ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

Caeser Cipher Cryptography in Data Security

S. Sundara Mohan[1], Konakalla Khasyap Surya Saketh[2], aasrithaaasritha09@gmail.com[3], Ganesh Gowtham[4].

[1] Assistant Professor, Department of Computer Science Engineering, Chalapathi Institute of Engineering and Technology, Chalapathi Rd, Nagar, Lam, Guntur, Andhra Pradesh- 522034

[2,3,4,] Students, Department of Computer Science Engineering, Chalapathi Institute of Engineering and Technology, Chalapathi Rd, Nagar, Lam, Guntur, Andhra Pradesh- 522034

Email id: sundaramohans@gmail.com[1], kkssaketh@gmail.com[2], kokaakhil99@gmail.com[3], ganeshgowtham797@gmail.com [4]

Abstract

This paper explores the Caesar Cipher cryptography method as a foundational approach to data security, emphasizing its role in safeguarding information from unauthorized groups. The Caesar Cipher, a classic and straightforward encryption technique, substitutes each letter in the plaintext with another letter, shifted by a fixed number within the alphabet. Despite its simplicity, the Caesar Cipher can effectively protect data integrity by transforming sensitive information without altering the plaintext structure. The study demonstrates how Caesar Cipher encryption can contribute to data protection in secure communications and data recovery, highlighting its adaptability as a stepping stone for more complex encryption mechanisms in information security.

Keywords: Caesar Cipher, Cryptography, Data security, Information protection

1.Introduction

In the ever-evolving landscape of digital communication, ensuring data security has become a paramount concern. Cryptography, the science of encoding and decoding information, plays a vital role in safeguarding sensitive data. Among the earliest and most well-known cryptographic techniques is the Caesar Cipher, named after Julius Caesar, who is believed to have used this method to protect his military messages.

The Caesar Cipher is a substitution cipher that shifts each letter in the plaintext by a fixed number of positions down the alphabet. For example, with a shift of 3, 'A' becomes 'D', 'B' becomes 'E', and so on. This simple method transforms readable data into an unintelligible format, thus maintaining confidentiality during transmission.

While the Caesar Cipher is relatively easy to break by modern standards, it laid the foundation for more complex encryption algorithms used today. It illustrates the basic principles of data obfuscation and symmetric key encryption, where both the sender and the receiver must know the encryption key (i.e., the shift value). In data security, even such rudimentary encryption techniques can be useful for educational purposes and understanding the fundamentals of more advanced cryptographic systems. The Caesar Cipher continues to be a valuable tool in teaching the core concepts of encryption, key management, and data protection.

2. Related work:

Cryptanalysis refers to techniques for breaking into cryptographic security systems and gaining access to the contents of encrypted messages, even when the cryptographic key is unknown by Kirchhoff's principle, which is the paramount principle in modern cryptosystems [1]. It attempts to discover a meaningful pattern inside a given ciphertext to recover the corresponding plaintext or key. Such cryptanalysis techniques have been studied and gradually improved with the evolution of cryptosystems over past decades. This cryptanalysis technology can be applied to encryption for audio encryption and face recognition [2] used in real life.

Recently, a new class of artificial intelligence (AI)-based cryptographic attacks on digital cryptosystems have been proposed to gain efficient, meaningful cipher cracking and classification results [3]. The AI algorithms based on deep neural networks have led to huge improvements in computer vision medical image processing machine translation and the generation of virtual data [4] that are almost identical to the real data, and the attacks on cyber information security over the past decade [5]. More recently, generative adversarial networks (GANs) have produced great results for image generation, translation, resolution enhancement, and synthesis [6]. These advances in GAN models can even generate realistic face images from virtual people In addition, few studies break and emulate ciphertexts using GANs [7]. Reference [8] demonstrated that CipherGAN was capable of providing the underlying cipher mapping between unpaired ciphertext and plaintext for automated cryptanalysis without any prior knowledge, such as the character frequency distribution observed in natural language. This new class of AI-based cryptanalysis has been proposed to gain efficient and meaningful cipher cracking results. Especially, generative adversarial networks (GANs) have produced great results for image generation and translation. These advances in GANs models can even generate realistic data. A language model must represent both the feature distributions at sequential data point, and the possibly intricate temporal dynamics of those variables [9]. In particular, we want to properly represent the conditional distribution of temporal transitions in multivariate sequential data. The recently introduced StarGAN is an efficient method for multi-domain image-to-image translation, which takes in as input images in different domains and learns to flexibly translate the input image into the output image in the target domain, instead of learning a fixed translation, such as black-to-blond hair [10]. As a result, Star GAN can learn image-to-image translations with a single learning model that covers multiple domains.

3.Methodology

To propose a UC−GAN network model for multi-domain cryptanalysis that is based on Star GAN using a single unified generator and discriminator model. In our model, as in Cipher GAN, our objective is to train the generator on multiple cipher and plain domains without any prior knowledge. However, the two models differ in the number of generators. Figure shows the process of UC-GAN and Cipher GAN generating plaintext from the ciphertext. As the Cipher GAN model is based on the Cycle GAN model, each generator must be used for each ciphertext domain. However, the proposed UC-GAN model uses only a single unified generator for multiple ciphertext domains based on the Star GAN model.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

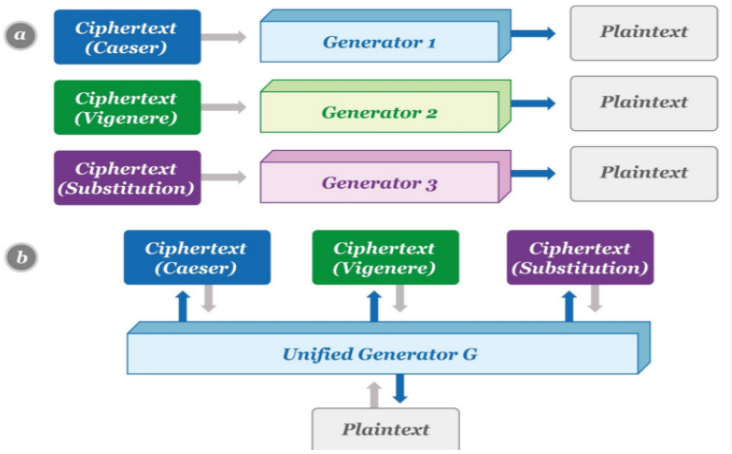

Figure: The process of generating the plaintext from the ciphertext with (**a**) the Cipher GAN model; (**b**) the proposed UC-GAN model.

Furthermore, we suggest an adversarial loss that supports the conversion of all source domain data into the target single data (multi-cipher to single-plain) or (single-plain to multi-cipher) with only one training process. Similar to Cipher GAN, the goal of our approach is to train the generator G on various cipher and plain domains with no previous information. The fundamental distinction is that we employ a single unified generator (G) and discriminator (D) across several cipher domains. The primary distinction between our model and Cipher GAN is that the Cipher GAN requires a ratio of the number of G and D to the number of ciphertext/plaintext domains. Therefore, our method is greatly improved in representing multi-domain on discrete random variables. The experiment results show that our proposed model can break and emulate multiple substitution ciphers (Caesar, Vigenere, and substitution cipher) in only one training process.

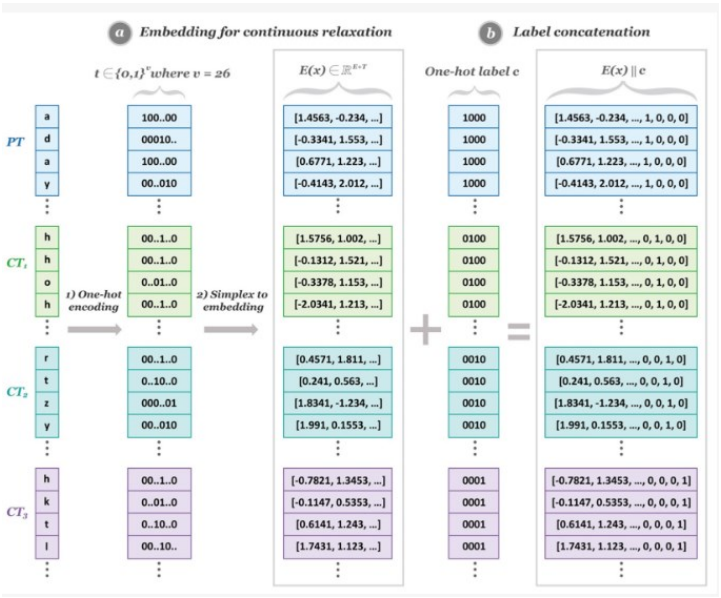

Figure: The process of creating data used for model training. (**a**) Embedding for continuous relaxation; (**b**) label concatenation. Finally, concatenated $E(x)\|cE(x)\|c$ is an input for the unified generator G.

The process of creating data used for model training. To build UC−GAN, we use embedding E for continuous relaxations before the main process as embedding space $Wemb$Wemb and timing space $Wtime$Wtime. The embedding E consists of a one-hot vector step and a simplex to embedding step. In the one-hot vector step, each character is represented by a one-hot vector with the length of v ($v = 26$ in our setting). To make embedding vector $E(x)$E(x), it computes $r=t\cdot Wemb$r=t·Wemb and $E(x)=r||Wtime$E(x)=r||Wtime, where the original one-hot vector tt from the original data xx, with embedding space $Wemb$Wemb and additional timing space $Wtime$Wtime. Such continuous relaxation makes it possible to back-propagate when we are training neural network models with preserved information. In our setting, we assume that the original data x consists of the number of N characters.Therefore, $t\in\{0,1\}N\times vt\in\{0,1\}$N×vt∈{0,1}N×v and $WEmb\in\mathbb{R}v\times E$WEmb∈ℝv×E and $W time\in\mathbb{R}N\times T$Wtime∈ℝN×T are trained parameters which we have to update when the training is processed. Finally, the input data of the model are created by concatenating the embedding data and the target domain label.

Datasets

a collection of digitized text samples in American English, for our experiments. There are four domains in the training dataset and the test dataset plaintext (PT), Caesar cipher (CT1), Vigenere cipher (CT2), and substitution cipher (CT3). The dataset is encrypted with Caesar, Vigenere, and Substitution ciphersFurthermore, each data row in our dataset consists of N = 100 characters. As a result, we can extract 4,537,600 characters. For the training dataset, each domain has a number of 9600 data rows to consider both the known-plaintext attack (KPA) scenario and the ciphertext-only attack (COA) scenario. In KPA settings, the attacker can access the encryption method, which means the attacker has the number of n pairs. Otherwise, in COA settings, the attacker has only ciphertexts, which means they have the number of n ciphertexts. We extract 9600 data rows for each 4 domains to show accurate unsupervised learning results. Therefore, the training dataset has 9600 × 4 × N(100) = 3,840,000 characters.

4. Results

As described above, we demonstrate that our UC-GAN can successfully break the Caesar, Vigenère, and Substitution ciphers using only one unified generator. In addition, our unified method can learn these discrete distributions for all multi-ciphers-to-plain domains Our encryption emulation can reconstruct all types of ciphertexts from a single plaintext, such as $PT\rightarrow CT_1$, $PT\rightarrow CT_2$, and $PT\rightarrow CT_3$ for the plain-to-multi-ciphers domain In addition, the proposed model can emulate all types of ciphertext to a single plaintext for multi-cipher-to-plain. We measured the accuracy of the model using test data per each epoch. To test the model, the test data was used as the input of the model, and one generator generates three ciphertexts according to the labeled target. The model accuracy was calculated by comparing the generated ciphertext with the target ciphertext. However, the convergence speed was different for different ciphers and Caesar and Substitution were broken faster than Vigenère. the results of the ciphertext generated by the model from the plaintext and the target ciphertext created by

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

the three cipher methods. Texts that do not match between the target ciphertext and the generated ciphertext are marked in red, showing that most of the texts match

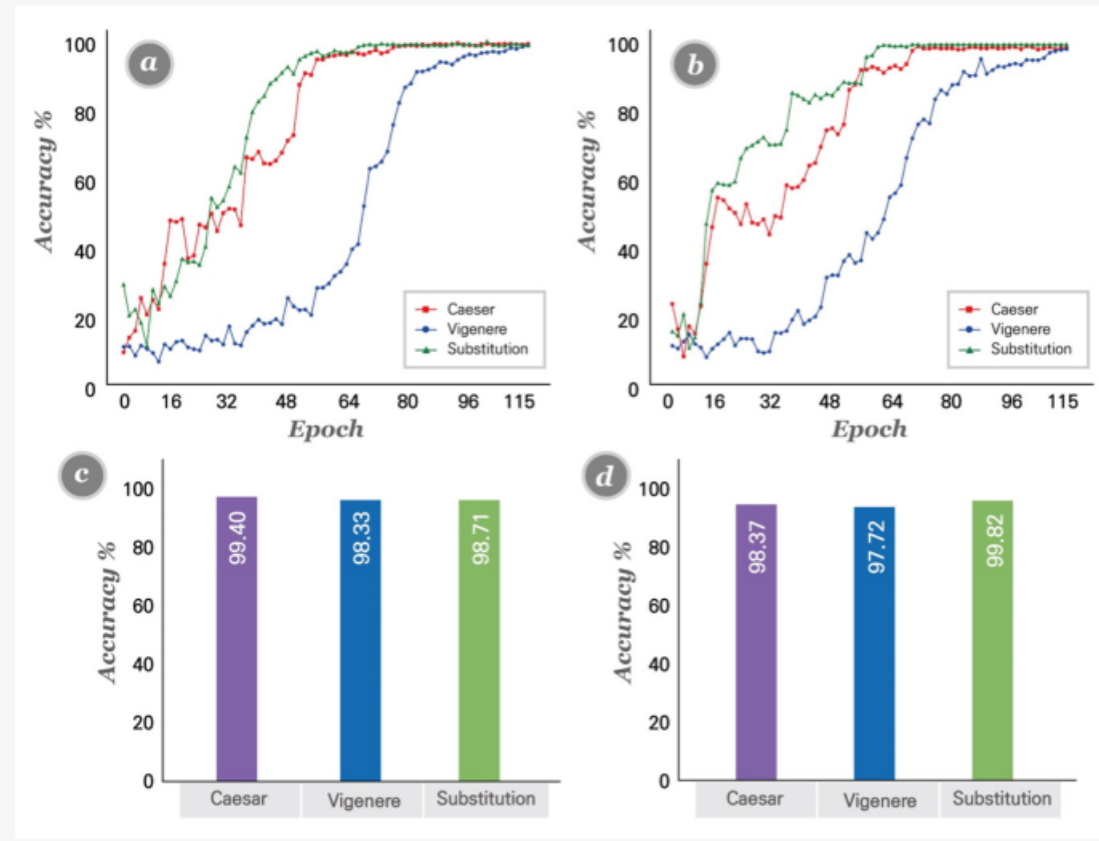

Figure: The proposed method tests the process and accuracy results. (**a**) proposed method test curves for multi-cipher-to-single plain recovery; (**b**) proposed method test curves for single-plain-to-multi-cipher recovery; (**c**) multi-cipher-to-single-plain recovery test results; (**d**) single-plain-to-multi-cipher recovery test results.

Table: Hyper parmeter tuning experiments.

Parameter	Default Setting	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Learning rate (*lr*)	1.8×10^{-4}	1.8×10^{-4}	1.8×10^{-4}	1.8×10^{-4}	1.8×10^{-4}
Batch size (*bs*)	32	8	128	32	32
Embedding space W_{emb} and W_{ttme}	256	256	256	128	512
Lambda for classification loss function (λ_{cls})	1	1	1	1	1
Lambda for reconstruction function (λ_{rec})	10	10	10	10	10

In the first round of the experiment, we investigated the impact of batch size on the different cipher attacks. First, we set the batch size to 8 and ran the training process for both ciphers to plain and plain to cipher attacks. and b demonstrate the network training process on a cipher to plain and plain to cipher, respectively.

Discussion

We suggested a novel unsupervised deep-learning model that is more flexible and efficient than existing information extraction techniques for translating ciphertext-to-plaintext across various cipher domains. The proposed model is based on generative adversarial networks (GANs). As described above, by competing for a generative deep neural network against a discriminative deep neural network, the GAN model creates samples that seem to come from the training set. Especially, Cipher GAN is a GAN-based model that is an unsupervised cryptanalytic method to break substitution ciphers without any prior knowledge. However, Cipher GAN requires the number of generators and discriminators to be in proportion to the number of ciphertext/plaintext domains. Unlike Cipher GAN, we proposed UC−GAN, which consists of a unified generator and a discriminator using only a single deep neural network for multiple domains. The proposed UC−GAN model was able to perform extremely well on multiple substitution ciphers, achieving near-flawless accuracy. The experimental results were carried out using three types of classical ciphers Caesar, Vigenere, and Substitution ciphers. In addition, we compared the model performance of our model with that of the Cycle GAN model.

Conclusion

This paper examined the underlying privacy dangers that behavioural identity detection technologies in VR potentially pose. First, we demonstrated that even simple classification approaches could identify the participants with a high detection rate utilising behavioural features in VR. Following that, we investigated how reliably these trained classifiers can be used to identify the same person performing different tasks. The study then demonstrated how ineffective intentional user behaviour changes are to circumvent these classifiers. Then, we investigated the impact of the user's physical attributes on the classification of behavioural identity. Finally, we demonstrated the behavioural data types with the highest variance, which should be assigned a higher priority when creating behavioural privacy protection solutions such as behaviour filters. These findings highlight the importance of offering greater privacy protection tools for VR users to benefit both VR consumers and the VR industry. Also, while numerous models exist in the literature for detecting identity from user behaviour, our main objective in this study was not to reproduce and evaluate all of them. Instead, we focused on highlighting the potentially harmful capabilities of some of this field's most widely used models. Specifically, we sought to demonstrate the dangers these models pose to user privacy and underscore the importance of further research to mitigate these risks.

References:

1. Courtois, N.T.; Oprisanu, M.-B.; Schmeh, K. Linear cryptanalysis and block cipher design in East Germany in the 1970s. *Cryptologia* 2019, *43*, 2–22.
2. Al-Shabi, M. A survey on symmetric and asymmetric cryptography algorithms in information security. *Int. J. Sci. Res. Publ. (IJSRP)* 2019, *9*, 576–589.
3. Wu, R.; Gao, S.; Wang, X.; Liu, S.; Li, Q.; Erkan, U.; Tang, X. Aea-NCS: An audio encryption algorithm based on a nested chaotic system. *Chaos Solitons Fractals* 2022, *165*, 112770.

4. Gao, S.; Wu, R.; Wang, X.; Liu, J.; Li, Q.; Tang, X. EFR-CSTP: Encryption for face recognition based on the chaos and semi-tensor product theory. *Inf. Sci.* 2023, *621*, 766–781
5. Ahmadzadeh, E.; Kim, H.; Jeong, O.; Moon, I. A novel dynamic attack on classical ciphers using an attention-based LSTM encoder-decoder model. *IEEE Access* 2021, *9*, 60960–60970.
6. Ahmadzadeh, E.; Kim, H.; Jeong, O.; Kim, N.; Moon, I. A deep bidirectional LSTM-GRU network model for automated ciphertext classification. *IEEE Access* 2022, *10*, 3228–3237.
7. Chan, T.-H.; Jia, K.; Gao, S.; Lu, J.; Zeng, Z.; Ma, Y. PCANet: A simple deep learning baseline for image classification? *IEEE Trans. Image Process.* 2015, *24*, 5017–5032.
8. Ahmadzadeh, E.; Jaferzadeh, K.; Shin, S.; Moon, I. Automated single cardiomyocyte characterization by nucleus extraction from dynamic holographic images using a fully convolutional neural network. *Biomed. Opt. Express* 2020, *11*, 1501–1516.
9. Fomicheva, M.; Sun, S.; Yankovskaya, L.; Blain, F.; Guzmán, F.; Fishel, M.; Aletras, N.; Chaudhary, V.; Specia, L. Unsupervised quality estimation for neural machine translation. *Trans. Assoc. Comput. Linguist.* 2020, *8*, 539–555.
10. Sirichotedumrong, W.; Kiya, H. A gan-based image transformation scheme for privacy-preserving deep neural networks. In Proceedings of the 2020 28th European Signal Processing Conference (EUSIPCO), Virtual, 18–22 January 2021; pp. 745–749.