Evaluation of Machine Learning Models to Determine the Privacy Level of Various Cryptosystems

[1]Lavanya Modhugu, [2]Aruna Kumari Akula, [3]Praveen Kumar Gajjella, [4]Jogula Chandrika
[1,2,3]Assistant Professor, [4]UG Student, [1,2,3,4]Department of Computer Science Engineering, Rishi MS Institute of Engineering and Technlogy for Women, Kukatpally, Hyderabad.

ABSTRACT
Recent developments in multimedia technology have made the security of digital data a vital concern. To address the shortcomings of the current security mechanisms, researchers frequently concentrate their efforts on altering the existing protocols. However, during the past few decades, a number of proposed encryption algorithms have been shown to be insecure, posing a major security risk to sensitive data. It is crucial to use the best encryption method to defend against such attacks, but which algorithm is best in a certain situation will depend on the type of data being secured. However, evaluating various cryptosystems one at a time to determine the optimal choice can consume a significant amount of processing time. We provide a security level detection approach for image encryption algorithms by incorporating a support vector machine (SVM) for quick and precise selection of relevant encryption techniques. In this work, we also create a data set using standard encryption security parameters, such as entropy, contrast, homogeneity, peak signal to noise ratio, mean square error, energy, and correlation. These parameters are taken as features extracted from different cipher images. Dataset labels are divided into three categories based on their security level: strong, acceptable, and weak. To evaluate the performance of our proposed model, we have calculated accuracy and our results demonstrate the effectiveness of this SVM-supported system.
Keywords: DDoS attack, machine learning, Deep learning, Volumetric attacks, protocol attacks.

INTRODUCTION
Due to the exponential increase in multimedia data transmissions through unsecure networks, security has become a prominent study area (most notably the Internet). To shield data from snoopers and unauthorized users, several researchers have turned to creating novel encryption techniques. When encrypting digital photos, diffusion and misunderstanding are two essential components (also known as scrambling). According to a hypothesis put forth by Claud Shannon, a cryptosystem with confusion and diffusion techniques can be regarded as secure. On digital photos, the scrambling process can be applied directly to the pixels or to the rows and columns, whereas diffusion modifies the original pixel values. In other words, the replacement process replaces each distinct pixel value with the value of the S-unique box. However, data transmission in an encrypted format is insufficient to preserve its privacy. Although the information which is to be transmitted is in encrypted form, it can still be visualized by unauthorized users due to the weak security of the encryption algorithm. The security level of the encryption algorithm used to encrypt the image has a significant impact on its robustness. The plain image will be entirely encrypted using a highly strong encryption method, allowing it to withstand attacks on its integrity, secrecy, and availability. Along with security, temporal complexity is another key element to consider when choosing an encryption technology. Because different types of data have different security priorities, choosing a cryptosystem is dependent on the nature of the

application to be encrypted. As the image encryption algorithm is very important, we propose a security level detection approach for image encryption algorithms by incorporating a support vector machine (SVM).

We have categorized the security of encryption algorithms into three different levels (strong, moderate and weak) based on standard security parameters of the encryption algorithms. Below is the detail of how we divided the encryption algorithms into three said security levels based on the security parameters such as entropy, homogeneity, contrast, correlation, energy, PSNR and MSE.

As we are targeting those encryption algorithms, which are used to encrypt the 8– bit images. For the 8–bit images, the maximum entropy cannot be exceeded by 8. Likewise, for the binary images, the maximum entropy that can be obtained is 2. So, in the case of 8-bit images, we have divided the whole entropy interval for 8-bit images into three intervals. The range of the whole interval is 0 to 8.The average entropy value of any plain image may vary from 7.600 to 7.700. Whereas, an enciphered image encrypted generated using a weak encryption algorithm such as a single Substitution-box (S-box) algorithm may produce the average entropy value between 7.9503 to 7.9799. While for an acceptable and strong encryption algorithm, the average entropy value may vary from 7.9800 to 7.9900 and 7.9901 to 8.000 respectively. Similarly, the values for other security parameters may vary accordingly.

We obtain the security parameter values for different enciphered images which are generated from different encryption algorithms. Weak and moderate encryption algorithms are not able to encrypt the images properly. The enciphered images encrypted with weak and moderate encryption algorithms.

For the security level detection, we have considered all types of image encryption algorithms whether it is based on the frequency domain, transform-based or chaotic maps based schemes The main objective of the proposed work is to find the security level of the encryption algorithms. To generate a dataset, we considered a bunch of enciphered images and extract the feature values of those images. The size of the dataset is not restricted; it can be of any size. Feature values for strong and acceptable security level must be properly mentioned in the dataset. Take entropy values as an example; for the entropy values, we have taken the step size of 0.0001. we have divided the entropy values into three said intervals. For strong security, there are one hundred values ranges from 7.9901 to 8.000. Likewise, for the acceptable security level, there are one hundred and two values ranges from 7.9900 to 7.9800. All the other values which are below 7.9800 will be for weak security status. Similarly, we have divided the other parameter values into three intervals by selecting an appropriate step size accordingly.

LITERATURE SURVEY

Automated detection and classification of cryptographic algorithms in binary programs through machine learning

Threats from the internet, particularly malicious software (i.e., malware) often use cryptographic algorithms to disguise their actions and even to take control of a victim's system (as in the case of ransom ware). Malware and other threats proliferate too quickly for the time-consuming traditional methods of binary analysis to be effective. By automating detection and classification of cryptographic algorithms, we can speed program analysis and more efficiently combat malware. This thesis will offer different ways for automatically discovering and classifying cryptographic algorithms in compiled

binary programmers using machine learning. While more research is needed to fully test these methods on real-world binary programmers, the findings in this paper imply that machine learning may be used to detect and identify cryptographic primitives in compiled code with success. These techniques are now being used to discover and categories cryptographic algorithms in small single-purpose programmers, and more work is being suggested to apply them to real-world situations.

Applications in Security and Evasions in Machine Learning

In recent years, Machine Learning (ML) has become an important part to yield security and privacy in various applications. ML is used to address serious issues such as real-time attack detection, data leakage vulnerability assessments and many more. ML extensively supports the demanding requirements of the current scenario of security and privacy across a range of areas such as real-time decision-making, big data processing, reduced cycle time for learning, cost-efficiency and error-free processing. Therefore, in this paper, we review the state of the art approaches where ML is applicable more effectively to fulfill current real-world requirements in security. We examine different security applications' perspectives where ML models play an essential role and compare, with different possible dimensions, their accuracy results. By analyzing ML algorithms in security application it provides a blueprint for an interdisciplinary research area. Even with the use of current sophisticated technology and tools, attackers can evade the ML models by committing adversarial attacks. Therefore, requirements rise to assess the vulnerability in the ML models to cope up with the adversarial attacks at the time of development. Accordingly, as a supplement to this point, we also analyze the different types of adversarial attacks on the ML models. To give proper visualization of security properties, we have represented the threat model and defense strategies against adversarial attack methods. Moreover, we illustrate the adversarial attacks based on the attackers' knowledge about the model and addressed the point of the model at which possible attacks may be committed. Finally, we also investigate different types of properties of the adversarial attacks

Machine learning approaches to IOT security

With the continuous expansion and evolution of IoT applications, attacks on those IoT applications continue to grow rapidly. In this systematic literature review (SLR) paper, our goal is to provide a research asset to researchers on recent research trends in IoT security. As the main driver of our SLR paper, we proposed six research questions related to IoT security and machine learning. This extensive literature survey on the most recent publications in IoT security identified a few key research trends that will drive future research in this field. With the rapid growth of large scale IoT attacks, it is important to develop models that can integrate state of the art techniques and technologies from big data and machine learning. Accuracy and efficiency are key quality factors in finding the best algorithms and models to detect IoT attacks in real or near real-time.

Secure, privacy-preserving and federated machine learning in medical imaging

The broad application of artificial intelligence techniques in medicine is currently hindered by limited dataset availability for algorithm training and

validation, due to the absence of standardized electronic medical records, and strict legal and ethical requirements to protect patient privacy. In medical imaging, harmonized data exchange formats such as Digital Imaging and Communication in Medicine and electronic data storage are the standard, partially addressing the first issue, but the requirements for privacy preservation are equally strict. To prevent patient privacy compromise while promoting scientific research on large datasets that aims to improve patient care, the implementation of technical solutions to simultaneously address the demands for data protection and utilization is mandatory. Here we present an overview of current and next-generation methods for federated, secure and privacy-preserving artificial intelligence with a focus on medical imaging applications, alongside potential attack vectors and future prospects in medical imaging and beyond.

Machine Learning and Cryptographic Algorithms –Analysis and Design in Ransom ware and Vulnerabilities Detection

The AI, deep learning and machine learning algorithms are gaining the ground in every application domain of information technology including information security. Information security domain knows for traditional password management systems, autoprovisioning systems and user information management systems. There is another raising concern on the application and system level security with ransomware. On the existing systems cyber-attacks of Ransom ware asking for ransom increasing every day. Ransomware is the class of malware where the goal is to gain the data through encryption mechanism and render back with the ransom. The ransomware attacks are mainly on the vulnerable systems which are exposed to the network with weak security measures. With the help of machine learning algorithms, the pattern of the attacks can be analysed. Create or discuss a workaround solution of a machine learning model with combination of cryptographic algorithm which will enhance the effectiveness of the system response to the possible attacks. The other part of the problem, which is hard part to create intelligence for the organizations for preventing the ransomware attacks with the help of intelligent system password management and intelligent account provisioning. In this paper I elaborate on the machine learning algorithms analysis for the intelligent ransom ware detection problem, later part of this paper would be design of the algorithm.

Prediction of DDoS Attacks using Machine Learning and Deep Learning Algorithms

With the emergence of network-based computing technologies like Cloud Computing, Fog Computing and IoT (Internet of Things), the context of digitizing the confidential data over the network is being adopted by various organizations where the security of that sensitive data is considered as a major concern. Over a decade there is a massive growth in the usage of internet along with the technological advancements that demand the need for the development of efficient security algorithms that could withstand various patterns of the security breaches. The DDoS attack is the most significant network-based attack in the domain of computer security that disrupts the internet traffic of the target server. This study mainly focuses to identify the advancements and research gaps in the development of efficient security algorithms addressing DDoS attacks in various ubiquitous network environments.

Now a day's with the advent of 4G, 5G networks and economic smart devices there is a massive growth in the usage of the internet that has become a part of daily life. A vast range of services provided over the internet in diverse application areas such as business, entertainment, and education, etc. made it a vital component in framing various business models. This context made security over wireless networks as the most important factor while using the internet from unsecured connections[1]. Different security algorithms and frameworks are developed to enable protection from Internet-based attacks while devising high performance IDS (Intrusion detection systems) which act as a defensive wall while confronting the attacks over internet based devices. Distributed architecture based computing environments like cloud computing and IoT are more prone towards DDoS attacks in which multiple devices are coordinated to launch attacks over distributed targets. DDOS attacks are primarily launched in the context of exhausting the connectivity and the processing of the target server resources in which it enables the access constraints to the legitimate users to utilize the services provided by the target server that leads towards the partial unavailability or total unavailability of the services. The phenomenon of distributed computing is based on the one-to-many dimension in which these types of attacks may cause a possible amount of damage to the server resources [3]. It is observed from the previous research studies that the damage capacity, as well as the disrupting nature of the DDoS attacks, is gradually increased with the rate of internet usage.

EXISTINGSYSTEM

Substitution boxes (S-boxes) are the vector Boolean functions used commonly in cryptographic applications. A function of the form S : GFð2Þ n ! GFð2Þ m is called an n m S-box which takes n bits as the input and gives m bits as the output. If each output bit is called the n-variable Boolean function fi; then SðxÞ¼ðf1ðxÞ; ... ; fmðxÞÞ; where xGFð2Þ n [1]. S-box is the only nonlinear part of the block cipher and is a source to create confusion. There are many S-box construction methods available in the literature [2–5]. In cryptographic applications, the performance of different S-boxes can vary from one another and depends upon nature of data and their application. A major performing criterion of the S-box in encryption techniques is its non-linearity. A foremost research development in the past few years for the construction of S-boxes has been done mainly to increase the non-linearity of these S-boxes [6–8]. However, in case of highly auto-correlated message data, the S-box exhibits poor substitution results despite its high non-linearity. Chaotic dynamics are the behavior exhibit by some nonlinear dynamical system and can be used as a source of diffusion in substitution techniques. It has been observed by many researchers that there exists the close relationship between chaos and cryptography; many properties of chaotic systems have their corresponding in traditional cryptosystems. Chaotic systems have several compelling features favorable to secure communications, such as sensitivity to initial condition, ergodicity, control parameters and random like behavior, which can be correlated with some conventional cryptographic properties of good ciphers, such as confusion and diffusion proposed by Shanon. In the proposed substitution algorithm, chaos is being used with the S-box to strengthen the projected algorithm via applying the Shannon idea of sequential application by combining the confusion and diffusion properties. Most of the number theory or algebraic concepts based traditional ciphers such as Advanced Encryption Standard (AES), Data Encryption Standard (DES) and so on not appear to be ideal for multimedia applications due to certain and justified

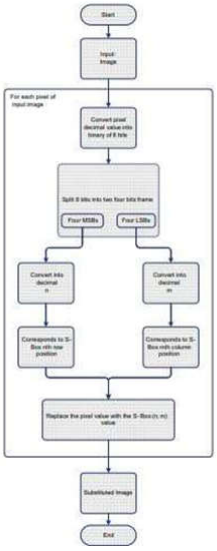

Fig 1:Substitution Algorithm for Two Dimensional Digital Image using Substitution Box

PROPOSED SYSTEM

The main objective of the proposed work is to find the security level of the encryption algorithms. To generate a dataset, we considered a bunch of enciphered images and extract the feature values of those images. The size of the dataset is not restricted, it can be of any size. Feature values for strong and acceptable security level must be properly mentioned in the dataset. Take entropy values as an example; for the entropy values, we have taken the step size of 0.0001. we have divided the entropy values into three said intervals. For strong security, there are one hundred values ranges from 7.9901 to 8.000. Likewise, for the acceptable security level, there are one hundred and two values ranges from 7.9900 to 7.9800. All the other values which are below 7.9800 will be for weak security status. Similarly, we have divided the other parameter values into three intervals by selecting an appropriate step size accordingly. For the visualization of the dataset, some portion of the proposed dataset is shown in Table 2 in which the first twenty feature vectors of each category of security level are displayed

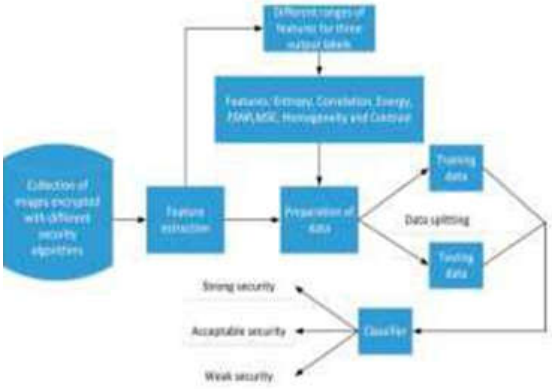

Fig 2: Block Diagram

WORKING

For the classification of each category, the decision will be based on the values of the security parameter. We have divided the range of each of the parameters into three intervals defined for weak, acceptable and strong security. For the weak security level, below 50 percent feature values must lie in the acceptable interval values. For acceptable security, atleast 65 percent feature values must lie in the acceptable interval values. encryption algorithms are taken as features, while three different levels of security – ‘‘strong,’’ ‘‘acceptable,’’ and ‘‘weak’’ . We have developed a new model using a support vector machine (SVM) to identify the security level of various cryptosystems.

S.No	Entropy	Energy	Contrast	Correlation	Homogeneity	MSE	PSNR	Security Status
1	7.926705	0.004527	10.197	-0.500593	0.0268	102.155	28.037	Strong
2	7.929994	0.004613	10.197	-0.500017	0.0214	101.297	28.074	Strong
3	7.9887	0.0157	10.18	0.00023	0.4034	108	11.5	Acceptable
4	7.9836	0.01575	10.179	0.00024	0.4035	107	11.6	Acceptable
5	7.9799	0.101	9.74	0.0012	0.4122	20	20.3	Weak
6	7.9797	1.1202	9.73	0.0014	0.4134	18	20.3	weak

Table 1: Data set for the proposed model

STATISTICAL ANALYSIS OF THE PROPOSED MODEL

CONFUSION MATRIX

The confusion matrix is a two-dimensional array that can be utilized to find accuracy, recall. and precision. The generalized confusion matrix for our proposed model.

1. TRUE POSITIVES When the system predicts ‘‘strong security’’ while the real output was also ‘‘strong security’’.
2. TRUE NEGATIVES When the system predicts ‘‘acceptable security’’ while the real output was also ‘‘acceptable security’’.
3. FALSE POSITIVES When the system predicts ‘‘strong security’’ while the real output was ‘‘acceptable or weak security’’.
4. FALSE NEGATIVES When the system predicts ‘‘acceptable security’’ or ‘‘weak security’’ while the real output was ‘‘strong security’’.

$$\text{Accuracy} = \frac{\text{Addition of all the values of first diagonal}}{\text{Total number of samples}}$$

According to table the percentage of accuracy from the proposed work will be:

Percentage Accuracy= (21+21+56/21+21+56+1+1) x 100%

Percentage Accuracy = 98%

CLASSIFICATION ACCURACY

The accuracy of this system reveals the information about how many correct predictions have been made by the model. The more correct predictions made, the higher the resulting accuracy. This classification accuracy can be measured as:

Classification Accuracy= (no of correct predictions/total number of predictions)

percentage Classification accuracy= (21+21+56/21+21+56+1+1) x 100%

percentage Classification accuracy = 98%
percentage classification accuracy= (TP+TN)/Total samples) x 100%
proposed work, the percentage of classification accuracy will be:

$$= \frac{T.P+(T.N)(1)+(T.N)(2)}{\text{Total Samples}} \times 100\%$$

Percentage of classification Accuracy= (21+21+56/21+21+56+1+1) x 100%
Percentage of Classification accuracy = 98%
PRECISION AND RECALL
Precision is the ratio between the true positive predicted observations and the total number of positive predicted observations. Mathematically, this can be expressed as:

$$\text{Precision} = \frac{T.P}{T.P+F.P}$$

In the case of our proposed work, the precision will be:

$$\text{Precision} - \frac{T.P}{T.P + (F.P)(1) + (F.P)(2)}$$

According to the values given in Table 5, the precision value for our proposed model
Will be precision= (21/(21+0+0)=1)
Recall= (T.P/(T.P+F.N)) (16) in the case of our proposed work, the equation 16 can be written as:

$$\text{Recall}= \frac{T.P}{T.P + (F.N)(1) + (F.N)(2) + (F.N)(3) + (F.N)(4)}$$

According to the values given in Table 5, the recall value for our proposed model will
F1 SCORE
Accuracy and F1 score both are important metrics when evaluating the performance of machine learning models. Accuracy is important when true positive and true negative samples are more valuable, while the F1 score is important when false positive and false negative samples are more important. F1 score can be calculated as:

$$\text{F1 Score} = \left[\frac{(\text{Recall})^{-1}+(\text{Precision})^{-1}}{2}\right]^{-1}$$

$$= 2\left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\right)$$

When the proposed model is tested on a 20% sample of the total data, the F1 score calculated (using equation 18) will be:

$$\text{F1 Score} = 2 \times \frac{1 \times 0.91}{1+0.91} = 0.94$$

SIMULATION RESULTS

As shown in below Fig,Describes the website of our project by running the code without errors,website link will be generate and then click on it.Then the website will be open in that mainly 3 Icons will be there.First one is Home Section,we want to return the page click that icon,the second one is the user we have to create a USER ID in that it will be possible next in third icon which photo we have to encrypt select the file from our system.

Fig: This page describes the website of the project.

As shown in below Fig, after we selecting the image then after that click on the upload icon and after that below the picture our algorithms will be there. They are DNA Encoding, Logistic Map, Rubik's Cube and Lorenz. Then select any one of the algorithm. And upload the image.

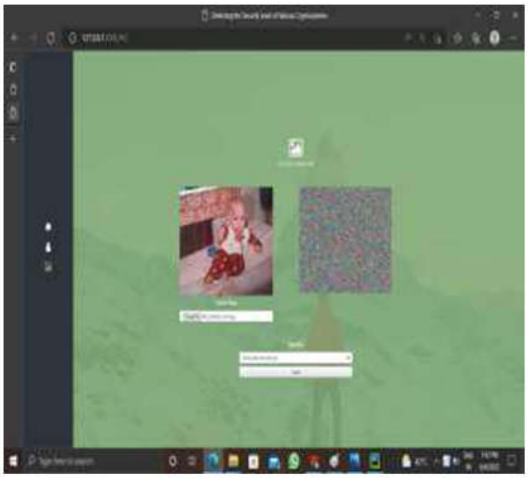

Fig: This page describes the inserting a picture.

As shown in below Fig, The final output will be displayed the encrypted image will be obtain and the accuracy of the image also displayed in the webpage.

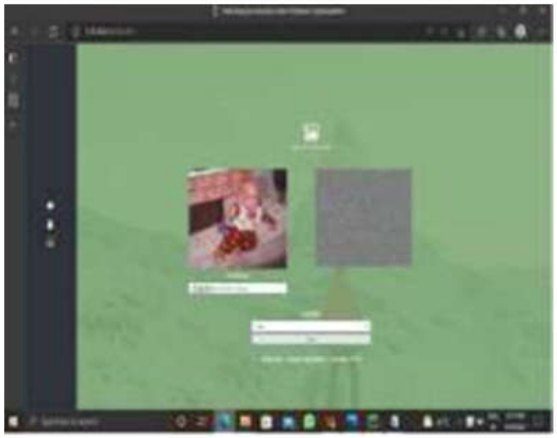

Fig: This page describes the accuracy of the model.

CONCLUSION

We have created and put out a model that can rapidly and accurately determine the security level of different encryption algorithms. We started by building a dataset and adding characteristics that represented the security parameters shared by different encryption techniques. We have separated the values of all attributes into three intervals—strong, acceptable, and weak—that correspond to the resulting security levels in order to generate a dataset. The various encryption techniques are then evaluated on our suggested model to determine the level of security that each one delivers. By calculating the statistical statistics of each one, we can manually determine the security level of these encryption techniques. With traditional testing methods, this process takes a great deal of time to accomplish but with our proposed model, testing can be achieved within a few seconds. To conclude, we also tested our proposed model using different experiments to evaluate its performance, and we found that it produces 94% correct predictions at much faster speeds than other models currently available

REFERENCES

1. Arslan shafique, Jameel ahmedi, Wadii boulila, Hamzah ghandora, Jawad ahmad and Mujeeb ur rehman.‘‘Detecting the security level of variuos cryptosystem using machine learning models”.
2. Hussain, A. Anees, A. H. Alkhaldi, M. Aslam, N. Siddiqui, and R. Ahmed, ‘‘Image encryption based on Chebyshev chaotic map and S8 S-boxes”.
3. Anees, I. Hussain, A. Algarni, and M. Aslam, ‘‘A robust watermarking scheme for online multimedia copyright protection using new chaotic map’’.
4. A. Shafique and J. Ahmed, ‘‘Dynamic substitution based encryption algorithm for highly correlated data”.
5. F. Ahmed, A. Anees, V. U. Abbas, and M. Y. Siyal, ‘‘A noisy channel tolerant image encryption scheme’’.
6. M. A. B. Farah, R. Guesmi, A. Kachouri, and M. Samet, ‘‘A novel chaos based optical image encryption using fractional Fourier transform and DNA sequence operation”.

7. C. E. Shannon, ‘‘Communication in the presence of noise”. S. Heron, ‘‘Advanced encryption standard (AES)’’.
8. H. Liu, A. Kadir, and X. Sun, ‘‘Chaos-based fast colour image encryption scheme with true random number keys from environmental noise’’.
9. Y.-L. Lee and W.-H. Tsai, ‘‘A new secure image transmission technique via secret-fragment-visible mosaic images by nearly reversible color transformations’’.
10. Anees, A. M. Siddiqui, and F. Ahmed, ‘‘Chaotic substitution for highly autocorrelated data in encryption algorithm’’.
11. L. Liu, Y. Lei, and D. Wang, ‘‘A fast chaotic image encryption scheme with simultaneous permutation-diffusion operation’’.
12. M. Khalili and D. Asatryan, ‘‘Colour spaces effects on improved discrete wavelet transform-based digital image watermarking using Arnold transform map”.