International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Leveraging Multicollinearity and Regression to Predict Advanced Persistent Threat (APT) Attacks

Veena R. C.[1], Dr. Brahmananda S. H.[2]

Submitted:14/03/2024 **Revised**: 29/04/2024 **Accepted**: 06/05/2024

Abstract: Hackers are breaking into businesses employing an Advanced Persistent Threat (APT) strategy to wreak havoc, demand ransom, and malign the company website. The APT attacker breaches an enterprise's firewall using several techniques. As technology develops, hackers are employing cunning strategies to compromise an organization's security. The goal of an APT threat is typically driven by personal, business, or government-backed organizations with vested interests in achieving it. The results of the literature review indicate that the US, India, Russia, and the UK are the main targets of an APT attack. The time has come to concentrate on predictive analysis to foresee APT threats as a preventive strategy, in addition to methods and technology to secure an organizational network. Using statistical methods like regression and multicollinearity analysis on the available threat data, this paper gives predictive modeling for APT Attacks. Secure data transfer is accomplished using hash-based cryptographic methods like MD5, SA1, and SHA1. These protocols are more harmful than previously thought, according to a study, because attackers can impersonate a client while hashing for handshake transcripts by the server. Based on previous research on APT assaults, the current paper gives the prediction of an APT attack. Based on a study of 4,296 hash keys, the distribution is 52% MD5, 11% SHA1, 28% SHA256, and 8% SHA1 according to the.exe download (Unknown). This analysis seeks to stop APTs from quickly expanding from infiltrating a single computer to controlling several computers or the entire organization. The designed model got trained with 60 types of APTs with varying signatures of MD5, SHA1, SHA256, and Unknown variants. The total number of threats used to analyze is 4,296. The proposed analysis significantly outperformed in comparison to the published accuracy of 91.80 percent [4] for early detection of APT from an unknown domain by 98.14 percent.

Keywords: *APT, Hashing, MD5, SHA1, SHA256, Regression, Multi Collinearity, Antivirus, Network Security, Hackers, Machine Learning, Threat Hunting, NIDS, EDR, Python, PyTorch. R.*

1. INTRODUCTION

More security threats have been created for company IT and its users due to the increased use of networked computers. Unauthorized access to computers is what leads to most malicious attacks. It's important to make sure that computers have safe, approved internet connectivity to mitigate cyber dangers. Work-from-home, personal, and office desktop PCs are among the machines used. The primary goal is to resolve internet-specific identity access issues, guarantee access for remote workers, and ensure acces s to other online resources. The most harmful attacks on commercial business systems are some of the APTs. Since the methods used by hackers to attack are always changing, many businesses rely on identity and access control tools that are insufficient to fend against future APT versions. More than before, APT versions and distribution are improvised and arbitrary. APTs quickly move from a single computer to control all connected computers inside an organization in a matter of minutes. Malicious software such as Backdoor Trojan loads is typically installed by APT handlers on infected PCs within the captured company network. Hackers have been utilizing both known and unknown attack patterns ever since the first global network was connected. They are becoming more aggressive in putting their information online to use the patterns as a tool to perpetrate crimes including stealing significant amounts of classified material and defacing websites with data from important national institutions. The categories of APT attackers include the following:

- **Cybercriminals:** Software developers using tools available on the dark web.
- **Competitor: They intend to access classified information to be ahead of coemption.**
- **Cyber-mercenaries: Software tool developers selling malicious paid services.**
- **Hacktivists:** Causes dangerous security threats to an enterprise.
- **Government Institutions:** Spy activities to gather inside information from institutions related to homeland security.

A prediction model for assessing APT threats can be created by properly classifying, gathering information on attack patterns, and determining the purpose of the attack. Robust methodologies can also be established to evaluate and identify the attackers.

A relatively recent APT detection technique is called

1Research Scholar, Dept. of Computer Science and Engineering GITAM University, Bengaluru, Karnataka, India
2Professor, Dept. of Computer Science and Engineering GITAM University, Bengaluru, Karnataka, India
1vchalapa@gitam.edu

malicious file hash detection (MFHD). The novel approach that is suggested uses analysis of MD5, SHA1, and SHA256 hashing to foresee dangerous assaults. A review of the literature revealed that protocols that include parts that enable MD5, SHA1, and SHA256 are more frequently the target of authentication and impersonation attacks. The hashing used in the encrypted communication protocols increases the security risk.

2. LITERATURE SURVEY

The literature survey started with the study of the book Code E. to develop an understanding of the threat and ways to defend against APT [1]. Hyunjoo et al. presented Behavior-based anomaly detection on big data [2]. Luh et al. published their research on the development of anomaly-based threat detection and interpretation system. [3]. Ibrahim Ghafir et al. proposed a novel intrusion detection approach for APT prediction. The approach involves the attack scenario reconstruction and the attack decoding. Prediction accuracy of at least 91.80% was achieved [4]. An article presented different aspects of APT and collected, sources of information on the topic. This work provides a framework to detect multi-stage APT attacks and is an early work on APT attack analysis [5]. However, during the initial phases of an APT's lifecycle, analysis with an understanding of each phase, and with the appropriate security analytics solutions in place, the threat can be mitigated, explained Ross Brewer of LogRhythm [6]. An interesting finding states that 60% of security budgets will be spent on rapid detection and response approaches [7]. The most challenging part of detecting an APT is keeping track of and relating the various steps logged over months of surveillance. Guillaume Brogi and Valerie Viet Triem Tong described Termin APT or, an APT detector that detects and emphasizes the traces left by attackers in the monitored system at various stages of an attack campaign. This uses Information Flow Tracking (IFT). to highlight APTs [8]. The book on Cryptography and Consensus dealt with extensive knowledge of patterns for encryption during information flow [9]. J Vijaya Chandra et al. worked on data and information storage security from APT attacks in cloud computing. This work is one of the initial studies involving APT and cloud computing [10]. Another article proposes a novel anomaly detection approach for modern intrusion detection systems. This applies a kind of black-list approach and considers only actions and behavior that match well-known attack patterns [11]. Diego et al. presented a thorough survey on Internet of Things (IoT) security and privacy challenges. This work deals with IoT intrinsic vulnerabilities and their implications [12]. Nurul Nuha et al. presented 4 case studies examining the reason for employee leakage behavior. The essence of these case studies is a maturity framework for organizational OSN Leakage Mitigation Capability (OSN-LMC) and lessons learned from the case analysis [13]. Terry Nelms et al. did a detailed analysis of APT risks and developed a categorization system to identify and organize the patterns used for the malicious purpose to gain unauthorized access. The work includes the characteristics of the network infrastructure used for attacks and uncovers several features that can be leveraged to distinguish between malicious and non-malicious software downloads [14]. Saurabh Singh et al. presented a survey on cloud computing security: Issues, threats, and solutions. This is a comprehensive collection of APT risks, mitigation strategies, and a comparison of published work by other researchers [15]. Xu Wang et al. analyzed the features of APT and found that the HTTP-based approach is widely used. They presented independent access, to differentiate between malicious and normal HTTP requests and validated the result based on a public dataset [16]. Another work presented an overview of different IoT operating systems, supported hardware, and future research directions in threat research [17].

Hasan, M.et al. discussed the performances of several machine learning (ML) models and compared them to predict APT attacks accurately. The ML algorithms developed are based on Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and Artificial Neural Network (ANN). The metrics applied are accuracy, precision, recall, F1 score, and area under the Receiver Operating Characteristic Curve [18].

The reason for the proposed method stems from the above findings. It suggests that most of the work is reactive. That means analysis carried out post-attack incident. The proposed work is more preventive. The proposed approach is an ML technique to learn from past incidents and apply the latest predictive modeling approach to forecast future potential threat areas.

3. APT THREAT

Researchers from the INRIA presented that the MD5 is significantly deadlier than earlier thought [19]. Hashing using MD5 (TLS 1.1) and improved hash functions like SHA-256 and SHA-512 (TLS 1.2). This is causing a high risk of APT. Table I shows the MD5, SHA1, and SHA256 Hash signature samples.

Table 1: Apt Signatures

APT	Type	Hash Pattern
Dark Hotel	MD5	8aa069860d591119af2859856ad5f063
Dragonfly	SHA256	ace12552f3a980f1eed4cadb02afe1bfb851cafc8e58fb130e1329719a07dbf0
Dark Caracal	SHA1	b0151434815f8b3796ab83848bf6969a2b2ad721

A survey paper provides more about intrusion detection systems [20] and the implementation of ML and data mining to improve vulnerabilities for APT [21]. An APT attack goes undetected initially and differs widely between regions.

Recent examples of APT havoc caused by intruders used a hacking tool previously associated with different country operatives and embedded some of their malicious code [22, 23].

I. THEORY

3.1. Multicollinearity

Multicollinearity is a statistical occurrence that occurs when two or more predictor variables in Multiple Linear Regression (MLR) models are highly correlated, implying that one can be predicted linearly from the others with a non-trivial degree of accuracy.

$$P = \alpha T + \beta T_R + e \quad (1)$$

$$P = \alpha T_R + (\alpha + \beta) T_{TR} + e \quad (2)$$

$$P = (\alpha + \beta)\, Y - \beta\, T_{TR} + e \quad (3)$$

Where,

P= real threat probability, T = real current threat

T_R = real threat from intrusion, T_{TR} = real transitory threat

T_R and T_{TR} are uncorrelated.

All these equations are equivalent. However, the correlations between the explanatory variables will be different depending on which of the three equations is considered.

In equation (1), since T and T_R are often highly correlated, we would say that there is **high multicollinearity.**

A measure that considers the correlations of the explanatory variable with the explained variable is Theil's measure [24], which is a measure of inequality among values of a distribution. which is defined as

$$m = R^2 - \sum_{i=1}^{k} (R^2 - R^2_{-i}) \quad (4)$$

Where,

R^2 = squared multiple correlations from a regression of y on $x_1, x_2, \ldots, x_k$

R^2_{-i} = squared multiple correlations from a regression of y on $x_1, x_2, \ldots, x_k$ with x_i omitted

$(R^2 - R^2_{-i})$ is termed the "incremental contribution" to the squared multiple correlations by Theil.

If $x_1, x_2, \ldots, x_k$ are mutually uncorrelated, then m will be 0 because the incremental contributions all add up to R^2.

In other cases, m can be negative as well as highly positive depending on α and β (coefficients). The variance inflation factor (VIF) is a measure to analyze the magnitude of multicollinearity of model terms and the following inferences are drawn from a predictor with other predictors:

- VIF < less than 5 - a low correlation
- VIF > 5 and <10 - moderate correlation
- VIF > 10 - high, not tolerable correlation.

The Standard Error (SE) in the output indicator is used to know how much larger the prediction error is due to the correlation with other predictors.

3.2. Regression

Regression is a statistical method to determine the strength and character of the relationship between one dependent variable and a series of independent variables. Regression analysis is used to:

- Predict the value of a dependent variable using at least one independent variable.
- Describe how changes in an independent variable affect the dependent variable.
- Dependent variable: the variable to be explained.
- Independent variable: the variable that is being utilized to explain the dependent variable.

4. PREDICTIVE MODELING FOR APT

Predictive modeling uses statistical algorithms to improve automated decisions through historical data.

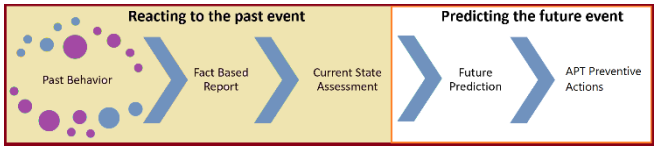

Fig 1: APT Prediction Flow

In the present work, an implementation is proposed as shown in Fig. 1 to predict malicious APT using MD5, SHA1, and SHA256 signatures. The proposed approach is to build a knowledge base using different signatures from public data [25] and do predictive modeling using R programming-based statistical analysis. The predicted model and the updated database can be used to validate a new authentication request from an unknown requester using a simple application using python. The algorithm is shown in Algorithm 1. Overall implementation of the algorithm involves a threat module comprising of an "Analysis module" using R programming. This includes multicollinearity test and regression analysis of a pattern, post analysis, and use of the finding to add in the alarm using the python-based ML implantation [26].

Algorithm 1 APT prediction using MD5, SHA1, and SHA256

For Each Authentication Request

Check if *ask* using MD5, SHA1, and SHA256 patterns

If *ask* is Yes Then

Then activate the "Threat module"

Else

Proceed to the subsequent level of validation for several other negotiating protocols.

End

For Each *pattern*

Analyze using the "Analysis module"

Check the *existence* in the database.

If *existence* is Yes Then

Send Alarm with details & history of attacks, multicollinearity.,

Else

Search in the cloud for the *existence*

If the *Search* is True then

Send Alarm with details & history of attacks, multicollinearity.,

Else

Send an Alarm stating "Unknown"

Decline Request

End

End

Update History and Status information in the database.

End

---

5. PREDICTIVE MODELING DATA

Fig. 2 shows Malicious threats pattern was used to train the algorithm and perform statistical analysis. Each of the different types of attack has additional signature variants.

Type	Hash Count	Type	Hash Count	Type	Hash Count
A01_E40	38	A001_F013	12	TA505	118
A42_D50	9	APT3	9	TEMP_Veles	54
Dark Caracal	37	APT30	45	A31_D57	27
A01_F13	11	A01_F46	45	A01_F29	27
Dragonfly	87	APT32	43	Threat Group-33	261
Elderwood	66	A01_F44	43	A001_E088	86
FIN6	157	APT37	10	Thrip	6
A001_F049	47	A01_F11	10	A1_F8	6
A051_C104	53	APT39	5	Tropic Trooper	215
A106_C161	56	A1_F6	5	A001_F028	26
NEODYMIUM	101	APT41	121	A030_C089	60
A001_E063	61	Darkhotel	40	A091_C219	129
A084_C106	22	A01_F08	7	Turla	395
PLATINUM	178	A01_F74	73	A001_F327	325
A001_F056	54	Deep Panda	84	A329_D365	37
APT18	43	Gamaredon Grou	217	A368_D462	95
A01_F37	36	A001_C219	217	A01_F07	5
APT19	8	APT-C-36	103	A09_D12	4
A1_F9	8	A001_F104	102	A14_D20	7
APT28	112	A032_C053	22	Whitefly	16
		Number of #	**4296**		
		Type of APT	**60**		

Fig 2: APT patterns used for analysis and training

Fig. 3 shows the implementation methodology of the APT threat module. Naive Bayes algorithm used for prediction. The threat module is trained with signatures of MD5 (TLS 1.1), SHA-256, and SHA-512 (TLS 1.2) to address a high risk of APT. Table I shows some of the sample signatures used. The model is utilized to know and understand the signatures, and the algorithm generates an alarm for any new requests based on the learning.

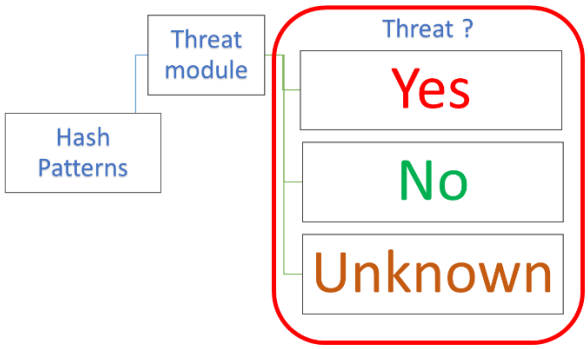

Fig 3: Methodology of the ML

Fig. 4 shows the number of variants based on APT. It also implies that, while the attack could take the same approach to breach the system, the signatures for the attack may differ.

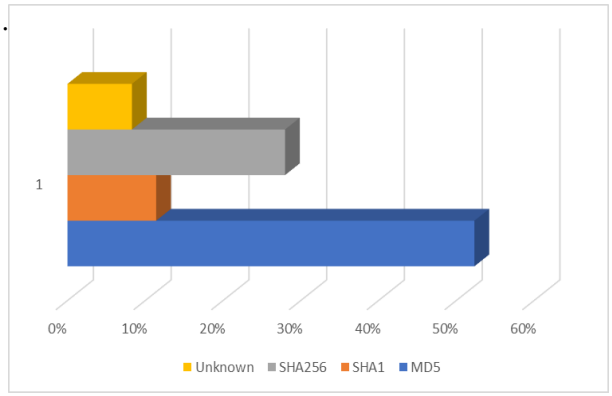

Fig 4: Hash-wise variants of training data

6. THREAT MODULEAPPLICATION

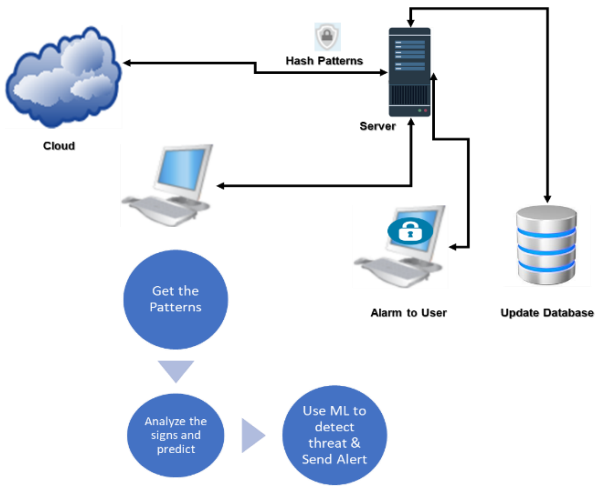

Fig 5: APT Application Workflow

Fig. 5 shows the designed application workflow. When a malicious request is received by the server, it sends information for analysis to the treat module. The threat module validates the signature and as detailed in Algorithm 1, necessary action is initiated. To test a pattern one can, get the public database created read from the DNS log.

7. RESULTS

7.1. Multicollinearity

The objective of multicollinearity analysis is to validate the dependency of a predicting variable for an APT pattern on the targeted environment (Windows, OS X, Android, Linux) and Age (Years). The APT attack goes undetected and differs widely between regions with the mean dwell time for 2018 in the Americas as 71 days, EMEA as 177 days, and APAC as 204 days [27]. Hence, age is important in analysis and detection. Table II shows input data used for multicollinearity analysis.

Fig. 6 shows the designed application output using R.

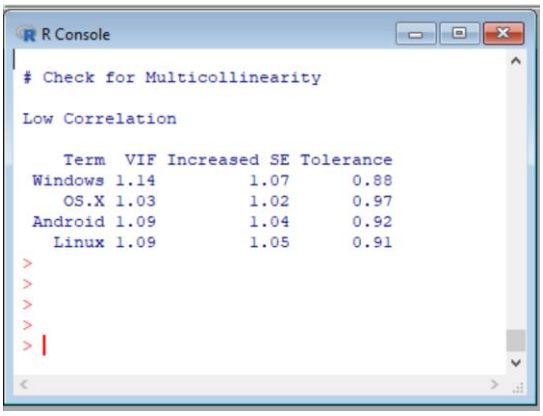

Fig 6: Multicollinearity Result

Table 2: Apt Signatures

APT_NAME	FIRST_KNOWN	Age (Years)	Windows	OS X	Android	Linux
TOPINAMBOUR	2019	3	1	0	0	0
TAJMAHAL	2013	9	1	0	0	0
SNEAKYPASTES	2018	4	1	0	0	0
OCTOPUS	1990	32	1	0	0	0
FRUITYARMOR	2018	4	1	0	0	0
MUDDYWATER	2017	5	1	0	0	0
OLYMPIC DESTROYER	2017	5	1	0	0	0
ZOOPARK	2015	7	0	0	1	0
WHITEBEAR	2016	6	0	0	0	0
SKYGOFREE	2014	8	1	0	1	0
SHADOWPAD	2017	5	1	0	0	0
SATELLITE TURLA	2007	15	1	0	0	0
PENQUIN TURLA	2010	12	0	0	0	1
LAMBERTS	2008	14	1	1	0	0
EXPETR	2017	5	1	0	0	0
BLACKOASIS	2015	7	1	0	0	0
ATMITCH	2016	6	1	0	0	0
WANNACRY	2017	5	1	0	0	0
SPRING DRAGON	2012	10	1	0	0	0
BLUENOROFF	2016	6	1	0	0	0
SHAMOON 2.0	2016	6	1	0	0	0
STONEDRILL	2016	6	1	0	0	0
STRONGPITY	2016	6	1	0	0	0
SAGUARO	2009	13	1	0	0	0
DROPPING_ELEPHANT	2016	6	1	0	0	0
CARBANAK 2.0	2015	7	1	0	0	0
PROJECT SAURON	2011	11	1	0	0	0
Animal Farm	2007	15	1	0	0	0
Kimsuky	2011	11	1	0	0	0
CROUCHING YETI	2010	12	1	0	0	0
COSMICDUKE	2012	10	1	0	0	0
BLACK ENERGY	2010	12	1	0	0	0
DESERT FALCONS	2011	11	1	0	0	0
HACKING TEAM RCS	2008	14	0	1	1	0
NETTRAVELER	2004	18	0	0	0	0
MINIDUKE	2008	14	0	0	0	0
EQUATION	2002	20	0	0	0	0
Naikon's Aria	2009	13	0	0	0	0
TURLA	2007	15	0	0	0	0
BLUE TERMITE	2013	9	0	0	0	0
SOFACY	2008	14	1	0	0	0
ADWIND	2012	10	0	0	0	0
POSEIDON	2005	17	0	0	0	0
CLOUD ATLAS	2014	8	0	0	0	0
CARBANAK	2013	9	0	0	0	0
REGIN	2003	19	0	0	0	0
DARK HOTEL	2007	15	0	0	0	0
EPIC TURLA	2012	10	0	0	0	0
FINSPY	2007	15	0	0	0	0
MINIFLAME	2010	12	1	0	0	0
WINNTI	2009	13	1	0	0	0
SABPUB	2012	10	1	0	0	0
WILD NEUTRON	2011	11	1	1	0	0
COZYDUKE	2014	8	1	0	0	0
DUQU 2.0	2014	8	1	0	0	0
HELLSING	2012	10	1	0	0	0
LAZARUS	2009	13	1	0	0	0
PROJECT SAURON	2011	11	1	0	0	0
CARBANAK 2.0	2015	7	1	0	0	0
DROPPING	2016	6	1	0	0	0
SAGUARO	2009	13	1	0	0	0
STRONGPITY	2016	6	1	0	0	0
STONEDRILL	2016	6	1	0	0	0
SHAMOON 2.0	2016	6	1	0	0	0
BLUENOROFF	2016	6	1	0	0	0
SPRING DRAGON	2012	10	1	0	0	0
WANNACRY	2017	5	1	0	0	0
ATMITCH	2016	6	1	0	0	0
BLACKOASIS	2015	7	1	0	0	0
EXPETR	2017	5	1	0	0	0
LAMBERTS	2008	14	1	1	0	0
PENQUIN TURLA	2010	12	0	0	0	1
SATELLITE TURLA	2007	15	1	0	0	0
SHADOWPAD	2017	5	1	0	0	0
SKYGOFREE	2014	8	1	0	1	0
WHITEBEAR	2016	6	1	0	0	0
ZOOPARK	2015	7	0	0	1	0

for analysis percent distribution of load patterns used is shown in Table III. That is the hash pattern used for download.

Table 3: Load Signatures

Load 1	Load 2	Load 3	Load 4
MD5	SHA1	SHA256	Unknown
52%	11%	28%	8%

As shown in Fig. 6 and 8, the result shows low collinearity amongst the identified environment parameter (Windows, OS X, Android, Linux) and Age (Years).

This an important information as the predominantly conception was that there is a high degree of collinearity with environment variables and age of APT threat.

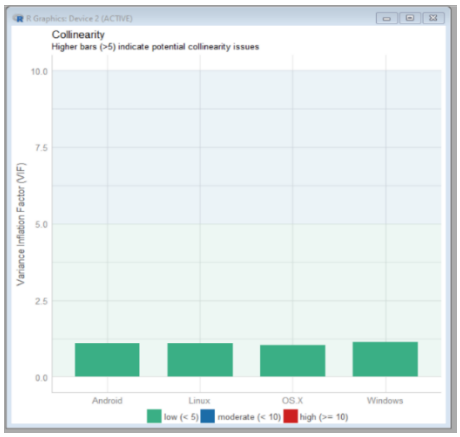

Fig 8: Multicollinearity plot using R

7.2. Regression

APT Type	Load Count	Type
A001_F011	2	1
A001_F013	2	1
A001_F327	34	1
A01_E40	1	1
A01_F29	5	1
APT28	2	1
APT37	2	1
TEMP_Veles	15	1
Turla	38	1
A030_C089	6	2
A032_C053	3	2
A368_D462	2	2
TA505	3	2
Tropic Trooper	6	2
Turla	2	2
A42_D50	5	3
A001_E088	1	3
A001_F6	2	3
A31_D57	13	3
APT39	2	3
TEMP_Veles	30	3
Threat Group-3390	1	3
A001_C219	3	4
A001_F74	2	4
APT3	6	4
Deep Panda	2	4
Gamaredon Group	3	4
	193	

Fig 9: Downloads used for different load types

Figure 9 shows the count of an executable-based APT load of different types (Load 1 to 4) used for regression.

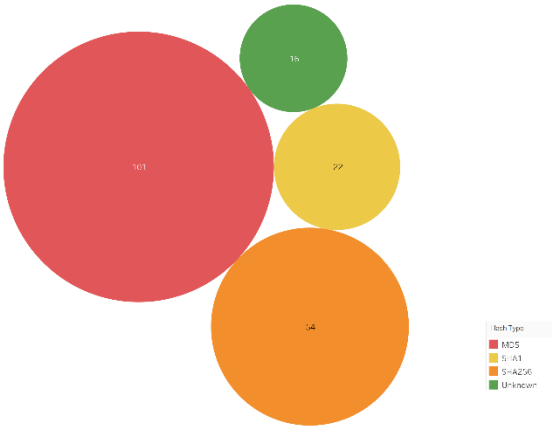

Fig 9: Load count distribution of 193

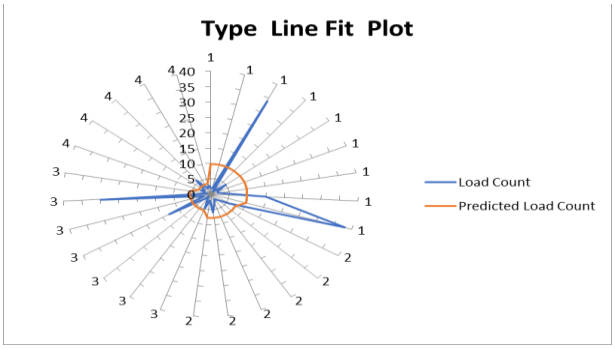

Fig 10: Current Vs. Predicted load count

Based on the identified multicollinearity, age, and available public data insight a regression analysis was carried out on 193 loads (.exe downloads). The analysis predicted that the chances of further attack due to such loads are significantly low because of the historical insight into the patterns.

The prediction accuracy of 98.15% is obtained as in Fig. 11.

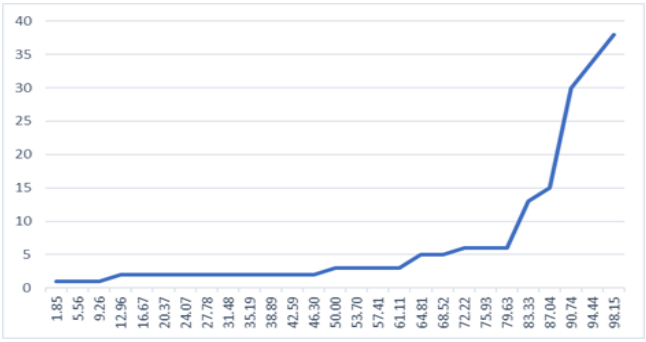

Fig 11: Predicted load Vs. Prediction Accuracy

8. ANALYSIS OF RESULT

We have taken 4296 test signatures for analysis and further testing the application with the breakup of the result as shown in Fig. 12.

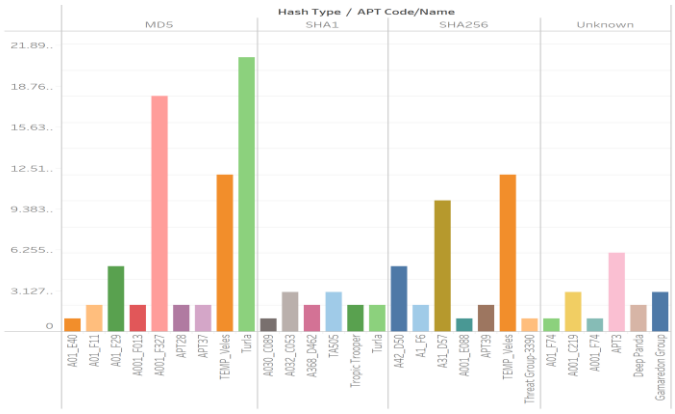

Fig 12: APT-specific distinct load

As shown in Fig. 11, Turla, A001_F327, and Temp_Veles pose the highest probability of threat risks.

9. COMPARISON ANALYSIS

Table 4: Load Signatures

Reference	Reported Accuracy
I. Ghafir et al. [4]	4%
Wu et al. [28]	76%
T. Javaheri et al. [29]	91.66%
Present work	98.15%

As shown in Fig 2, 4296 test signatures for analysis were used with 60 types of APT. This provided better training for the algorithm and hence obtained better accuracy in prediction compared to others as shown in Table IV. It shows a comparison of the obtained result with other published APT work [4] and regression-based prediction work [28,29]. From the comparative analysis, we can observe that the prediction accuracy of 98.15% for the present approach is significantly higher than the published result.

10. CONCLUSION

APT attacks use several mechanisms to breach the cyber security of an enterprise. Threat generators can impersonate clients on servers that support hashing for handshake transcripts. This work provides insight and prediction based on experimental data on APT attacks. The analysis of 4,296 hash keys based on the .exe download shows a distribution of 52%(MD5), 11% (SHA1), 28% (SHA256), and 8% (Unknown). The developed model is trained with 60 types of APTs with varying signatures of MD5, SHA1, SHA256, and Unknown variants the total number of threats used to analyze is 4,296. The proposed analysis has higher accuracy of 98.14% compared to the published accuracy of 91.80% [4] for early detection of APT from an unknown domain. Also, a comparison of the regression approach shows the present approach is well ahead of 76%, and 91.66% [28,29].

11. FUTURE

The present work can be further extended to predict APT campaigns based on two, three, and four correlated alerts. Further work can be done to assess collinearity amongst the other environmental parameters such as processors, data lake used, IoT integration points, and firewall type.

DECLARATION:

Ethical Approval

Institutional Review Board approval was not required.

Consent for Participate

All contributors agreed and given consent to participate.

Consent for Publication

All contributors agreed and given consent to Publish.

Data availability

No data, models, or code were generated or used during the study

Competing interests

None

Funding

The authors state that this work has not received any funding.

Author Contribution

The authors confirm contribution to the paper as follows and all authors reviewed the results and approved the final version of the manuscript.

Acknowledgements

The authors would like to thank the Deanship of Universiti Teknologi Malaysia for supporting this work.

Referencs

[1] Code E. Advanced Persistent Threat. Understanding the Danger and How to Protect Your Organization. 1st Edition, Amsterdam: Elsevier 2012.

[2] Hyunjoo Kim, Jonghyun Kim, Ikkyun Kim, Tai-myung Chung. Behavior-based anomaly detection on big data. Australian Information Security Management Conference 2015; 13: 73-80.

[3] Luh R, Schrittwieser S, Marschalek S, Janicke H. Design of an Anomaly-based Threat Detection & Explication System. International Conference on Information Systems Security and Privacy 2017; l3: 397-402.

[4] Ghafir I, et al. Hidden Markov Models and Alert Correlations for the Prediction of Advanced Persistent Threats. in IEEE Access 2019; 7: 99508-99520. doi: 10.1109/ACCESS.2019.2930200.

[5] Merete Ask, Petro Bondarenko, John Erik Rekdal, Andre´ Nordb, Pieter Bloemerus, and Dmytro Piatkivskyi. Advanced persistent threat (apt) beyond the hype. Project Report in IMT4582 Network Security at Gjøvik University College, Springer 2013: 168

[6] Parth Bhatt, Edgar Toshiro Yano, and Per Gustavsson. Towards a framework to detect multi-stage advanced persistent threats attacks. In 2014 IEEE 8th International Symposium on Service Oriented System Engineering. IEEE 2014.

[7] Ross Brewer. Advanced persistent threats: minimizing the damage. Network Security 2014; 4(4): 5–9.

[8] Guillaume Brogi and Valerie Viet Triem Tong. TerminAPTor: Highlighting advanced persistent threats through information flow tracking. In 2016 8th IFIP International Conference on New Technologies, Mobility and Security (NTMS). IEEE 2016.

[9] Weigold. Blockchain, cryptography, and consensus 2016.

[10] Vijaya Chandra J, Narasimham Challa, and Mohammed Ali Hussain. Data and information storage security from advanced persistent attack in cloud computing. International Journal of Applied Engineering Research 2014; 9(20): 7755–7768,

[11] Ibrahim Ghafir, Vaclav Prenosil, Mohammad Hammoudeh, Francisco Aparicio-Navarro J, Khaled Rabie, and Ahmad Jabban. Disguised executable files in spear-phishing emails. In Proceedings of the 2nd International Conference on Future Networks and Distributed Systems -ICFNDS. ACM Press 2018.

[12] Diego Mendez Mena, Ioannis Papapanagiotou, and Baijian Yang. Internet of things: Survey on security. Information Security Journal: A Global Perspective 2018; 27(3): 162–182.

[13] Nurul Nuha Abdul Molok, Atif Ahmad, and Shanton Chang. A case analysis of securing organisations against information leakage through online social networking. International Journal of Information Management 2018; 43: 351–356.

[14] Terry Nelms, Roberto Perdisci, Manos Antonakakis, and Mustaque Ahamad. Towards measuring and mitigating social engineering software download attacks. In USENIX Security Symposium 2016: 773–789.

[15] Saurabh Singh, Young-Sik Jeong, and Jong Hyuk Park. A survey on cloud computing security: Issues, threats, and solutions. Journal of Network and Computer Applications 2016; 75: 200–222.

[16] Xu Wang, Kangfeng Zheng, Xinxin Niu, Bin Wu, and Chunhua Wu. Detection of command and control in advanced persistent threat based on independent access. In 2016 IEEE International Conference on Communications (ICC). IEEE 2016.

[17] Zikria YB, Kim SW, Hahm O, Afzal MK, Aalsalem MY. Internet of Things (IoT) Operating Systems Management: Opportunities, Challenges, and Solution. Sensors 2019; 19: 1793.

[18] Hasan M, Islam MM, Zarif MII, Hashem MMA. Attack and Anomaly Detection in IoT Sensors in IoT Sites Using Machine Learning Approaches. Internet Things 2019; 7: 100059.

[19] Lucian C. Ongoing MD5 support endangers cryptographic protocols. https://www.computerworld.com/article/3020066/ongoing-md5-support-endangers-cryptographic-protocols.html

[20] Khraisat A, Gondal I, Vamplew P, Kamruzzaman J. Survey of Intrusion Detection Systems: Techniques, Datasets and Challenges. Cybersecur. 2019; 2: 20.

[21] Alloghani M, Al-Jumeily D, Hussain A, Mustafina J, Baker T, Aljaaf AJ. Implementation of Machine Learning and Data Mining to Improve Cybersecurity and Limit Vulnerabilities to Cyber Attacks. In Nature-Inspired Computation in Data Mining and Machine Learning 2020; 3: 47–76.

[22] Web source: https://www.cyberscoop.com/chinese-iranian-hackers-front-companies/

[23] Web source: https://www.cyberscoop.com/china-israel-iran-fireeye-hacking/

[24] Plat D. IC2: Inequality and Concentration Indices and Curves. R package version 1.0-1. https://CRAN.R-project.org/package=IC2 2012.

[25] https://github.com/CyberScienceLab/Our-Datasets

[26] Veena RC, Brahmananda SH. A Significant Detection of APT using MD5 Hash Signature and Machine Learning Approach. Web source: https://www.mandiant.com/resources/m-trends-2021 2021.

[27] Wu X, Hui H, Niu M, Li L, Wang L, He B, and Yang X. Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: A multicentre study. Eur. J. Radiol. 2020; 128: 109041

[28] Javaheri T, et al. Covid CTNet: An open-source deep learning approach to identify COVID-19 using CT image. arXiv:2005.03059. [Online], Available: http://arxiv.org/abs/2005.03059 2020.

[29] Liu Y, Chen Y, Yu H, Fang X, Gong C. Real Time Expert System for Anomaly Detection of Aerators Based on Computer Vision Technology and Existing Surveillance Cameras. arXiv 2018, arXiv:1810.04108 2018.

[30] Glossary: Common DDoS Attack Types, Corero. Available online: https://www.corero.com/blog/glossary/ 2019.

[31] Rajendran B. DNS amplification & DNS tunneling attacks simulation, detection and mitigation approaches. 2020 International Conference on Inventive Computation Technologies (ICICT) 2020: 230-6.

[32] Zhang K, Ji W, Li N, Wang Y, Liao S. Detection of malicious domain name based on DNS data analysis. J Phys Conf Ser. 2020; 1544: 012169.

[33] Palau F, Catania C, Guerra J, Garcia S, Rigaki M. DNS tunneling: a deep learning based

lexicographical detection approach. Cryptography and Security 2020.

[34] Vissers T, Barron T, van Goethem T, Joosen W, Nikiforakis N. The wolf of name street: hijacking domains through their nameservers. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security 2017: 957-70.

[35] Alharbi F, Chang J, Zhou YC, Qian F, Qian ZY, et al. Collaborative client-side DNS cache poisoning attack. IEEE INFOCOM 2019-IEEE Conference on Computer Communications 2019.

[36] Diego Mendez Mena, Ioannis Papapanagiotou, and Baijian Yang. Internet of things: Survey on security. Information Security Journal: A Global Perspective 2018; 27(3): 162–182.

[37] Nurul Nuha Abdul Molok, Atif Ahmad, and Shanton Chang. A case analysis of securing organisations against information leakage through online social networking. International Journal of Information Management 2018; 43: 351–356.

[38] Daesung Moon, Hyungjin Im, Jae Lee, and Jong Park. MLDS: Multi-layer defense system for preventing advanced persistent threats. Symmetry 2014; 6(4): 997–1010.

[39] Kara Nance and Matt Bishop. Introduction to deception, digital forensics, and malware minitrack. In Proceedings of the 50th Hawaii International Conference on System Sciences 2017.

[40] Terry Nelms, Roberto Perdisci, Manos Antonakakis, and Mustaque Ahamad. Towards measuring and mitigating social engineering software download attacks. In USENIX Security Symposium 2016: 773–789.

[41] Protecting Your Critical Assets Lessons Learned from "Operation Aurora" By McAfee Labs and McAfee Found stone Professional Services, 2010. International Journal of Computer Applications 2016; 141(13): 0975 -8887.

[42] Adelaiye OI, Showole A, & Faki SA. Evaluating Advanced Persistent Threats Mitigation Effects: A Review. International Journal of Information Security Science 2018; 7(4): 159-171.

[43] Chen W, Helu X, Jin C, Zhang M, Lu H, Sun Y, & Tian Z. Advanced persistent threat organization identification based on software gene of malware. Transactions on Emerging Telecommunications Technologies 2020; 31(12): e3884.

[44] Joloudari JH, Haderbadi M, Mashmool A, GhasemiGol M, Band SS, & Mosavi A. Early detection of the advanced persistent threat attack using performance analysis of deep learning. IEEE Access 2020; 8: 186125-186137. https://ieeexplore.ieee.org/abstract/document/9214817.

[45] Steffens T. Attribution of Advanced Persistent Threats. Springer Berlin Heidelberg. https://link.springer.com/book/10.1007%2F978-3-662-61313-9 2020.

[46] Yan D, Liu F, & Jia K. Modeling an information-based advanced persistent threat attack on the internal network. In ICC 2019-2019 IEEE International Conference on Communications (ICC) 2019: 1-7.

[47] Zou Q. An Approach for Detection of Advanced Persistent Threat Attacks. Computer. IEEE Computer Society. https://www.researchgate.net/publication/347261373_An_Approach_for_Detection_of_Advanced_Persistent_Threat_Attacks 2020.

[48] Surange G, Khatri P. Integrated intelligent IOT forensic framework for data acquisition through open-source tools. Int. j. inf. Tecnol. https://doi.org/10.1007/s41870-022-01025-5 2022.

[49] Kataria S, Batra U. Co-clustering neighborhood—based collaborative filtering framework using formal concept analysis. Int. j. inf. tecnol. 2022; 14: 1725–1731. https://doi.org/10.1007/s41870-022-00913-0

[50] Lekhraj, Kumar A, & Kumar A. An approach based on modified multiple attribute decision making for optimal node deployment in wireless sensor networks. Int. j. inf. tecnol. 2022; 14: 1805–1814. https://doi.org/10.1007/s41870-022-00919-8

[51] Sharma A, Mishra PK. Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis. Int. j. inf. tecnol. 2022; 14: 1949–1960. https://doi.org/10.1007/s41870-021-00671-5

[52] Song, Ch. A hybrid SEM and ANN approach to predict the individual cloud computing adoption based on the UTAUT2. Int. j. inf. tecnol. 2022. https://doi.org/10.1007/s41870-022-00936-7

[53] Itoo F, Meenakshi & Singh S. Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection. Int. j. inf. tecnol. 2021; 13: 1503–1511. https://doi.org/10.1007/s41870-020-00430-y

[54] Dymora P, Mazurek M. Anomaly Detection in IoT Communication Network Based on Spectral Analysis and Hurst Exponent. Appl. Sci. 2019; 9: 5319.