background and Emergence of Advanced Persistent Threat (APT) 24

2.1.1 Definitions of terms 27

2.1.2 The Anatomy of APT 28

2.1.2(a) Phase one: Incursion 28

2.1.2(b) Phase two: Discovery, and Lateral Movement 31

2.1.2(c) Phase three: Capture, Command and control 32

2.1.2(d) Phase four: Exfiltration 34
2.1.2(e) The APT long period of attack on the target 35
2.1.3 Recent Occurrences of APT 36
2.1.4 Attack Vectors of APT 37
2.2 State-of-the-Art Techniques of Detecting APTs 44
2.2.1 Taxonomy of APT Detection Approaches 44
2.2.2 APT Capability (background) 45
2.2.2(a) APT Capability Capability 45
2.2.3 Proactive Methods against APT (background) 46
2.2.3.(a) Proactive Methods 46
2.2.4 Honeypot and Honeynet Strategies (background): 46
2.2.4(a) Strategies: 47
2.2.5 Moving Target Defense (background): 47
2.2.5(a) MTD: 48
2.2.6 Network indicators of APT attack (background) 53
2.2.6(a) Network indicators 54
2.3 Threat Awareness (background) 55
2.3.1 Threat Awareness (TA) 55
2.3.2 Data Breaches and Loses 56
2.3.3 Threat Awareness Shortcomings 57
2.3.4 Summary of Literature on APT Awareness 59
2.4 APT's Detection Approaches 59
2.4.1 Detection 61
2.4.2 Shortcomings of APT Approaches on Detection 66
2.4.3 Critiques on APT Detection 69
2.4.4 Summary of APT Detection Approaches 72
2.4.4(a) APT Infrastructure-based detection challenges 73

2.4.4(b) APT Attack-based detection challenges 74

2.5 Data Collection and Data Analysis Methodology on Features of APT attack. 74

2.5.1 Behaviour Analysis (background) 74

2.5.1(a) Behavior Analysis (BA) 75

2.5.2 Sandbox Analysis (background) 77

2.5.2(a) Sandbox Analysis (SA) 77

2.5.2(b) Cloud Sandbox Submission Limits 78

2.5.2(c) Steps in Setting Up Cloud Sandboxing 79

2.5.3 Network Traffic Analysis (background) 79

2.5.3.(a) Network Traffic Analysis (NTA) 80

2.5.4 Pattern Analysis (background) 81

2.5.4(a) Pattern Analysis (BA) 81

2.5.5 Anomalous Behaviour (background) 82

2.5.5(a) Anomalous Behaviour (AB) 83

2.5.6 Graph Analysis (background) 83

2.5.6(a) Graph Analysis (GA) 84

2.5.7 Network User Profiling (background) 85

2.5.7(a) Network User Profiling (NUP) 85

2.5.8 Packet Monitoring (background) 86

2.5.8(a) Packet Monitoring (BM) 86

2.5.9 Log Monitoring (background) 86

2.5.9(a) Log Monitoring (LM) 87

2.5.10 Machine learning (background) 89

2.5.10(a) Disguised exe File Detection (DeFD) 89

2.5.10(b) Malicious File Hash Detection (MFHD) 90

2.5.10(c) Malicious Domain Name Detection (MDND) 90

2.5.10(d) Malicious SSL Certificate Detection (MSSLD) 90

2.5.10(e) Random Forest 91
2.5.10(f) XGBoost 91
2.5.10(g) Support Vector Machine 91
2.5.11 Critiques of Research on Data Collection and Analysis 92
2.5.12 Summary of Data Collection and Data Analysis 94
2.6 Summary of Literature 96
2.7 Summary of APT attacks Detection Techniques 97
2.7.1 Conclusion of the Literature Review 98
2.7.2 Other Identified Areas of Future Research 98
2.7.3 The Contribution in the Area of Literature Reviewed 98
2.7.4 Approach Comparisons and interpretation 99
2.8 Summary 102
CHAPTER 3 RESEARCH METHODOLOGY 102
3.1 Introduction 103
3.2 Research Methodology 103
3.3 Justification of using experimental research 104
3.4 Understanding the gap and formulate the research problem 104
3.5 Simulate APT attack 106
3.6 Data Collection 106
3.7 Data analysis 107
3.8 Creating the CDRBA (Framework and Rules) 108
3.8 Summary 114
CHAPTER 4 PROPOSED CDRBA 115
4.1 Introduction 115
4.2 Overview of proposed approach (Rule-Based Approach). 115
4.2.1 Potential victim features selection 118
4.2.2 APT Detection 119

4.3 Requirements and characteristics of the proposed rule-based approach 119
4.4 Proposed approach 119
4.4.1 Work independently / Parallel 121
4.4.2 Why and how the features were selected 121
4.5 Potential victim machine feature extracted (Stage 1) 123
4.5.1 Random Access Memory features 125
4.5.2 Central processing Unit features 127
4.5.3 Registry system features 128
4.5.4 File system features 129
4.6 APT Detection (Stage 2) 132
4.6.1 Rule-Based for detecting abnormality behavior in RAM. 133
4.6.2 Rule-Based for detecting abnormality behavior in CPU. 134
4.6.3 Rule-Based for detecting abnormality behaviour in Registry 135
4.6.4 Rule-Based for detecting abnormality behaviour in File systems 135
4.7 Evaluation Metrix 136
4.8 Tools, and Programming languages 139
4.8.1 Python programming library PSutil 139
4.8.2 Kali Linux Distribution 140
4.8.3 DumpIt, ... 141
4.8.4 WinHex 141
4.8.5 Forensic Toolkit FTK Imager 142
4.8.6 Encase 143
4.8.7 Task Manager 143
4.8.8 ATT&CK 144
4.9 Summary 144
CHAPTER 5 CDRBA IMPLEMENTATION 146
5.1 Introduction 146

5.2 Analysis of credential dumping technique 146
5.2.1 Mimikatz 147
5.3 Testbed and link with the data collection 147
5.3.1 Setup 147
5.3.2 Implementation of APT Live Attack 148
5.3.3 Pre and during credential dumping 153
5.3.4 Implementation of features selected. 154
5.3.4(a) Implementation of the Abnormality in RAM 156
5.3.4(b) Implementation of the Abnormality in CPU 159
5.3.4(c) Implementation of the Abnormality in Registry 160
5.3.4(d) Implementation of the Abnormality file Systems 162
5.3.5 Implementation of Rules-Based behaviour detection 163
5.3.5(a) Implementation of Rules against CDT 163
5.4 Dataset Collection on Features of APT attack 164
5.4(a) Elements of CDRBA Dataset (2019) 166
5.4(b) Benchmark Dataset 168
5.5 Summary 168
CHAPTER 6 EXPERIMENTAL RESULT, AND ANALYSIS 169
6.1 Introduction 169
6.2 Test Scenarios 170
6.3 Experimental RAM abnormality behavior test scenario 170
6.3.1 Exploit, backdoor, and Trojan 175
6.3.2 CredManager 177
6.3.3 Mimikatz Scripts 177
6.3.4 Malwares 178
6.4 Experimental CPU abnormality behavior test Scenario 179
6.5 Experimental Registry abnormality behavior test Scenario 181

6.6 Experimental File Systems abnormality behavior test Scenario 182
6.6.1 Investigate DLL files 184
6.7 Overall Observations Detection Accuracy Scenario 187
6.8 CDRBA Comparison Scenario 190
6.9 Summary 193
CHAPTER 7 CONCLUSION 194
7.1 Overview 194
7.2 Introduction 194
7.3 Research Findings Implications 195
7.4 Research Contribution to Knowledge 197
7.5 Contribution of the Research to the Extant Literature 198
7.6 Mechanisms of rule-based approach 199
7.7 Study limitation 199
7.8 Recommendations 199
7.9 Future Work 199
REFERENCES 201
APPENDICES
LIST OF PUBLICATIONS

LIST OF TABLES

Page

Table 1.1 The Biggest Data Breaches of the 21st Century 8

Table 1.2 The relation between research questions and research objectives 16

Table 1.3 Research scope 18

Table 2.1 Comparison of traditional attacks and APT attacks 26

Table 2.2 Taxonomy of APT detection Techniques 45

Table 2.3 Moving Target Defense Techniques against APT's 53

Table 2.4 Critical Review of Threat Awareness Shortcomings 58

Table 2.5 Critical review of APT attacks and Shortcomings 66

Table 2.6 Behaviour and APIs as Features in the form of APT attacks 76

Table 2.7 APTs stages and corresponding AI/ML roles 84

Table 2.8 Summary of the Critical Review on Data Collection on APT 88

Table 2.9 Summary of Features on APT attacks 95

Table 2.10 Comparison rule-based approach with State-of-the-Art 101

Table 4.1 MITRE matrix table's header and entries 118

Table 4.2 Describe the whole features 124

Table 4.3 Describe the whole commands of Mimikatz 130

Table 4.4 Evaluation Metrics 135

Table 5.1 Present each behaviour feature related with resource 156

Table 6.1 Described accuracy and false positive in RAM 179

Table 6.2 The percentage of CPU usage per each application 180

Table 6.3 Describes accuracy and false positive in CPU 181

Table 6.4 Described accuracy and false positive in registry 182

Table 6.5 Describe the relation between DLL files, and CVE 185

Table 6.6 Described accuracy and false positive in file systems 187

Table 6.7 Described accuracy and false positive of CDRBA 188

Table 6.8 Describe the comparison with related studies.....192

LIST OF FIGURES

Page

Figure 1.1 Cybercriminals and their APT methods 3

Figure 1.2 Social Engineering (Drive by compromise) 5

Figure 1.3 Research Methodology 19

Figure 2.1 The Literature Reviewed Organisation 24

Figure 2.2 An APT discovery phase 31

Figure 2.3 An APT capture phase 33

Figure 2.4 An APT exfiltration causes data leak 34

Figure 2.5 Common APT attack Indicators 60

Figure 2.6 ATT&CK Matrix 70

Figure 2.7 An architecture of Sandbox analysis in the Cloud 78

Figure 3.1 CDRBA (Independent) RAM, and CPU 111

Figure 3.2 CDRBA (Independent) Windows registry, and file systems 112

Figure 3.3 The whole recourses of CDRBA (Parallel) 113

Figure 4.1 General component of proposed approach (CDRBA) 117

Figure 4.2 CDRBA stages 120

Figure 4.3 RAM evidence (during APT attack) 127

Figure 4.4 CPU before using CD application, (using Python PSUTIL) 128

Figure 4.5 CPU after using CD application, (using Python PSUTIL) 128

Figure 4.6 APT got the root and standard users 129

Figure 4.7 Mimikatz source code to filtrate file system at victim 132

Figure 4.8 The relation between the experimental results and rule-based 138

Figure 4.9 ROC Curve to measure the accuracy detection 139

Figure 4.10 Python PSutil....140
Figure 4.11 DumpIT Application....141
Figure 5.1 APT Attack....149
Figure 5.2 Payload has been created....150
Figure 5.3 Network activity has been changed....150
Figure 5.4 Windows defender, and firewall cannot detect the payload....151
Figure 5.5 Whole users and passwords, on the first victim....152
Figure 5.6 Whole super user, password, and tickets on the second victim....153
Figure 5.7 LSA system become R/W in second victim....153
Figure 5.8 APT reak Attack....154
Figure 5.9 Behaviour features of APT attack....159
Figure5.10 Implementation structure of the Abnormality in RAM....158
Figure 5.11 Implementation structure of the Abnormality in CPU....160
Figure 5.12 Implementation structure of the Abnormality in Registry....161
Figure 5.13 Implementation of the Abnormality in file systems....163
Figure 5.14 Dataset formats....165
Figure 5.15 First victim of dataset....166
Figure 5.16 Second victim of dataset....167
Figure 5.17 Third victim of dataset....167
Figure 6.1 Structure of chapter 6....169
Figure 6.2 RAM space does not affected after execute Mimikatz....170
Figure 6.3 RAM space does not affected after execute CD applications....171
Figure 6.4 RAM containing dump command after execute CD applications....171
Figure 6.5 RAM containing Trojan file after execute CD applications....172
Figure 6.6 APT try to exploit CVE-2019-0803....173
Figure 6.7 APT try to exploit CVE-2014-0322....174

Figure 6.8 APT try to exploit CVE-2012-0507 174
Figure 6.9 APT try to run number of Trojans 176
Figure 6.10 APT try to take advantage 176
Figure 6.11 APT using Exploit, backdoor, and Trojan 176
Figure 6.12 APT try to use CredMan 177
Figure 6.13 APT using Scripts to take high privilege 178
Figure 6.14 APT using other malwares 178
Figure 6.15 One CPU usage before using CD 179
Figure 6.16 One CPU usage during execute CD application 180
Figure 6.17 APT has another sessions under same user 182
Figure 6.18 File system filtered by APT after run CD Application 183
Figure 6.19 File system filtered by APT after run CD Application 183
Figure 6.20 File system filtered by APT after run CD Application 184
Figure 6.21 Accuracy, and false positive 190
Figure 6.22 Accuracy, and false positive 190
Figure 6.23 The accuracy at the level of 35 features 190

LIST OF ABBREVIATIONS

USM	Universiti Sains Malaysia
CDRBA	Credential Dumping Rule-Based Approach
APT	Advanced Persistent Threat
ATT&CK	Adversarial Tactics Techniques and Common Knowledge
CKC	Cyber Kill Chain
CDT	Credential Dumping Technique
CDA	Credential Dumping Application
CA	Credential Access
TTP	Tactics Techniques and Procedures
CAR	Cyber Analytic Repository
MTD	Moving target defines
IP	Internet protocol
CPU	Central Processing Unit
RAM	Random Access Memory
CVE	Common Vulnerability Exposure
OSINT	Open-source intelligence
ROC	Receiver Operating Characteristic
IoC	Indicators of Compromise
NLP	Natural Language Processing
TI	Threat Intelligence
ARP	Address Resolution Protocol
DNS	Domain Name System
TAL	Threat Agent Library
PTH	Using Pass-the-hash
API	Application Programming Interface
ML	Machine Learning
HIDS	Host-based intrusion detection systems
EDR	Endpoint Detection and Response
DLL	Dynamic Link Libraries
IFF	Identification Friend or Foe
OTP	One Time Password

MAC	Media Access Control
C&C	Command & Control
CIA	Confidentiality Integrity and Availability
ETDR	Endpoint Threat Detection and Response
DLP	Data Loss Prevention
SIEM	Security Information and Event Management
HIDS	Intrusion Detection Systems
TTI	Technical Threat Intelligence
CA	Content Analysis
PTH	Pass the Hash
SOC	Security Operations Centre
GPO	Group Policy object
LSASS	Local Security Authority Subsystem Service
CWE	Common Weakness Enumeration
CCE	Common Configuration Enumeration
TAXII	Trusted Automated Exchange of Indicator Information
STIX2	Structured Threat Information Expression
MAEC	Malware Attribute Enumeration and Characterization
CAPEC	Common Attack Pattern Enumeration and Classification
CybOX	Cyber Observable Expression

LIST OF APPENDICES

APPENDIX A SURVEY

APPENDIX B APPLICATIONS AND PROGRAMMING LANGUAGES

APPENDIX C THE INFECTED SYSTEM / APPLICATION AND CVEs

PENDEKATAN BERASASKAN PERATURAN UNTUK MENGESAN ANCAMAN BERTERUSAN TERMAJU MENGGUNAKAN FITUR KELAKUAN TEKNIK LAMBAKAN TAULIAH

ABSTRAK

Pergeseran dari pendekatan manual memproses data ke kaedah digital telah menjadikan data organisasi rentan terhadap pelbagai serangan oleh penjenayah siber. Advanced Persistent Threat (APT) merupakan ancaman baru-baru ini yang telah merosakkan banyak industri dan kerajaan. APT menyebabkan kerosakan besar untuk kehilangan data, pengintipan, sabotaj, kebocoran, atau pembayaran wang tebusan secara paksa kepada penyerang. Langkah-langkah keselamatan semasa menangani serangan APT melibatkan pengesanan serangan lama setelah ia berlaku dan gagal memberikan tindak balas proaktif. Penyelesaian keselamatan semasa digunakan untuk mengesan tandatangan dan tingkah laku APT setelah APT memintas keseluruhan perlindungan dan mencapai teknik pergerakan lateral, yang menjadikan penyelesaian semasa tidak berkesan untuk menyelesaikan masalah APT. Kajian ini memperkenalkan pendekatan berasaskan peraturan baru untuk mengesan serangan APT. Pendekatan ini mengenal pasti APT pada mesin mangsa yang berpotensi pertama ketika penyerang menggunakan teknik pembuangan bukti. Pendekatan berdasarkan peraturan menggabungkan, menyiasat dan memantau fitur perilaku APT dalam CPU, RAM, windows windows, dan sistem file. Pendekatan berdasarkan peraturan dicadangkan untuk mengesan serangan APT pada mesin mangsa berpotensi pertama. Pendekatan berdasarkan peraturan yang dicadangkan dalam penyelidikan ini dibina berdasarkan matrik kitaran hidup APT, Teknik Taktik Adversarial dan Pengetahuan Umum (ATT & CK) dari akses awal sehingga pengelupasan dan impak. Oleh itu, ini dapat mengesan serangan tersebut sebelum berlanjutan ke tahap yang lebih maju.

Pendekatan yang dicadangkan telah dinilai, disahkan, dan dibandingkan dengan kajian canggih lain di lapangan. Hasil penyelidikan ini disempitkan ke Memori Akses Rawak, Unit Pemprosesan Pusat, Sistem Pendaftaran Windows dan Sistem Fail. Hasilnya menunjukkan bahawa pendekatan yang dicadangkan dapat mengesan serangan APT dengan ketepatan 99.8%, dan 0,2 palsu positif berdasarkan set data Pendekatan Aturan Berdasar Kredensial (CDRBA) terpakai. Kecekapan pendekatan yang dicadangkan dinilai melalui kontras dengan pendekatan yang terkenal.

RULE-BASED APPROACH FOR DETECTING ADVANCED PERSISTENT THREAT USING BEHAVIORAL FEATURES OF CREDENTIAL DUMPING TECHNIQUE

ABSTRACT

The shift from the manual approach of processing data to the digitized method has made organizational data prone to various attacks by cybercriminals. Advanced Persistent Threat (APT) is a recent threat that has ravaged many industries and governments. APT causes enormous damages for data loss, espionage, sabotage, leak, or forceful pay of ransom money to the attackers. Current security measures of addressing APT attack involve detecting the attacks long after it has happened and failed to provide proactive responses. The current security solutions are deployed to detect APT signature and behaviour after APT bypasses the entire protections and accomplishes lateral movement technique, which makes the current solutions ineffective to resolve APT problem. This thesis introduces a new rule-based approach for detecting APT attack. This approach identifies the APT on the first potential victim's machine when the attackers use credential dumping technique. A rule-based approach incorporates, investigates and monitors APT behavioural features in the CPU, RAM, windows registry, and file systems. Rule-based approach is proposed to detect APT attack. The rule-based approach proposed in this research is built on APT lifecycle, Adversarial Tactics Techniques and Common Knowledge (ATT&CK) matrix from initial access until exfiltration and impact. This, thus, detect the attack before it develops to more advanced phases. The proposed approach has been evaluated, validated, and compared with other state-of-the-art studies in the field. The results of this research are narrowed down to Random Access Memory, Central Processing Unit, Windows Registry and File Systems. The results show that the

proposed approach is able to detect APT attack with 99.8% accuracy, and 0.2 false positive based of used Credential Dumbing Rule Based Approach (CDRBA) dataset. The efficiency of the proposed approach is evaluated through a contrast with renowned approaches.

CHAPTER 1

1.1 BACKGROUND

Since 2009, there have been many breaches of many public and private institutions worldwide, and those incidents were behind them the Advanced Persistent Threat (Zhu & Rass, 2018). APT uses sophisticated tactics and techniques to achieve the goals, espionage or sabotage (Li et al., 2016). The intrusion detection systems face great challenges to detect APT attacks compared with traditional attacks because the APT attackers know very well how these defences work and what are the methods used in these detection devices, so they implement tactics and techniques to bypass all intrusion detection devices (Li et al., 2016).

Unlike other types of attacks, the APT attack uses a modern pattern to achieve the planned objective. The attack constantly trails its target for a long period. Numerous attack cases have shown that these kinds of attacks are stealthy. They extend over a very long period until the victim or network administrator detects the malicious activities. Penetration and insertion into the network vary from one month to twenty-eight months (Siddiqi & Ghani, 2016; Alperovitch, 2011).

The original APT attack component's lifecycle consists of a twelve-stage model. This thesis summarizes the model into an abstraction of a six-stage model that encompasses the most critical stages, namely: "reconnaissance, initial compromise, credential access, lateral movement, data exfiltration, and impact. This thesis sheds light on exacerbating the credential dumping technique under the credential access tactic, which is considered the most significant central point for APT attacks at public and private institutions (Li et al., 2016).

1.2 INTRODUCTION

The evolution of the Internet and computer networks has initiated new and sophisticated types of attacks called the Advanced Persistent Threat (APT). The (APT) term was created by Colonel Gerg Rattary who was serving in the United States Air Forces in 2006 (Ghazi & Adam, 2016). APT attacks, which are furtive and orchestrated, target organizations and governments to exfiltrate confidential data (Alshamrani et al., 2019; Zhu et al., 2018; Zhang et al., 2017; Ezefosie, 2016; Alperovitch, 2011). Without being detected, the APT attackers hide in the network for a long time to steal data and critical information (Strom et al., 2018). Such attacks used lateral movement technique to leverage multiple vectors and entry points to navigate around defenses to breach the network in minutes and evade the radar of traditional security measure and detection for months.

APT attackers are well-funded and highly skilled in hacking. Meanwhile, governments typically sponsor military and intelligent units, or other highly organized groups. These groups work methodically to gain access to the target organizations (like military and economic value), and they have the capability to take advantage of zero-day vulnerability to perform their attacks. A wide range of techniques are used to exploit the earmarked organization. This includes deception to download materials, and the use of tools and techniques such as malware, Microsoft SQL injection, spam, spyware, phishing, etc. (Alperovitch, 2011).

Recent reports on APT attacks have shown that this menace continues to soar, there are many popular APT groups for example, APT38 sponsored by North Korea, APT33 sponsored by Iran, APT16 sponsored by China, APT19, APT28, and APT29 sponsored

by Russian Government, USCYBERCOM -Equation Group- sponsored by United States of America, and Unit 8200 sponsored by Israel (Boot et al.,2019).

Recent years have witnessed a drastic increase in the scale of APT attacks. Attackers exploit loopholes and vulnerabilities to control businesses to their advantage. One approach that exemplifies APT attacks is the utilization of social engineering techniques whose aim is tricking people to violating normal security procedures or deceiving employees of the targeted organization to violating or abusing legitimate access rights. while leveraging ghost net techniques to stay for a long period inside the infrastructure. This, thus, creates a scenario of an insecure environment for many online based organizations (Alshamrani et al., 2019; Luh et al., 2017; Ezefosie, 2016). APT attack activities is described in Figure 1.1.

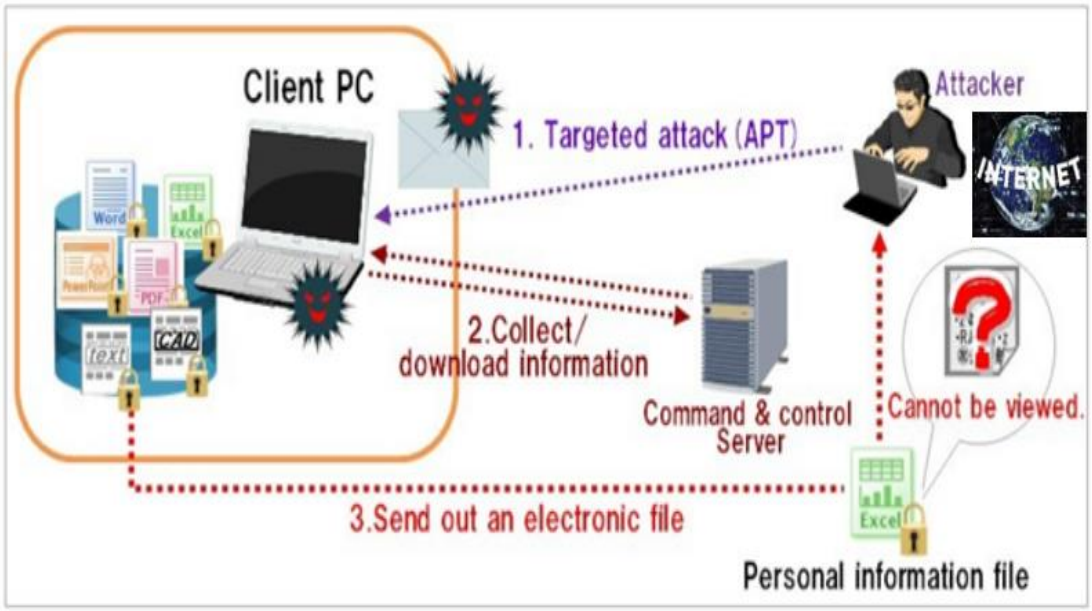

Figure 1.1. Cybercriminals and their APT methods (Source:securityaffairs.co)

Figure 1.1 demonstrates the way an APT attacker employs a complex mixture of attack methods by targeting multiple vulnerabilities within an organization. The operation may involve identifying primary persons of the victim's organization and then carrying out multiple techniques as follows:

- Social engineering attacks - Telephone based-attack
- Internet malware - infection like phishing emails to install Remote Access Tools (RAT).
- Physical malware – such as using infected USB sticks and memory cards.
- External exploitation - through injecting customized code into privileged hosts and mass vulnerability exploits (Aldawood & Skinner, 2020).

Regrettably, business organizations are not the only target of APT attacks, most government sensitive offices are on the radar of attacker eyes by APT attackers (Siddiqi & Ghani, 2016).

After successfully gaining access into the network, the attacker installs malware on the victim's computer. The attacker deepens the search to find other vulnerable hosts to pivot to and hide (their) presence to grant themselves the highest privilege in order to reach their goal (Zhu & Rass, 2018). The attacker grants themselves administrative rights to remotely control the network infrastructure and to observe and steal sensitive data. Command and Control (C&C) is a separate channel used to execute the command, which infiltrates the breached sensitive data. Using several savvy techniques, the attacker continues to (re)write codes to maintain access to the network and evade detection while carrying out malicious actions (Zhu & Rass, 2018).

The implication of APT attack causes significant data and monetary losses that may cause bankruptcy to the business. APT is increasingly recognised as a serious, worldwide concern affecting all sectors in the target countries including banks, health, education, army, water, and electricity. Hence, the problem has attracted the cyber security community to put an end to this APT problem since this may get a country involved in war electronically. Some of the ATP attacks against institutions use

machine learning (ML) algorithms, and zero-day vulnerability (Zhu et al., 2018; Corporation, 2016; Bilge & Dumitras, 2012). A frequent technique having been widely used is drive-by compromise technique that is categorized under the social engineering technique. The drive-by compromise technique is part of the Adversarial Tactics Techniques and Common Knowledge (ATT&CK), which is used as a theoretical framework in this thesis. The adversary using drive by compromise technique to take advantage of the potential victim by exploring and checking the websites which are visited daily and allowed by the organization (Golushko & Zhukov, 2020).

By injecting malware silently on the victim's machine, the APT attackers are able to use the information to attack the website. Then, an open shell is used to remote control the user potential victim's machine. This process is often tracked using a credential dumping technique by installing any applications such as Mimikatz, gsecdump, secretsdumppy, and pwdumpx.exe (Niakanlahiji et al., 2018). Figure 1.2 shows a description of the drive by downloading technique.

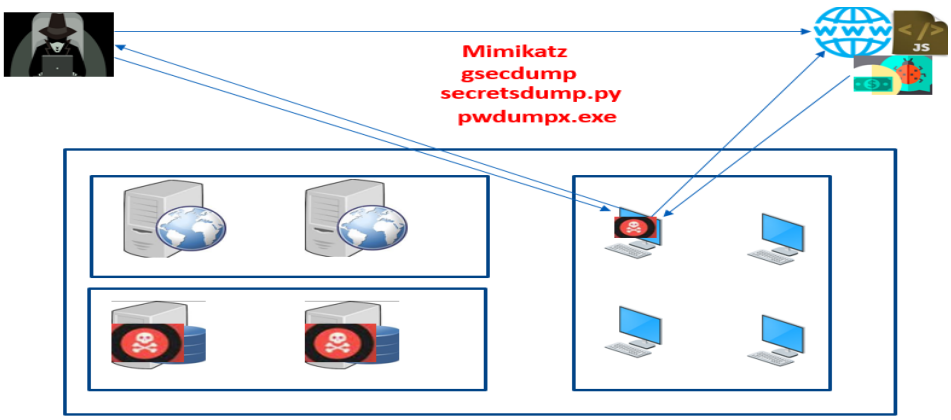

Figure 1.2. Social Engineering (Drive by compromise)

Credential dumping obtains the whole login credentials of the organization. It constitutes usernames and passwords information. This is generally in the form of a

hash or a plaintext password. In some cases, the APT attack uses the acquired credentials to perform Lateral Movement and access restricted information.

Current measurements of using signature-based, heuristic-based, behavioural-based, sandbox, and data mining techniques for addressing the APT attacks have not been able to effectively address the issue (Alshamrani, 2019; Ali, 2018; Daesung et al., 2017; Awais, 2016). Likewise, the use of intrusion detection systems has been widely reported in the literature that it is not able to protect systems and network infrastructures against APT attacks (Alshamrani, 2019; Ali, 2018).

One of the major reasons why the current approaches have failed to tackle the APT attack problem is because current methodologies deploy their solutions after lateral movement technique where there is no signature-based or suspicious behavior when the attack has already occurred. Such approaches have proven to be ineffective in that APT(s) employs advanced techniques that are difficult to differentiate from a legitimate process. Thus, there is a need to initiate the detection process from the earliest point of the attack before the attacker can get into the core process of the network infrastructure and execute the exfiltration commands.

1.3 Research Motivation

The motivation for this thesis emanated from the problems listed as follows;

Data Breaches:

The motivation for this thesis lies in the digitized nature of data stored in organization's servers, which makes the data a lucrative target for attackers. According to Verizon (2018), Data-Breach Investigation Reports (DBIR), the state-of-the-art measures for curtailing attacks are inadequate and unable to address current threats as such attacks

are executed by the use of sophisticated tools and techniques that evade detection under the radar of traditional detection methods (Widup et al., 2018).

According to the IBM, Ponemon Institute reports, 66% of incurred breaches took several months/years to be discovered, on the reverse, 84% of the same breaches took only a few hours to infiltrate the computer and network infrastructure. The report further stated that it was only 5% of the breaches being detected using traditional IDS. 69% being discovered by external entities (Verizon et al., 2018). Almeshekah & Spafford (2016) highlighted that the reality of these attacks may be significantly worse as there are unarguably some attacks that remain undiscovered and thus not reported. This research is significant because it targets APT attackers as the first potential victim in credential dumping technique before APT achieve lateral movement technique and be stealthy for a long period. This research is unique because it will eliminate prolonged characteristics of APT groups, which cost the public and private institutions 3.8 million dollars in one breach, according to Ponemon Institute and IBM 2019.

Recently, the city of Baltimore computer system was attacked with ransomware by cybercriminals who make it impossible for users in the community to access government emails, accounts, and computer system (BBC News, 2019). Also, the hackers disabled the online payment services i.e., paying water bill, purchasing properties, and many other services (BBC News, 2019). As estimated by Alexander Ivanov, the President of the Russian Association of Networks and Services, Internet operators lose around $55 million from the damage caused by spam (Adkins et al., 2013; Ramasubramanian & Prakash, 2013; Luo et al., 2005). Some of the biggest data breaches of the 21st century are highlighted in Table 1.1.

Table 1.1 Shows the Biggest Data Breaches of the 21st Century (Source: securitymagazine.com & https://www.csoonline.com)

S/N	Year	Incidence	Number of accounts / Records
1	2019	Social Media Profiles	4 Billion
2	2019	Orvibo	2 Billion
3	2019	TrueDialog	1 Billion
4	2019	First American	885 Million
5	2019	Verifications.io	808 Million
6	2019	Collection #1	773 Million
7	2019	Dream Market	620 Million
8	2019	Third-Party Facebook	540 Million
9	2019	Chinese Job Seekers	202 Million
10	2019	Canva	139 Million
11	2019	ElasticSearch Server	108 Million
12	2018	Ministry of health (Singapore)	1.5 million
13	2018	Facebook	50 Million
14	2017	Equifax	143 Million
16	2015	Anthem	78.8 Million
18	2014	JP Morgan Chase	76 Million
19	2014	Home Depot	56 Million
20	2013	Yahoo	3 Billion

In 2019, Diachenko and Troia found a trove of data flashed and easily obtainable to the public on an unsecured server, which included about 4 billion records. A total count of unique characters across all data sets transferred more than 1.2 billion users, which makes it one of the largest data leaks from a single source organization in history (Meier et al., 2019). In July 2019 Rotem and Locar found an open database connected to Orvibo Smart Home stocks, exposing more than 2 billion records. According to the researchers, Orvibo, which operates an IoT platform, claims to have about a million users. This includes private individuals who connected their smart homes, hotels and other businesses with Orvibo smart home devices (Dong et al., 2019). The data breach

affected users from several countries such as China, Japan, Thailand, United State, United Kingdom, Mexico, France, Australia, and Brazil. The data compromise includes email addresses, username, passwords, user ID, family name, and family ID (Dong et al., 2019).

A several other data security breaches reported during the year 2019 are Capital One 106 million records, Biometric Records 27 million records, Quest Diagnostics/AMA 24 million records, Ecuador Breach – 20 million records, Hostinger 14 million records, DoorDash Breach 5 million records, Choice Hotels 700,000 records, and European Hotel Group 600,000 records (Nelson & Kettani, 2020). Thus, 99% of the data was hacked (Garon, 2017; Whittaker, 2016). In 2016, the Monsoon APT attack occurred and so some Chinese industries and governmental establishments in southern Asia were targeted as victims.

Forcepoint (2016) analysed the report and stated that the attackers would send a doctored phishing email featuring captivating news concerning topics of interests to their potential victims (Nelson & Kettani, 2020). Unknown to the victims, a weaponised document is embedded in the email and the victim is enticed to open it (Fung et al., 2020). Having opened the document, the victim activates the exploits, which enables the attackers to access the victim's computer and find and extract sensitive information. It was reported that 2 malware tools "badnews and tinytyphon" were leveraged to commit the malicious act and the tools were said to be sophisticated enough to evade traditional detection methods (Fung et al., 2020).

Additionally, security countermeasures such as firewalls, or anti-virus program rely on different behavior features that can help in distinguishing intrusions and malware from regular user activity. However, current thesis efforts have not looked critically

into studying the internal behavior of core actors, and attack vector before lateral movement. In addition, it appears that in forensic investigations, information contained some resources in the potential victim such as RAM can contain sufficient evidence to fathom the complete case. And really be high level information for forensics examiner. (Hausknecht et al., 2015).

The APT attacks begins once a network is accessed by the attacker who silently discovers the network access points, exploits security implementations, and similar information. Such actions assist attackers to plan for the next step of identifying the target sensitive data. The next phase of action is either to steal or to destroy the data. The last phase is the exfiltration. The attacker tries to cover their trails and hide activities done during the attack. Such escape measures make it complex for the victim organization to trail the attacker and to identify the damage done (Siddiqi &Ghani, 2016; Symantec, 2011; Olszewski, 2018; Pakharenko, 2015; Brewer, 2014; Sanger et al., 2013; Berenbaum, 1996).

Recently, a number of 110 countries were affected by APT attack and the exact number of victims established (Thomas, 2017; Daitch, 2017; Wheatley, 2016). These incessant attacks encouraged many cyber security researchers including the researcher of this thesis to conduct more studies.

Espionage:

Recently, numerous public and private associations have focused on Advanced Persistent Threats (APTs), refined, designated and constant dangers meant to take data like intellectual property, association or state insider facts for financial, specialized political, or military reasons. Later, APTs will presumably proceed to increment and change their assault designs (Bahrami et al., 2019). APTs are undeniably challenging

to recognize and eliminate, and they can act undetected on the network for a long time, control the objective sitting tight for the chance to spilling out your data. Often, talented, and roused aggressors utilize tactics, techniques and procedures to spy on the public and private sector (Bahrami et al., 2019). Just an early identification and a solid reaction ability can help an organization to confront APTs assault. The ID of Threat Indicators and Techniques, Tactics and Procedures (TTP) of assaults just as information sharing and coordinated effort can improve counteraction and discovery abilities of association. At a similar time, a compelling employable joint effort requires the selection of normal strategies and standards (Lei et al., 2011).

In September 2015, Kaspersky Lab's Anti-Targeted Attack Platform found atypical organization traffic in an administration association organization. Investigation of this episode prompted the disclosure of a bizarre executable program library stacked into the memory of the space regulator worker. The library was enrolled as a Windows secret word channel and approached touchy information, for example, managerial passwords in cleartext. The extra examination uncovered indications of action of a formerly obscure danger actor, liable for huge scope assaults against key administrative substances (Bahrami et al., 2019). The name 'ProjectSauron' mirrors how the code creators allude to 'Sauron' in the arrangement documents. The danger actor behind ProjectSauron orders a top-of-the-top measured cyber-espionage platform in terms of technical sophistication as specialized complexity, intended to empower long-term sophistication espionage through subtle endurance systems combined with different exfiltration techniques (Ahmadvand et al., 2018). Specialized subtleties show how aggressors gained from other amazingly progressed actors to try not to rehash their missteps. In that capacity, all relics are tweaked per given objective, lessening their worth as markers of giving and take for some other casualty

(Ahmadvand et al., 2018). Typically, APT missions have a geological nexus, pointed toward removing data inside a particular locale or from a given industry. That typically brings about a few contaminations in nations inside that area or the designated business throughout the planet. Strangely, ProjectSauron is by all accounts devoted to only a few countries such as (Russia, Iran, and Rwanda), zeroed in on gathering high worth knowledge by compromising practically all key elements it might reach inside the objective region. (Ahmadvand et al., 2018).

1. ProjectSauron is a particular stage intended to empower long-term cyber-espionage campaigns.

2. All modules and organization conventions utilize solid encryption algorithms like RC6, RC5, RC4, AES, Salsa20, etc.

3. It utilizes an altered Lua engine to carry out the centre plat structure and its modules.

4. There are as many as 50 diverse module types.

5. The actor behind ProjectSauron has an exorbitant interest in communication encryption programming broadly utilized by designated administrative associations. It takes encryption keys, design records, and IP locations of the key foundation workers identified with the encryption programming.

6. It can exfiltrate information from air-gapped networks by utilizing specially-arranged USB stockpiling drives where information is put away in a space imperceptible to the activity framework.

7. The stage utilizes the DNS convention for information exfiltration and ongoing status revealing.

8. The infection vector used to infiltrate casualty networks remains obscure. The assailants use real programming appropriation channels for lateral development inside infected networks.

According to (Kaspersky), the organizations attacked are key entities that provide core state functions (government, scientific research centres, military, telecommunication, providers, and finance) (Ahmadvand et al., 2018).

1.4 Research Problem

The APT attack is so sophisticated that it has over 100 techniques and many high-level attacking tools for executing its malicious acts (Strom et al., 2018). In the wake of the attack, the attackers leverage one or more of the penetration techniques to evade networks detection. The difficulty of the APT attacks rises as a large number of APT attacks is still undiscovered by the security community (Strom et al., 2018). This obstacle of the APT attack discovery lies in finding the correct secret exploit that penetrates the machine (Liu et al., 2019). Moreover, more convoluted APT attacks are exemplified as zero-day exploit attack, that is, the attacks are put in action directly after the declaration of discovery giving little time to protectors to take proper reaction (Tang et al., 2018). A zero-day vulnerability, for instance, can be loaded by using Exploit Public-Facing Application technique to exploit SQL injection (Choi et al., 2015). However, the scanners allow it to pass through as it uses signature-based techniques. and does not know the zero-day exploit (Thanassis et al., 2014). The current detection solutions (i.e. intrusion detection, latest generation of firewall, moving target defence, anomaly threat intelligence, etc) are not able to detect APT because they all detect after lateral movement technique (Su et al., 2017) . This makes APT detection difficult for current solutions. As in (MITRE, 2018), APT attackers use numerous tactics, techniques and strategies such as (Lateral Tool Transfer, Remote Service Session Hijacking, SSH Hijacking, RDP Hijacking, Distributed Component Object Model, Windows Remote Management, SSH, VNC, Windows Remote

Management, Software Deployment Tools, Taint Shared Content, Use Alternate Authentication Material, Application Access Token, and Pass the Hash) to transport laterally inside the target enterprise environment (Strom et al., 2018). However, the present techniques and strategies have boundaries along with requiring extended permissions, growing new connections, acting new authentications. Based on those characteristics, the current solutions are applied after lateral motion to protect the infrastructure from APT attacks based on signature or suspicious behavior (Niakanlahiji et al., 2020). The purpose of this study is to address the gap in the literature which is, all detection approaches start their work after APT achieved lateral movement tactic, in which APT are free from any suspicious activity or signature after lateral movement (APT become legitimate user and normal traffic), in this case, it is impossible to detect the APT attack after a lateral movement tactic (Tanase, 2015). Despite there are many countermeasures and means of detection (i.e., using sandbox analysis, cloud sandbox analysis, network traffic analysis, moving target defense, honey boot, firewalls, anti-virus program, and intrusion detection techniques), the attackers can use their sophisticated techniques to evade detection and make it difficult to understand or analyse their intentions (Su et al., 2017). APT are able to stay a long period in the target infrastructure in the presence of these traditional security protections (Kiwia et al., 2018; Yao, 2016). To identify malicious activity on a computer network is very challenging since the attackers make their activities similar of that of a legitimate user (Matsuda et al., 2018). This challenge has prompted modern systems to continue to update defensive methods as a proactive measure, yet cases of APT attacks continue to soar (Li et al., 2019).

In summary, the problem addressed in this thesis emphasizes that APT attack is difficult detention. This is due to:

1-characteristics of APT, for instance, they use mor than 100 variant techniques and 12 tactics (Strom et al., 2018), 2- APT has capability to use Zero-Day vulnerability to target any vulnerable system over the world, 3- APT is a prolonged attack and slow motion especially after lateral movement technique, and 4- the security solutions try to detect APT based on signature, and suspicious behavior though APT does not have any signature, or suspicious behavior after lateral movement technique (Niakanlahiji et al., 2020). The rule-based approach is able to detect APT because it targets the only weak point in APT lifecycle specifically inside credential access tactic (credential dumping technique). This is the weak point because it is the only stage in which suspicious behavior occurs at the first target (Xiong et al., 2021).

1.5 Research Questions

This thesis presenting a rule-based approach for detection of APT attack at the earliest stage before data exfiltration. To investigate the main issue of the study, this thesis aims to answer these questions:

1. What are the relevant features in the RAM, CPU, Registry, and file systems is great to detect APT attack?

2. What is the most efficient technique to detection of APT attack?

3. What approach can be employed to detect the APT attack at the earliest stage to avoid data exfiltration?

1.6 Research objectives

The main goal of this research is to propose Credential Dumbing Rule-Based Approach (CDRBA) for detecting APT attacks in credential dumping technique at the first potential victim, before lateral movement technique, and avoid data exfiltration.

The research objectives derived from the research questions are as follows:

1- To identify a set of an effective features that contribute to detection of APT attack.

2- To propose rule-based mechanisms to detect APT attack in credential dumping technique.

3- To propose framework that integrate rule based with features for detecting APT attack based of using credential dumping technique.

Table 1.2. shows the mapping between research questions and objectives.

Research Questions	Research objectives
What are the relevant features in the RAM, CPU, Registry, and file systems is great to detect APT attack?	To identify a set of an effective features that contribute to detection of APT attack.
What is the most efficient technique to detection of APT attack?	To propose rule-based mechanisms to detect APT attack in credential dumping technique.
What approach can be employed to detect the APT attack at the earliest stage to avoid data exfiltration?	To propose framework that integrate rule based with features for detecting APT attack based of using credential dumping technique.

1.7 Research Scope

This thesis is designed to detect the APT attack at the earliest phase. It covers only APT detection. The study narrows down the investigation on the behavior of core actors of the computers (such as the CPU, RAM, file system) to design a rule-based approach. Credential Dumping Rule-Based Approach (CDRBA) in this thesis works in credential dumping technique, and does not target spear phishing attack, because spear phishing is a stage normally achieved prior to credential dumping.

The focus of the idea presented in this thesis is to detect malicious behavior in contrast to identifying attack signatures, where the later is very common for lateral movement technique.

In the context of this thesis, malicious behavior refers to unauthorized changes performed by person/software to critical resources of operating system such as registry, file systems, managed RAM apps, and user credential (Arnoth & Cserna, 2019; Marková et al., 2019). Whereas, a signature is a unique identifier that is associated with a known vulnerability or 'hole' in the targeted application code (Ahmadpour & Kabiri, 2020).

If APT approaches are restricted to rely only on signature-based detection, there is high possibility that APT-based malware would be fully missed. It will be very difficult to detect these malwares with detection methods that rely on signatures. As we mentioned in the research problem, these groups (APT) do not have a malicious behavior or signature.

Thus, the whole target infrastructure remains vulnerable. Regardless of how malicious code is, APT can stay months, or sometimes years without any detection, if signature-based security applications / hardware has not detected it (Stellios et al., 2019; Bhatt & Gustavsson, 2014). The recent hacked Twitter was in July 2020. The attackers were able to bypass the whole security signature based without any detection and tweet from accounts of the world famous.

The rule-based approach is tested with some live data for detection against APT attack using existing tools and techniques used by advanced persistent threat (APT) attacker groups.

Table 1.3 shows the study scope within APT identification in the victim's machine.

Table 1.3. Research scope.

	Items	Scope of research
1	Environment	Network (First potential victim)
2	Attack type	Advanced persistent threat (APT)
3	Resources and features	RAM, CPU, Registry, and file systems
4	Dataset	Created during APT attack in six formats
5	Victim Operating Systems	Windows 7, and Windows 10
6	Performance rule-based approach	Accuracy detection and false positive
7	APT Groups that used CD malicious tool (Mimikatz)	APT28, APT32, Axiom, Carbanak, HOMEFRY, Leviathan, OnionDuke, PinchDuke, Poseidon Group, Revenge RAT, Sowbug, Suckfly, Trojan.Karagany
8	Type of Dataset	Experimental (Hex, Decimal, Octal, Binary, Float, and Double)

1.8 Research Methodology

The thesis is divided into 7 chapters. Chapter 1 discusses the main research problem. Chapter 2 discusses the extant literature relevant to the current study. Chapters 3, and

4 discusses methodology adopted by this thesis as well as the development of the proposed approach to suggest practical solutions. Chapter 5- discusses the architecture of CDRBA. Chapter 6 discusses experimental results demonstrated. Chapter 7 presents conclusion, contribution, and future work. Figure 1.3 details out the methodology adopted in this thesis.

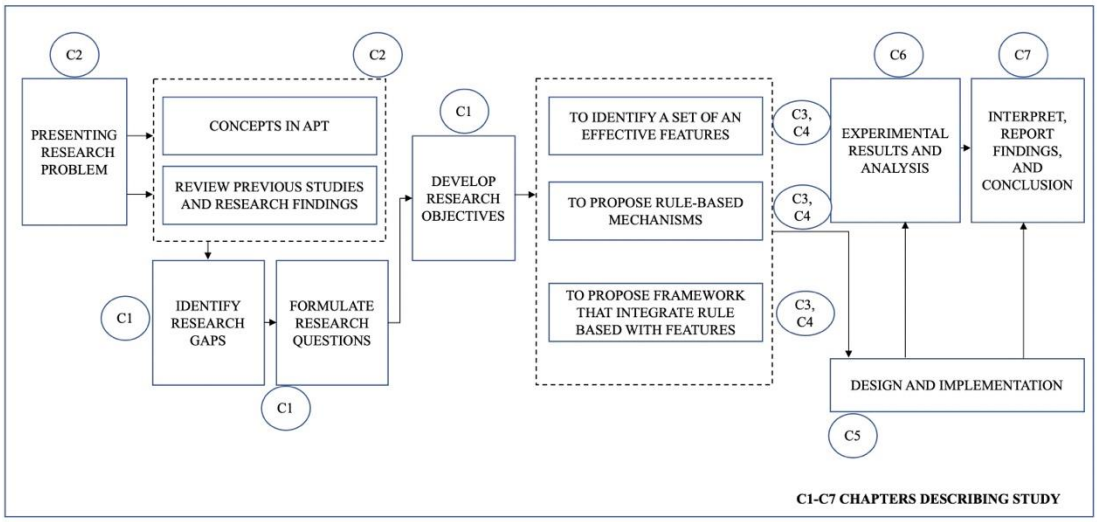

Figure 1.3. Research Methodology

This research is conducted using a combination of theoretical analysis and experiments to examine performance of rule-based approach that detects advanced persistent threat at first potential victim based on malicious behaviour of using credential dumping technique. The proposed approach includes the detection of APT attacks at earliest point in target environment to avoid data exfiltration. All the research chapters in phases as follows:

Phase 1 (C2) is the literature review chapter in which the concept of APT (attack phases, Adversarial Tactic Techniques and Common Knowledge (ATT&CK) is reviewed by discussing and analysing findings of previous studies (APT detection techniques, shortcoming, and critiques).

Phase 2 (C1) is the data analysis chapter that investigates the problem of the existing APT detection approaches through (presenting research problem, identifying the research gap, and formulating the research question). Then the approaches used to detect APT attacks are considered to identify the limitations of the existing approaches. As a result, the problem statement of this research is identified.

Phase 3 (C3, C4) is the research design and modelling. This phase presents how the proposed approach is designed. It discusses Random Access Memory (RAM) features, Central Processing Unit (CPU) features, Windows Registry features, and File Systems features. It proposes the rule-based of abnormality behaviour of each resource to achieve the research objectives.

Phase 4 (C5, C6, C7) is the performance evaluation. This phase discusses and compares the evaluation of the proposed approach with another approaches. It presents the experimental results from four resources (Random Access Memory (RAM), Central Processing Unit (CPU), windows registry, and file systems). Furthermore, it describes how to use these unique evidences to detect APT attack at first potential victim. It finally presents the research conclusions and future work recommendations.

1.9 Organization of Chapters

Chapter 2 reviews the extant literature on the detection of APT attack which encompasses many techniques that are adopted in the academia and industries on APT, spanning variants of APTs, and existing attack vectors for APTs, and current evasion methods used by attackers to avoid detections and others. This is done in accordance with the research questions and objectives. The background of the study was presented before the discussion of the literature. In the background, the definition of terms was introduced in the context of the thesis, rather than in the dictionary point of view. What

follows terms definition are state-of-the-art techniques, research gap discovery, critiques of research methods and findings and then a summary of the state-of-the-art.

Chapter 1 presents the research background, motivation, statement of the problem, questions, objectives, scope and methodology used in the study.

Chapter 3 demonstrates the research methodology phases of the proposed approach for detection APT attack at first potential victim based on the malicious behaviour resulting from using credential dumping technique.

C**hapters 4** addresses the research design and tools used for the proposed rule-based approach. It proposes the features selection and rule-based of rule-based approach. It also discusses the dataset gathered during the attack and presents the layout / topology of dataset.

C**hapters 5** presents the setup, APT attack and the 35 features.

Chapter 6 presents the findings discussion, experimental results and the comprehensive analysis of the rules achieved through sing rule-based approach. It also evaluates the performance of rule-based approach.

Chapter 7 presents the research summary, conclusions, recommendations, and future work.

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

Several components of APT are presented in this chapter to build the context and support the research performed during this thesis. The literature is organized into three major parts namely, the background knowledge of the study, state-of-the-art and the conclusion.

This background knowledge preludes with the background and emergence of Advanced Persistent Threat (APT) in section 2.1, followed by definitional terms used in APT in section 2.1.1. The anatomy of APT is presented in section 2.1.2. Section 2.1.3 presents recent occurrences and versions of APT, followed by the attack vectors in APT in Section 2.1.4. Other discussion in this section presents the mode of operations, and different types of detection techniques in full detail.

The state-of-the-art components are structure of methods, focused area, identified gaps, critiques of methods, and summary of each state-of-the-arts.

The conclusion part contains the discussion on identified flaws in past studies and gaps in APT detection, features analysis, and some contribution in respect to APT detection. Other composition in the conclusion are relationship between this thesis, existing knowledge and summary.

The literature is organized as highlighted in Figure 2.1.

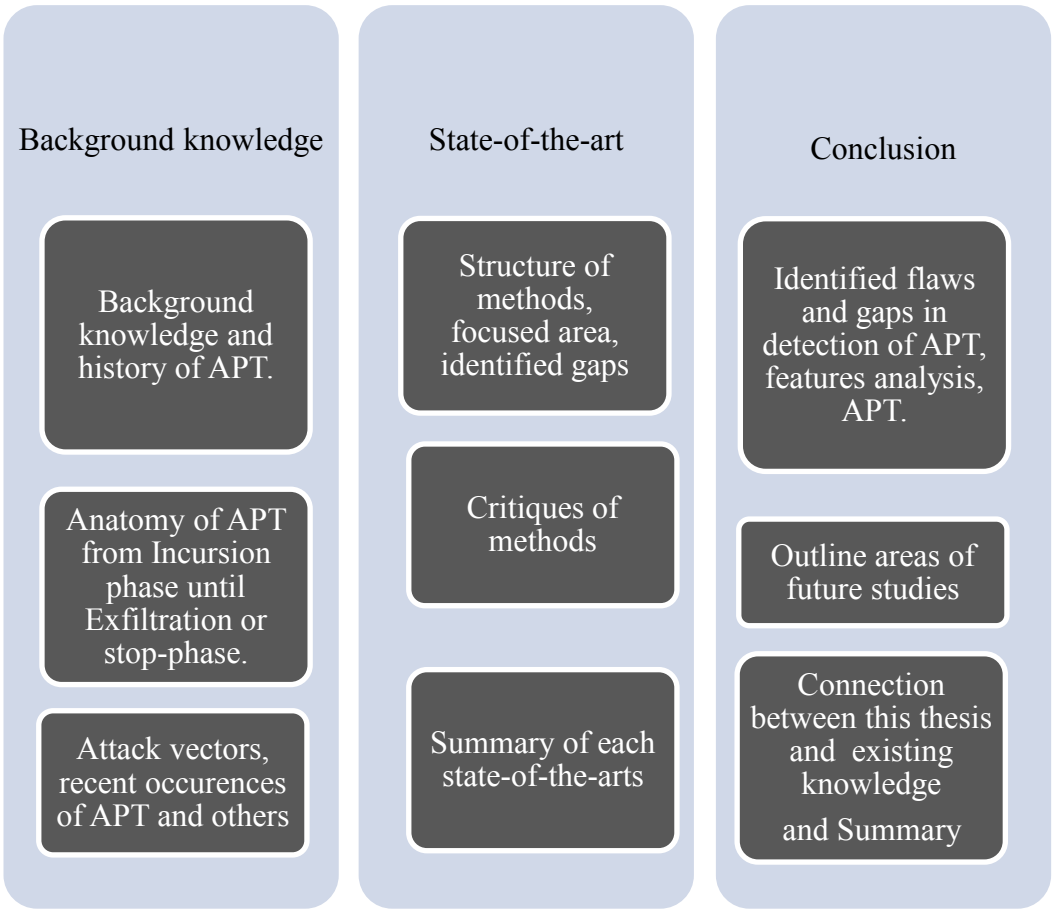

Figure 2.1. The Literature Reviewed Organisation

2.1 The Background and Emergence of Advanced Persistent Threat (APT)

Advanced Persistent Threats (APTs) are classified as a recent type of cyber security threat executed by highly skilled and well-resourced adversaries targeting precise information in governments and high-profile organizations, mostly in a long-term campaign entailing different phases/steps (Siddiqui et al., 2016).

APT is a dangerous form of attack launched through the internet by a human subject called the attacker or adversary from a computer located in a remote environment called attacker's or adversarial home base to a target computer network. The targeted computer network belongs to either the government, an organization or private persons. As mentioned in chapter 1 the (APT) term was created by Colonel