Great Place To Work. Certified MAR 2021–FEB 2022 INDIA

UPES
UNIVERSITY OF TOMORROW

MINOR PROJECT

Artificial Intelligence Based Security Countermeasures for Internet of Things Networks

Presented by:
Harshit Raheja, R2142201556, CSE AiMl (Hons.)
Jahanvi Arora, R2142201568, CSE CSF (Hons.)
Adarsh Singh, R2142201563, CSE CSF (Hons.)
Archit Nangla, R2142201943, CSE CSF (Non Hons.)

Guided by:
Dr. Sunil Gupta
Professor and Cluster Head
Department of Cybernatics
School of Computer Science, UPES

PROBLEM STATEMENT

A security model for the IoT sensors which will use various security algorithm and with the help of Machine Learning Algorithm it will decide which security algorithm is best suited for the particular attack and situation.

INTRODUCTION

Cryptography is a method to accomplish confidentiality of messages. In Symmetric key encryption the message is encrypted by utilizing a key and a similar key is utilized to decode the message which makes it simple to utilize yet less secure. Asymmetric Key Encryption depends on open and confidential key encryption procedures. It is safer than the symmetric key encryption method yet is a lot slower.

ABSTRACT

This study covers drone technology, application areas, citizen multi-objective uses, drone security, protection, and secrecy concerns. The security algorithm to run will be decided by a Machine Learning Algorithm. There are total of Three Hundred Forty - Eight security algorithms out of which we are going to use few of the security algorithms to make sure that our data is secure. Each algorithm has its own pros and cons for different attacks. The probability will be calculated from a Machine Learning model.

MOTIVATION

- Security is one of the main aspect in today's era. So, we decided to build a model which can provide security to different areas.
- As Artificial Intelligence is an emerging and trending area. Hence, we decided to use the Decision Making ML Algorithms which can be trained and tell us which security algorithm is best suited for which attack.
- To enhance the performance of our model we have used various algorithms of Cryptography and Hashing. We also have implemented some protocols.

OBJECTIVES

- The fundamental goal of our project is to carry out a security model with the help of different cryptographic algorithms and furthermore we will utilize different Machine Learning based Algorithms to conclude which Encryption Algorithm is best for a specific digital assault.
- Based on the Machine Learning algorithms we will provide security to IOT devices with the help of encryption algorithms.

Technology Stack

Platform Independent Software based on JAVA and can be executed on any system having JVM installed in it.

Development

Backend Language Used: Java
Backend Software Used: ecclipse and intellj

Dataset

Dummy Dataset

METHODOLOGY

RESEARCH

Detailed analysis of different data encryption technique through research papers.

REVIEW

Studying about the selected encryption technique their algorithm their pros and cons.

EVALUATING

Assessing few algorithms which are best suited for data encryption.

THREAT

Determining security threat for the drone and use encryption technique accordingly.

IMPLEMENTATION

Implementation of the selected security algorithm and using machine learning to determine which security algorithm is suited for cyber-attack.

MODEL

Working model of data encryption technique is made.

TRAINING

Using data sets to determine which algorithm is best suited for the encryption of the data by machine learning.

TESTING

Testing of encryption technique for data encryption on the model.

APPLICATIONS

Banking Sector

Networking Sector

Social Media

IOT Devices

IMPLEMENTATION

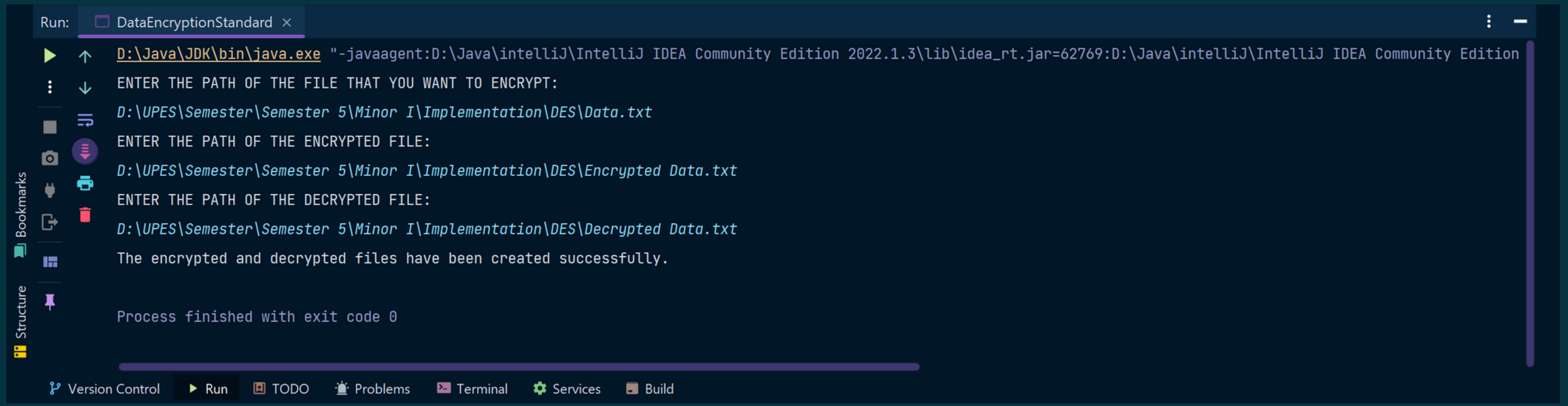
SECURITY ALGORITHMS

- Data Encryption Standard (DES)
- Triple Data Encryption Standard (3DES)
- Advance Encryption Standard (AES)
- Rivest-Shamir-Adleman (RSA)
- Elliptical curve cryptography (ECC)
- Blowfish
- TwoFish
- ElGamal

IMPLEMENTATION

SECURITY ALGORITHMS

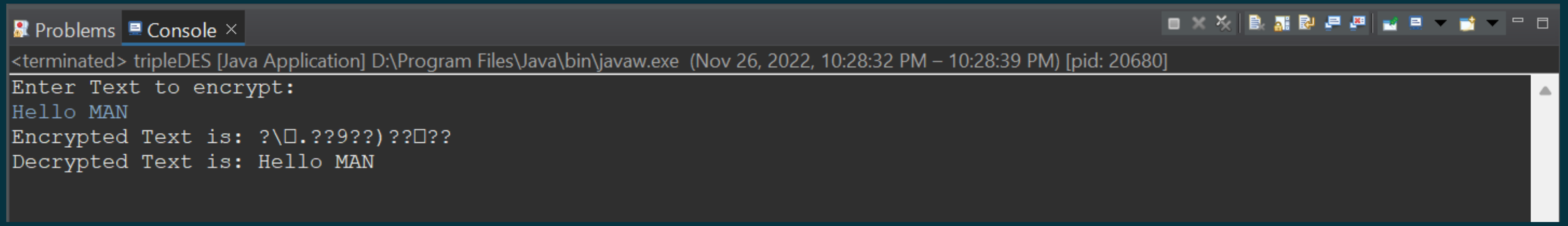
- Data Encryption Standard (DES): It is a block cipher used in the Data Encryption Standard (DES). It encrypts a plain text block of 64 bits with a key of 56 bits. It is made up of a fiestal network that divides a block into two equal halves, with the right half passing through a variety of functions. DES employs a series of S-boxes and P-boxes. The cypher text is obtained by the XOR operation after passing through these permutation and substitution boxes. DES employs 19 rounds.



IMPLEMENTATION

SECURITY ALGORITHMS

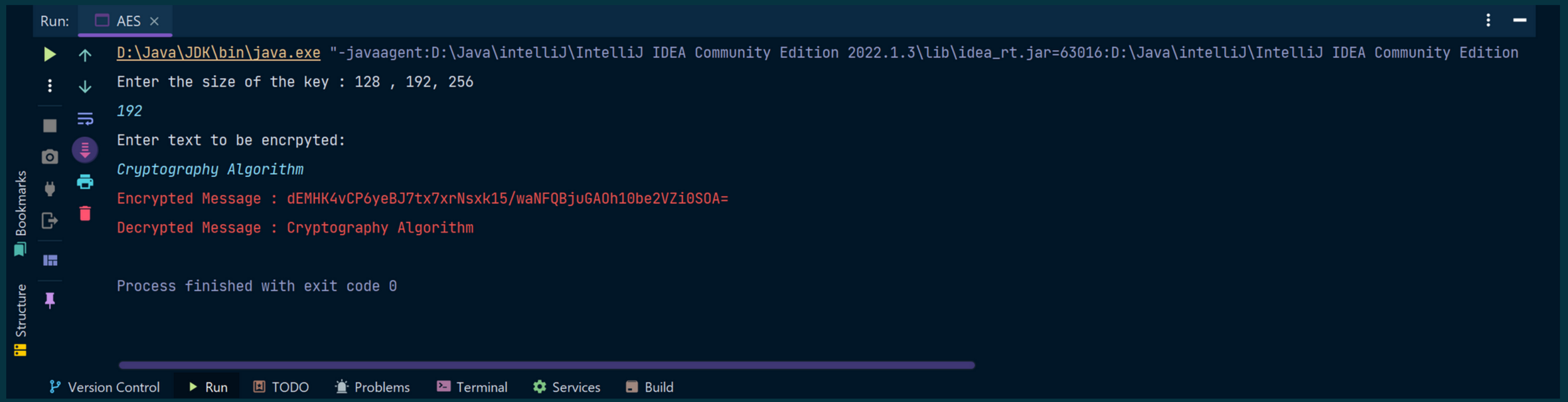
- Triple Data Encryption Standard (3DES): It is a symmetric-key encryption technique. It employs a block size of 64 bits and a key length of 56 bits to encrypt or decrypt any message or data. As the name implies, it applies the same DES algorithm to each data block three times.



IMPLEMENTATION

SECURITY ALGORITHMS

- Advance Encryption Standard (AES): It operates on blocks of three sizes: 128 bits, 192 bits and 256 bits. AES-128 employs 10 rounds, AES-192 employs 12 rounds, AES-256 employs 14 rounds to encrypt and decrypt the message. In each round different steps are there like: substitution byte, shift rows, mixed columns and add round key.



IMPLEMENTATION

SECURITY ALGORITHMS

- Rivest Shamir Adleman (RSA): This cryptographic algorithm is a widely-used method of public key encryption and was first described by Rivest, Shamir, and Adelman. It uses two keys: one key is used to encrypt the message and another key is used to decrypt the message.

Problems | Console

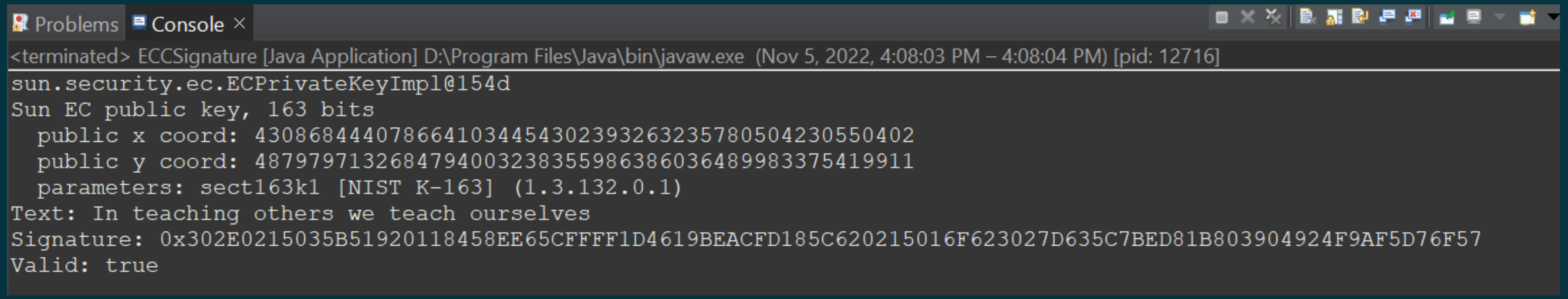
<terminated> rsa [Java Application] D:\Program Files\Java\bin\javaw.exe (Nov 5, 2022, 3:56:45 PM – 3:57:02 PM) [pid: 13436]

```
Enter Value of P = 11
Enter Value of Q = 13
Value of N = (P * Q) => 143
Value of Phi of N = (P-1) * (Q-1) => 120
value of E  = 7
Value of D = 103
        Your Public key = { 7 , 143 }
        Your Private key = { 103 , 143 }
Enter a Character = z
Plain Text = z
Cypher Text = 34
```

IMPLEMENTATION

SECURITY ALGORITHMS

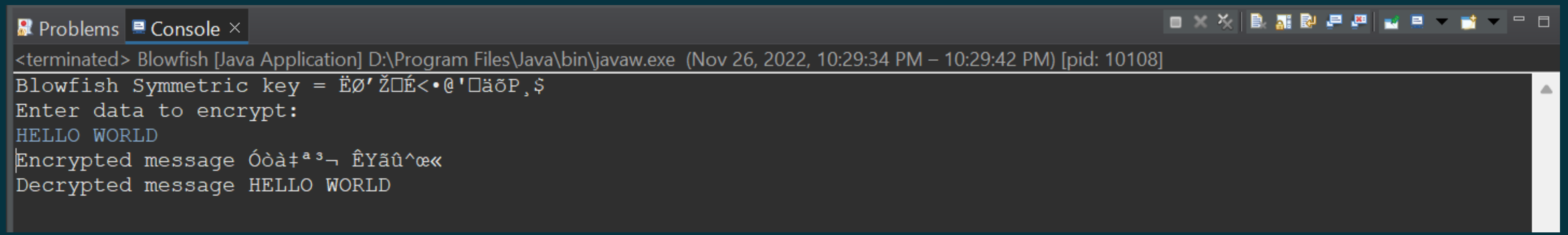
- Elliptic Curve Cryptography (ECC): This cryptographic algorithm is a widely used public key encryption system. Encryption is the process of transforming plaintext into ciphertext and it's often used to make sure that only intended recipients can access sensitive information. The ECC algorithm is based on two prime numbers a generator and a base that are multiplied together to form two pairs of numbers that are then multiplied together. The resulting product is called an Elliptic Curve.



IMPLEMENTATION

SECURITY ALGORITHMS

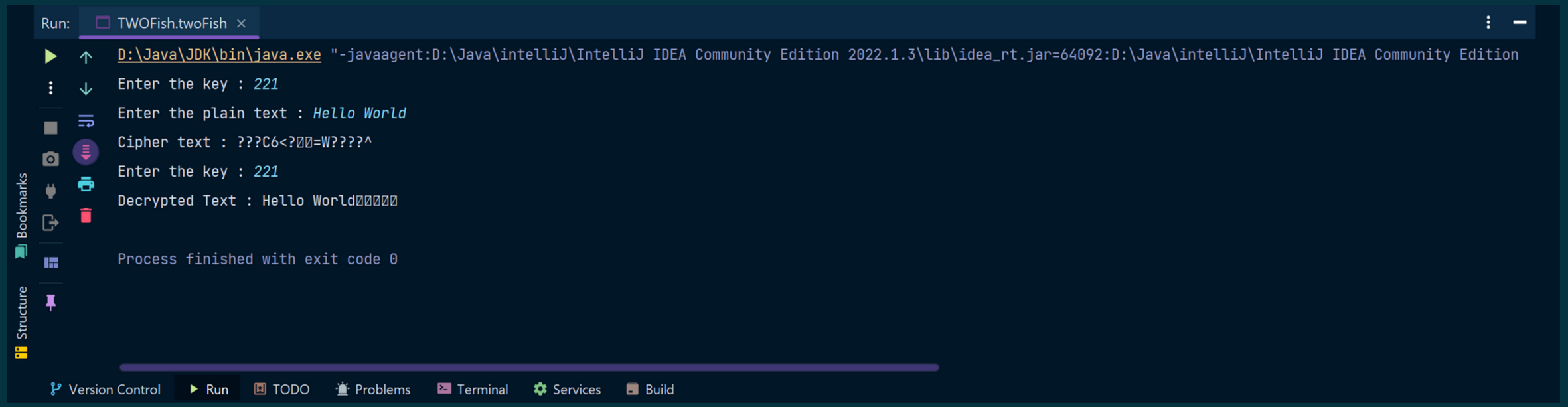
- BlowFish: It has a variable key length with a maximum of 448 bits. It has a 64-bit block size. The blowfish algorithm consists of two stages The first stage is the key expansion phase, which converts a 448-bit key into a number of sub keys bringing the total 4168 bytes. The second stage is the encryption phase, which involves iterating a function 16 times and obtaining the encrypted text via the XOR operation.



IMPLEMENTATION

SECURITY ALGORITHMS

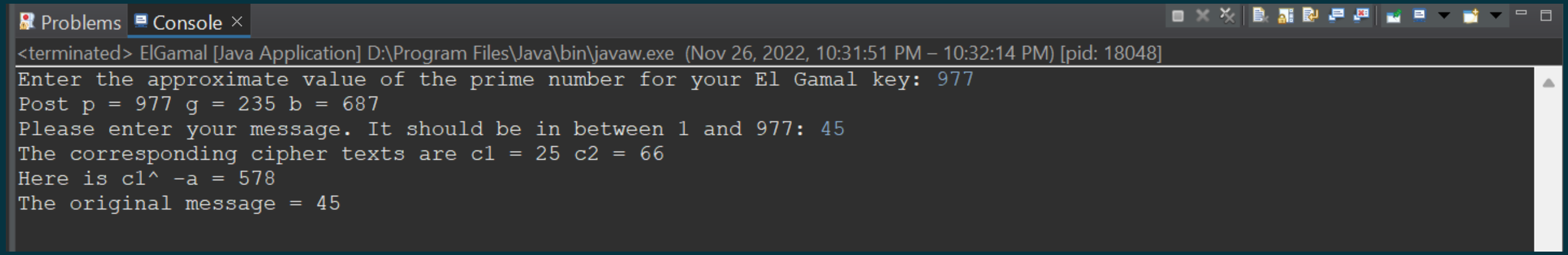
- TwoFish: This is a block-encrypted symmetric key cryptography. The size of block used in this cryptography is of 128 bits and the size of key is of variable length (128, 192 or 256 bits). It is Open Source (not licensed), patent free and freely available. 2 FISH is quite similar to older symmetric-key block cipher BLOWFISH. It also includes extended functionality to substitute the Data Encryption Standard (DES) algorithm.



IMPLEMENTATION

SECURITY ALGORITHMS

- ElGamal: It is a type of asymmetric cryptographic algorithm. The difficulty of using the cyclic group to find the discrete logarithm is its key concern. Even if the attacker knows the values of g^a and g^b, it will still be very challenging for him to determine the value of g^ab, which is simply the cracked value.



IMPLEMENTATION

HASHING ALGORITHMS

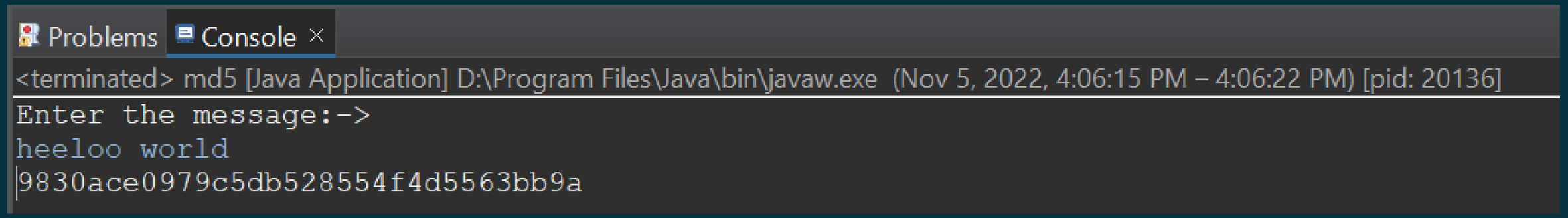
- Message-Digest Algorithm 5 (MD5)
- Secure Hash Algorithm 1 (SHA1)
- Secure Hash Algorithm 2 (SHA2)
- Secure Hash Algorithm 3 (SHA3)

IMPLEMENTATION

HASHING ALGORITHMS

- MD5: It is a widely used hash function producing a 128-bit hash value. Although it was initially designed to be used as a cryptographic hash function, it has been found to suffer from extensive vulnerabilities. It can still be used as a checksum to verify data integrity, but only against unintentional corruption.

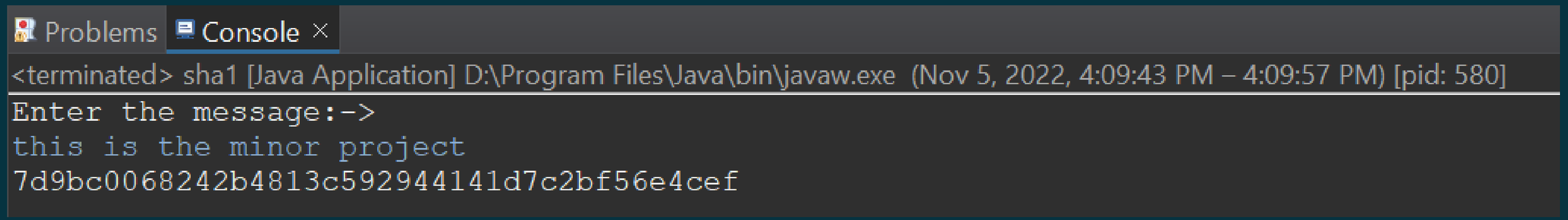
MD5 Function + Input String = 128-bit Message Digest



IMPLEMENTATION

HASHING ALGORITHMS

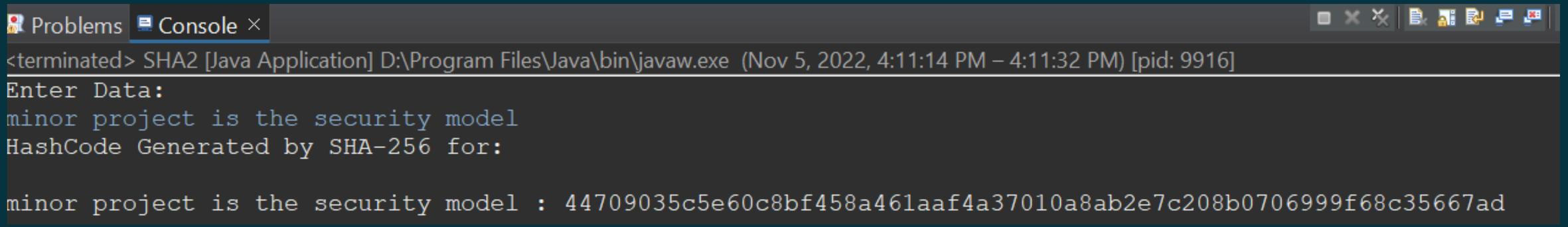
- SHA1: It is a cryptographic hash function which takes an input and produces a 160-bit hash value known as a message digest. It is also used to identify that during transmission of information from sender to receiver is any changes are occurring or not.



IMPLEMENTATION

HASHING ALGORITHMS

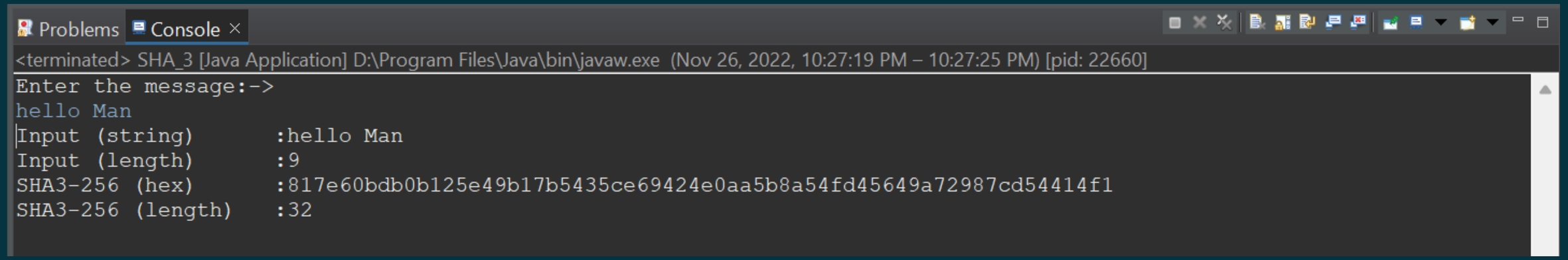
- SHA2: There are six distinct SHA-2 variations, which vary in direct proportion to the bit size being used to encrypt data. A 256-bit hash is produced by SHA-256, which also has a 512-bit block size. The initialization variables and constants are 32 bits long, and the message input is handled in 32-bit words.



IMPLEMENTATION

HASHING ALGORITHMS

- SHA3: The purpose of SHA-3 is that it can be directly substituted for SHA-2 in current applications if necessary, and to significantly improve the robustness of NIST's overall hash algorithm toolkit. To ensure the message can be evenly divided into r-bit blocks, padding is required.



IMPLEMENTATION

MACHINE LEARNING ALGORITHMS

- RANDOM FOREST ALGORITHM: We use random forest for forecasting data based on a large data collection and several decisions. An example of an ensemble is a random forest, which combines the results of various algorithms. To construct a random forecast, many decision trees are created. Each decision tree forecasts a value, and the average of the forecasted values is then calculated.

- ID3: The algorithm repeatedly divides characteristics into two or more groups at each step, hence the name "ID3" (Iterative Dichotomize 3). ID3, developed by Ross Quinlan, constructs a decision tree from the top down in a greedy manner. Simply said, the greedy technique means that we choose the best feature at the time of each iteration to produce a node, whereas the top-down approach indicates that we build the tree from the top down. ID3 is typically only applied to classification issues involving solely nominal features.

OUTPUT

ID3 DESICION TREE

Problems | Console ×

Main [Java Application] D:\Program Files\Java\bin\javaw.exe (Nov 26, 2022, 10:39:10 PM) [pid: 23104]

```
ï»¿Attacks=Ransomware
        Class: ELGAMAL
ï»¿Attacks=Trojans
        Class: BLOW_FISH
ï»¿Attacks=Viruses
        Class: AES
ï»¿Attacks=Malware Based
        Class: SHA3
ï»¿Attacks=Phishing
        Class: SHA1
ï»¿Attacks=man in the middle
        Class: RSA
ï»¿Attacks=password attacks
        Class: SHA3
ï»¿Attacks=SQL injections
        Class: DES
ï»¿Attacks=Dos
        Class: SHA2
ï»¿Attacks=Ddos
        Class: ELGAMAL
ï»¿Attacks=advanced persistent threat
        Class: SHA2
ï»¿Attacks=Watering Hole attacks
        Class: SHA1
ï»¿Attacks=Cross-site Scripting
        Class: AES
ï»¿Attacks=Cryptojacking
        Class: DES
ï»¿Attacks=URL manipulation
        Class: 3DES
```

Problems | Console ×

Main [Java Application] D:\Program Files\Java\bin\javaw.exe (Nov 26, 2022, 10:39:10 PM) [pid: 23104]

```
        Class: AES
ï»¿Attacks=Cryptojacking
        Class: DES
ï»¿Attacks=URL manipulation
        Class: 3DES
ï»¿Attacks=Zero-day exploits
        Class: DES
ï»¿Attacks=DNS-based
        Class: RSA
ï»¿Attacks=Rootkits
        Class: RSA
ï»¿Attacks=Session hijacking
        Class: RSA
ï»¿Attacks=Spear phishing attack
        Class: SHA3
ï»¿Attacks=angler phshing attacks
        Class: 3DES
ï»¿Attacks=Whaling attacks
        Class: ELGAMAL
ï»¿Attacks=Brute force attack
        Class: MD5
ï»¿Attacks=Dictionary attack
        Class: MD5
ï»¿Attacks=Password spraying
        Class: SHA1
ï»¿Attacks=DNS tunneling
        Class: SHA1
ï»¿Attacks=DNS spoofing
        Class: 3DES
```

COMPARISION TABLES

MACHINE LEARNING ALGORITHMS

Name of Algorithms	Theory	Mathematics	Advantages	Disadvantages	Time complexity
Logistic Regression	Finding the greatest fit line through data is the main goal of linear regression methods. When data is divided into groups, the linear regression algorithm is modified to forecast issues using logistic regression. In order to determine the likelihood that an event will occur, we utilize logistic regression. Due to the fact that it is a linear classification model, a relationship between the independent and dependent variables is discovered. Logistic functions are used to model the likelihood of outcomes.	$g(E(y)) = \alpha + \beta x1 + \gamma x2$ g() is the link function, E(y) is the expectation of target variable and $\alpha + \beta x1 + \gamma x2$ is the linear predictor (α,β,γ to be predicted). To "link" the expectation of y to the linear predictor is the function's primary function.	Fast to train and forecast. Good for small classification data problems. Easy to understand.	Not very accurate. Can't be used for nonlinear data. Not flexible to complex data. Model occasionally ends up overfitting.	The Logistic Regression's overall temporal complexity during training is n(O(d))=O(nd)
Decision Tree	A supervised machine learning technique called decision tree analysis can do classification or regression analysis. Decision trees produce outcomes that are highly interpretable and are simple to understand thanks to their graphical depiction. Predicting illness stage using clinical history, lab results, and biomarker levels, or predicting antibody concentration in response to vaccination based on patient and vaccine features, are a few instances pertinent to the subject of health.	Entropy is amount of information is needed to accurately describe some sample. $Entropy = -\sum_{i=1}^{n} p_i * \log(p_i)$ Gini index is measure of inequality in sample. It has value between 0 and 1. $Gini\,index = 1 - \sum_{i=1}^{n} p_i^2$	Decision trees take less work to prepare the data during pre-processing than other methods do. Data normalization is not necessary for a decision tree. Scaling of data is not necessary when using a decision tree. Additionally, the construction of a decision tree is not significantly impacted by missing values in the data. Technical teams and stakeholders can understand a decision tree model very quickly.	A slight change in the data can result in a big change in the decision tree's structure, which can lead to instability. When compared to other algorithms, a decision tree's calculations may become far more complicated. The model training process for decision trees typically takes longer. Because of its intricacy and lengthier training period, decision tree training is relatively expensive. Regression applications and continuous value predictions are insufficient for the Decision Tree algorithm.	The things we need while training a decision tree are the nodes which are typically stored as if-else conditions. Test time complexity would be O(n), where n is the depth. Since we have to move from root to a leaf node of the decision tree.
Id3	The algorithm iteratively (repeatedly) dichotomizes (divides) characteristics into two or more groups at each step, hence the name "ID3" (Iterative Dichotomize 3). ID3, developed by Ross Quinlan, constructs a decision tree from the top down in a greedy manner. Simply said,	Entropy is used by the ID3 algorithm to determine how homogeneous a sample is. Entropy E(s) = 0 denotes complete homogeneity or the leaf node of a tree, which precludes further division. ID3 splits the algorithm using the least entropy possible.	Understandable prediction rules are created from the training data. Builds the fastest tree. Builds a short tree. Only need to test enough attributes until all data is classified.	Data may be over fitted or over classified. Only one attribute at a time is tested for making a decision. Classifying continuous data may be computationally expensive.	Time complexity of id3 is O(y * n log(n))

COMPARISION TABLES

MACHINE LEARNING ALGORITHMS

	the greedy technique means that we choose the best feature at the time of each iteration to produce a node, whereas the top-down approach indicates that we build the tree from the top down. ID3 is typically only applied to classification issues involving solely nominal features.		Finding leaf nodes enables test data to be pruned, reducing number of tests. Whole dataset is searched to create tree.		
Random Forest	Use random forest if your forecasting data is based on a large data collection and several decisions. You can divide data into different groups, present it to various decision trees, merge different trees into a forest, and utilize majority voting to find the best possible option using random forest. An illustration would be determining the best-selling TV brand for the following year based on factors such as pricing, TVs sold the year prior, warranty, screen size, etc. An example of an ensemble is a random forest, which combines the results (decisions) of various algorithms. To construct a random forecast, many decision trees are created. Each decision tree forecasts a value, and the average of the forecasted values is then calculated. chosen. First, a tree must be created, and then the tree must be trained to foresee. Each tree in the ensemble is constructed using a sample from the training set that was drawn with replacement (i.e., a bootstrap sample). In addition, when splitting a node during tree construction, the split that is picked is not evenly distributed across all features. as opposed to the division that.	We could write the equation in terms of indicator functions for a single decision tree. Let us consider the following simple example: • y=1y=1 if x<5x<5 • y=2y=2 if 5≤x≤105≤x≤10 • y=3y=3 if x>10x>10 Then we could express the function as y=1×I(x<5)+2×I(5≤x≤10)+3×I(x>10)y=1×I(x<5)+2×I(5≤x≤10)+3×I(x>10) This wouldn't generalize to ensemble methods like Random Forest, though. It also doesn't accomplish anything that is not already expressed in pretty much any decision tree implementation; it's just a different expression of the same information.	High Precision. a good place to start when solving an issue. flexible and effective at fitting a wide range of data. swift in execution. simple to use helpful for classification and regression issues. be used to model missing values. It is quite effective.	Overtraining and Slow Training Tiny changes in training data can modify models, making them unsuitable for small samples. Sometimes, too easy a solution for extremely complex issues.	The computational complexity at test time for a Random Forest of size T and maximum depth D (excluding the root) is O (T ·D). However, the computational cost can be lower if trees are not balanced
Support Vector Machine	Among data scientists, Support Vector Machine (SVM) is undoubtedly one of the most widely utilized ML techniques. SVM is effective, simple to understand, and generally works well. I'll outline the justifications for SVM in this article and	If we know the weights and intercepts of the decision boundary, the boundary can be expressed by the following equation: Boundary:	When there is a large gap between classes, SVM performs comparatively well. In large dimensional spaces, SVM performs better.	Large data sets are not a good fit for the SVM algorithm. When the target classes are overlapping and the data set includes more noise, SVM does not perform very well.	The results of our research has proved that the complexity of SVM (LibSVM) is O(n3)

COMPARISION TABLES

MACHINE LEARNING ALGORITHMS

	demonstrate its Python implementation. I'll limit my attention to binary classification issues in this essay for simplicity's sake. SVM, however, allows for multiple classifications. In contrast to logistic regression, which measures optimality by total probability, SVM aims to maximize the size of the distance between the smallest data point and the decision boundary. In other words, SVM favors an 8-line freeway over a country road if you think of the decision boundary as the middle line of a street. The margin refers to the street's width.	$W_1X_1 + W_2X_2 + W_3X_3 + W_4X_4 + \ldots. + W_NX_N + B=0$	If there are more dimensions than samples, SVM works well in certain situations. SVM uses relatively little memory.	The SVM will perform poorly when there are more training data samples than features for each data point. There is no probabilistic justification for the classification because the support vector classifier places data points above and below the classifying hyperplane.	
Neural Network	We employ a neural network technique to construct voice recognition, self-driving cars, and self-trading traders. It draws inspiration from the brain's biological neural network. Each input that is received by a neuron (or node) has a weight assigned to it. In order to produce an output, the neuron then uses a function known as an activation function, such as RELU, SIGMOID, TANH, etc. The following layers, known as hidden layers, receive this output after which outputs are created. The term "multi-layer perceptron" refers to a network having numerous layers.	Neural Network is based on backpropagation and forward propagation method. $f\left(b + \sum_{i=1}^{n} x_i w_i\right)$ • b = bias • x = input to neuron • w = weights • i = a counter from 1to n • n = the number of inputs from the incoming layer	very precise a lot of variables to improve predictability can resolve challenging classification, deep learning, and non-linear issues.	Very slow forecasting and training. significant amount of data is needed. is conceivably a "black box." They are expensive to compute with and prone to overfitting.	For k⟶j, we have the time complexity $O(kt+klt+ktj+kj)=O(k*t(l+j))$. which is the same as the feedforward pass algorithm. Since they are the same, the total time complexity for one epoch will be $O(t*(ij+jk+kl))$. This time complexity is then multiplied by the number of iterations (epochs)
Convolution Neural Network	A neural network type called a convolutional neural network, or CNN or ConvNet, is particularly adept at processing input with a grid-like architecture, like an image. A binary representation of visual data is a digital image. It is made up of a grid-like arrangement of pixels, each of which has a pixel value to indicate how bright and what color it should be. The moment we perceive an image, the human brain begins processing a massive amount of data.	It uses the convolution operator.	Without any human oversight, it automatically recognizes the crucial characteristics.	Image classification according to positions negative examples Dimensional Frame additional small drawbacks, such as performance	

COMPARISION TABLES

MACHINE LEARNING ALGORITHMS

	Every neuron has a unique receptive field and is coupled to other neurons so that they collectively cover the whole visual field. Each neuron in a CNN processes data only in its receptive field, similar to how each neuron in the biological vision system responds to stimuli only in the constrained area of the visual field known as the receptive field. The layers are set up so that simpler patterns are detected early on and more complicated patterns later on. One can enable sight to computers by employing a CNN.				

COMPARISION TABLES

SECURITY ALGORITHMS

COMPARISION OF CRYPTOGRAPHIC ALGORITHMS

Name of Algorithm	Definition	Year	Developed By	Size		Rounds	Security	Avalanche Effect	Tunability	Vulnerabilities	Application
				Key	Block						
Data Encryption Standard [DES]	It is a block cipher used in the Data Encryption Standard (DES). It encrypts a plain text block of 64 bits with a key of 56 bits. It is made up of a fiestal network that divides a block into two equal halves, with the right half passing through a variety of functions. DES employs a series of S-boxes and P-boxes. The cypher text is obtained by the XOR operation after passing through these permutation and substitution boxes. DES employs 19 rounds. [2]	1977	IBM [International Business Machines]	56 Bits	64 Bits	16	Inadequate [1]	Less than AES [2]	No	Vulnerable to Linear and Differential Cryptoanalysis. [2]	Smart Card [4]
Triple Data Encryption Standard [3DES]	It is a symmetric-key encryption technique. It employs a block size of 64 bits and a key length of 56 bits to encrypt or decrypt any message or data. As the name implies, it applies the same DES algorithm to each data block three times. [2]	1978	IBM [International Business Machines]	168 Bits 112 Bits 56 Bits	64 Bits	48	Vulnerable [1]	Medium [2]	No	Vulnerable to differential brute force. Attackers can analyze plaintext. [2]	Microsoft OneNote Outlook 2007 [4]
Advance Encryption Standard [AES]	It operates on blocks of three sizes: 128 bits, 192 bits and 256 bits. AES-128 employs 10 rounds, AES-192 employs 12 rounds, AES-256 employs 14 rounds to encrypt and decrypt the message. In each round different steps are there like: substitution byte, shift rows, mixed columns and add round key. [2]	2000	NIST [National Institute of Standards and technology] John Daemen Vincent Rijmen	128 Bits 192 Bits 256 Bits	128 Bits	10 - 128 Bits 12 - 192 Bits 14 - 256 Bits	High [1]	Faster Encryption / Decryption. Less time than DES. [2]	No	Strong against truncated differential, linear, interpolation and square attacks. [2]	Password Manager [4]
Rivest-Shamir-Adleman [RSA]	This cryptographic algorithm is a widely-used method of public key encryption and was first described by Rivest, Shamir, and Adelman. It uses two keys: one key is used to encrypt the message and another key is used to decrypt the message. [2]	1977	Ron Rivest Adi Shamir Leonard Adleman	Depends on the number of bits in the modulus n **n = p * q** where p and q are prime numbers	Variable [Minimum 512 Bits]	N/A	High [1]	Slower Encryption / Decryption [2]	Yes	Brute Force Attack difficult to accomplish. [2]	Online Credit Card Security System RSA Signature Verification [4]
Elliptical Curve Cryptography [ECC]	This cryptographic algorithm is a widely used public key encryption system. Encryption is the process of transforming plaintext into ciphertext and it's often used to	1985	Victor Miller [IBM] Neil Koblitz	256 Bits	N/A	N/A	Very High	Fastest and efficient.	Yes	Pollard's rho Algorithm	Digital Signatures Mutual Authentication

COMPARISION TABLES

SECURITY ALGORITHMS

	make sure that only intended recipients can access sensitive information. The ECC algorithm is based on two prime numbers a generator and a base that are multiplied together to form two pairs of numbers that are then multiplied together. The resulting product is called an Elliptic Curve. [6]		[University of Washington]								Secure Data Transmission
International Data Encryption Algorithm [IDEA]	Using symmetric key block cyphers, the International Data Encryption Algorithm (IDEA)employs a fixed-length plaintext of 16 bits that is encrypted into a 16-bit ciphertext using 4 chunks of 4 bits each. The key being utilized is 32 bits long. Additionally, the key is split into 8 blocks, each with 4 bits. [5]	1991	Xuejia Lai James L. Massley	128 Bits	64 Bits	8.5	Low	Slow	No	Vulnerable to Brute Force Attacks.	Audio and video data for cable TV, video conferencing, etc. Smart Cards Email via public networks [5]
Blowfish	It has a variable key length with a maximum of 448 bits. It has a 64-bit block size. The blowfish algorithm consists of two stages The first stage is the key expansion phase, which converts a 448-bit key into a number of sub keys bringing the total 4168 bytes. The second stage is the encryption phase, which involves iterating a function 16 times and obtaining the encrypted text via the XOR operation. [2]	1993	Bruce Schiener	32 Bits upto 448 Bits	64 Bits	16	Moderate	Fastest. Except when changing keys. [2]	No	Vulnerable to differential Brute Force Attacks. [2]	IDS, Server.Sql Server 2000 [4]
TwoFish	This is a block-encrypted symmetric key cryptography. The size of block used in this cryptography is of 128 bits and the size of key is of variable length (128, 192 or 256 bits). It is Open Source (not licensed), patent free and freely available. 2 FISH is quite similar to older symmetric-key block cipher BLOWFISH. It also includes extended functionality to substitute the Data Encryption Standard (DES) algorithm. [3]	1998	Bruce Schiener	128 Bits 192 Bits 256 Bits	128 Bits	16	Moderate	Slower Encryption / Decryption compared to RSA.	Yes	Highly secure with still no cryptoanalysis found. [3]	GNU Privacy Guard Software Calenderscope .NET CEX
Elgamal	It is a type of asymmetric cryptographic algorithm. The difficulty of using the cyclic group to find the discrete logarithm is its key concern. Even if the attacker knows the values of g^a and g^b, it will still be very challenging for him to determine the value of g^ab, which is simply the cracked value.	1984	Taher ElGamal	512 Bits 1024 Bits 2048 Bits	514 Bits	N/A	Moderate	Efficient	No	Adaptive chosen ciphertext attacks.	Digital Signature Algorithm GNU Privacy Guard Software Pretty Good Privacy Versions

COMPARISION TABLES

HASHING ALGORITHMS

COMPARISION OF HASHING ALGORITHMS

Name of Algorithm	Definition	Year	Construction	Size			Rounds	Collision Level	Operations	Weakness	Successful Attacks	Security Level	Application
				Block	Digest	Word							
Message Digest 5 [MD5]	It is a widely used hash function producing a 128-bit hash value. Although it was initially designed to be used as a cryptographic hash function, it has been found to suffer from extensive vulnerabilities. It can still be used as a checksum to verify data integrity, but only against unintentional corruption. **MD5 Function + Input String = 128 - bit Message Digest** [6]	1992	Merkle - Damgard	512 Bits	128 Bits	32 Bits	64	High They can be found in seconds, even using an ordinary home computer. [4]	ADD XOR AND OR NOT SHIFT [5]	Vulnerable to Collisions.	Hash Collisions, Brute Force Attack, etc. [4]	Low	It used for verifying the integrity of files against involuntary corruption. [4]
Secure Hash Algorithm 1 [SHA-1]	It is a cryptographic hash function which takes an input and produces a 160-bit hash value known as a message digest. It is also used to identify that during transmission of information from sender to receiver is any changes are occurring or not. [6]	1995	Merkle - Damgard	512 Bits	160 Bits	32 Bits	80 [4 groups of 20 rounds]	Theoretical Attack Cheap and easy to find. [4] [6]	ADD XOR AND OR NOT ROTATE [5]	Vulnerable to collisions.	Chosen prefix collision attack, Collision Attack, etc. [4]	Low	Used in TLS, SSL, and HMAC for verifying the integrity of files against involuntary corruption. [4]
Secure Hash Algorithm 2 [SHA-2]	There are six distinct SHA-2 variations, which vary in direct proportion to the bit size being used to encrypt data. A 256-bit hash is produced by SHA-256, which also has a 512-bit block size. The initialization variables and constants are 32 bits long, and the message input is handled in 32-bit words. [5]	2002	Merkle - Damgard	512 Bits 1024 Bits	224 Bits 256 Bits 384 Bits 512 Bits	32 Bits 64 Bits	64 [224, 256 Bits] 80 [384, 512 Bits]	Low No known collisions found to date. [4]	ADD XOR OR AND SHIFT ROTATE [5]	Susceptible to preimage attacks. [4]	It has never been broken.	High	Security applications and protocols Cryptocurrencies Transactions Validation Digital certificates [4]
Secure Hash Algorithm 3 [SHA-3]	The purpose of SHA-3 is that it can be directly substituted for SHA-2 in current applications if necessary, and to significantly improve the robustness of NIST's overall hash algorithm toolkit. To ensure the message can be evenly divided into r-bit blocks, padding is required. [4]	2008	Sponge [Keccak]	1152 Bits 1088 Bits 832 Bits 576 Bits	224 Bits 256 Bits 384 Bits 512 Bits	64 Bits	24	None [6]	N/A	Susceptible to practical collision and near collision attacks. [4]	Few collision type attacks have been demonstrated.	High	Cryptocurrencies Transactions Validation

SWOT ANALYSIS

STRENGTH

The model provides the best solution for the attack. With the help of the Machine Learning algorithm the model invokes the best security algorithm for the specific attack and hence the security algorithm with the higher accuracy percent is invoked into action and the security is granted.

WEAKNESS

Due to the dummy dataset proper solution is not provided.

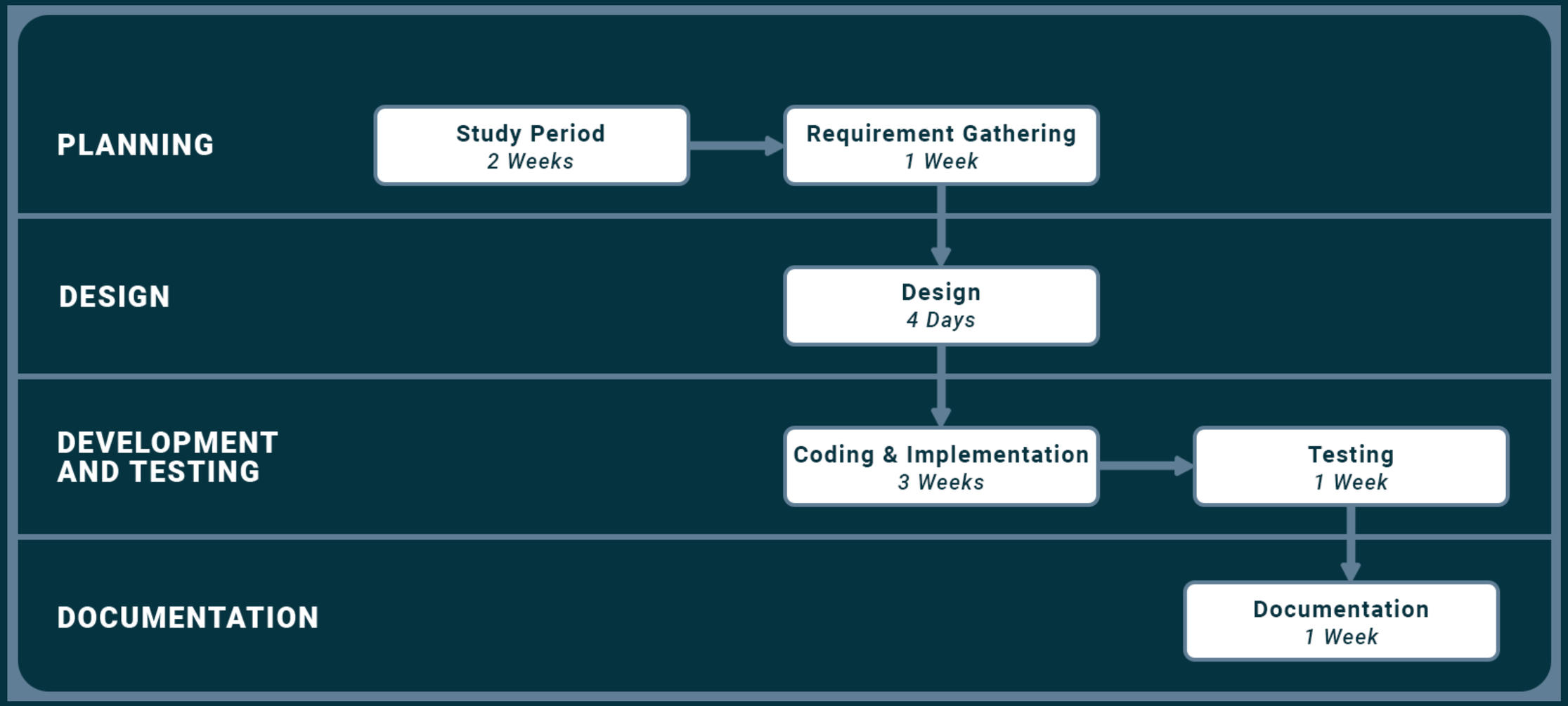
OPPORTUNITIES

Can be used in various field for the security purpose. Like in banking sector, Iot devices etc.

THREAT

If any change to dataset is made the prediction can go wrong.

PERT CHART


PLANNING
Study Period
2 Weeks
Requirement Gathering
1 Week
DESIGN
Design
4 Days
DEVELOPMENT AND TESTING
Coding & Implementation
3 Weeks
Testing
1 Week
DOCUMENTATION
Documentation
1 Week

OBJECTIVES COVERED

Objectives	Status
Implementation 1. Security Algorithms 2. Hashing Algorithms 3. Machine Learning Algorithms	Completed
Comparison Table 1. Security Algorithms 2. Hashing Algorithms 3. Machine Learning Algorithms	Completed
Compilation	Completed

OUTPUT

TEST CASE: 1

Problems | Console

<terminated> Main [Java Application] D:\Program Files\Java\bin\javaw.exe (Nov 26, 2022, 10:42:39 PM – 10:43:04 PM) [pid: 21812]

```
Enter the name of the attack :
MAN IN THE MIDDLE
The best security algorithm for the enetered attack is : RSA

Invoking The Security Algorithm:RSA

Enter Value of P = 3
Enter Value of Q = 7
Value of N = (P * Q) => 21
Value of Phi of N = (P-1) * (Q-1) => 12
value of E  = 5
Value of D = 5
        Your Public key = { 5 , 21 }
        Your Private key = { 5 , 21 }
Enter a Character = D
Plain Text = D
Cypher Text = 17
```

OUTPUT

TEST CASE: 2

Problems Console

<terminated> Main [Java Application] D:\Program Files\Java\bin\javaw.exe (Nov 26, 2022, 10:44:58 PM – 10:45:25 PM) [pid: 23260]

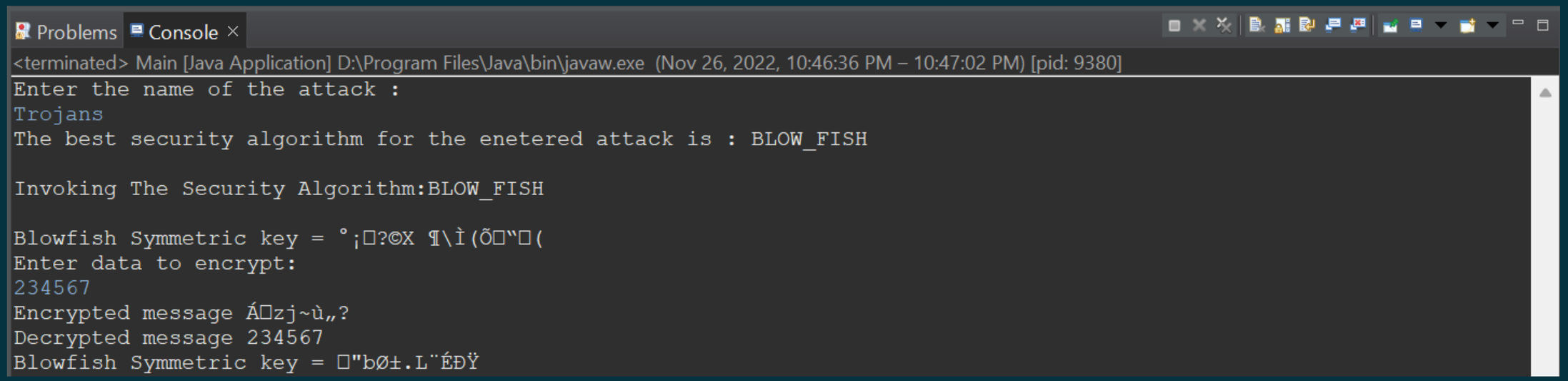
```
Enter the name of the attack :
RANSOMWARE
The best security algorithm for the enetered attack is : ELGAMAL

Invoking The Security Algorithm:ELGAMAL

Enter the approximate value of the prime number for your El Gamal key: 977
Post p = 977 g = 145 b = 801
Please enter your message. It should be in between 1 and 977: 57
The corresponding cipher texts are c1 = 873 c2 = 504
Here is c1^ -a = 634
The original message = 57
```

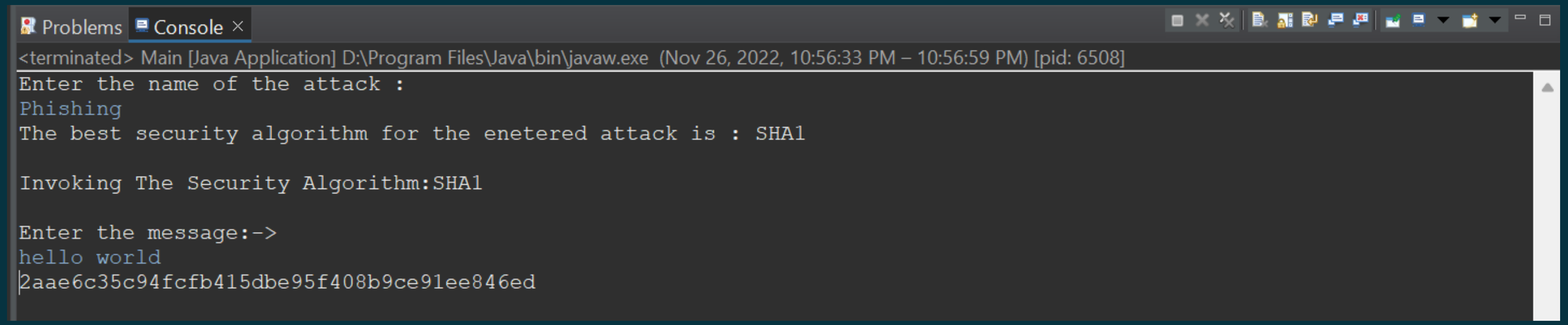
OUTPUT

TEST CASE: 3



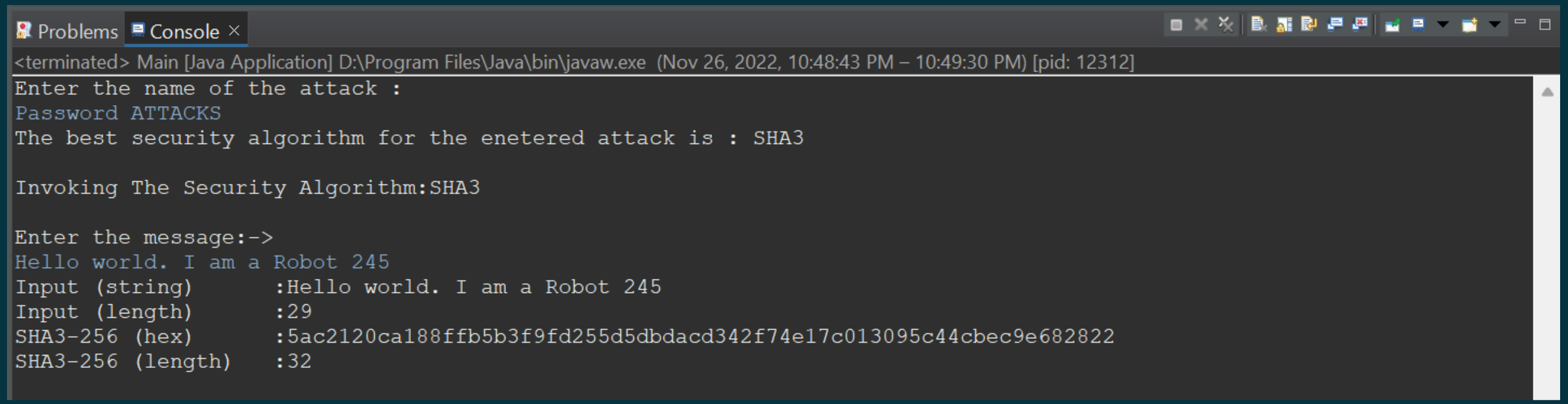
OUTPUT

TEST CASE: 4



OUTPUT

TEST CASE: 5



FUTURE WORK

- Our project can also be used on many IoT (Internet of things) devices like smart door.
- Motion detector or any activity tracker like CCTV to insure the confidentiality and integrity of the data.
- We can also use in banking sector to ensure the security of monetary transactions including the security of ATM cards.
- Computer passwords, and electronic commerce.
- Every operating system uses encryption in some of the core components to keep passwords secret.
- Email encryption is a method of securing the content of emails from anyone outside of the email conversation looking to obtain a participant’s information.

REFERENCES

https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.677.5654&rep=rep1&type=pdf

S. Zeadally, A.K. Das and N. Sklavos; Cryptographic technologies and protocol standards for Internet of Things, Internet of Things

Cong Pu, Andrew Wall, Kim-Kwang Raymond Choo, Imtiaz Ahmed, Sunho Lim; A Lightweight and Privacy-Preserving Mutual Authentication and Key Agreement Protocol for Internet of Drones Environment; Volume: 9, Issue: 12; Page(s): 9918 – 9933

Muhammad Wahid Akram, Ali Kashif Bashir, Salman Shamshad, Muhammad Asad Saleem, Ahmad Ali AlZubi, Shehzad Ashraf Chaudhry, Bander A. Alzahrani, Yousaf Bin Zikria; A Secure and Lightweight Drones-Access Protocol for Smart City Surveillance; Page(s): 1 – 10

Muktar Yahuza, Mohd Yamani Idna Idris, Ismail Bin Ahmedy, Ainuddin Wahid Abdul Wahab, Tarak Nandy, Noorzaily Mohamed Noor, Abubakar Bala; Internet of Drones Security and Privacy Issues: Taxonomy and Open Challenges; Volume: 9; Page(s): 57243 – 57270

Zhihan Lv, Yuxi Li, Jingyi Wu, Haibin Lv; Securing the Internet of Drones Against Cyber-Physical Attacks; Volume: 4, Issue: 4; Page(s): 74 – 78

Vincent Omollo Nyangaresi, M.A. Morsy; Towards Privacy Preservation in Internet of Drones

REFERENCES

- https://codesigningstore.com/hash-algorithm-comparison
- Ali Maetouq, Salwani Mohd Daud, Noor Azurati Ahmad, Nurazean Maarop, Nilam Nur Amir Sjarif, Hafiza Abas; Comparison of Hash Function Algorithms Against Attacks: A Review; IJACSA, Volume: 9, Number: 8, 2018
- Prashant P. Pittalia; A Comparative Study of Hash Algorithms in Cryptography; Volume: 8 Issue: 6, June - 2019, Page: 147-152
- Dr. Kiramat Ullah, Bibi Ayisha, Farrukh Irfan, Inaam Illahi, Zeeshan Tahir; Comparison of Various Encryption Algorithms for Securing Data; PIEAS
- Mr. Pradeep Semwal and Dr. MK Sharma; Comparative Study of Different Cryptographic Algorithms for Data Security in Cloud Computing; Volume: 8 Issue: 1, Pages: 746-750, NCETST-2017
- Shailendra Singh Gaur, Hemanpreet Singh Kalsi, Shivani Gautam; A Comparative Study and Analysis of Cryptographic Algorithms: RSA, DES, AES, BLOWFISH, 3-DES, and TWOFISH; IJRECE Volume: 7, Issue: 1 (JANUARY- MARCH 2019)
- Ms.Theres Bemila, Karan Kundar, Lokesh Jain, Shashikant Sharma, Nayan Makasare; Comparative study of various Security Algorithms applicable in Multi-Cloud Environment; IJARCCE, Volume: 5, Issue: 3, March 2016

REFERENCES

- How-Shen Chang; International Data Encryption Algorithm; CS-627-1, Fall 2004
- Zhihan Lv, Yuxi Li, Jingyi Wu, and Haibin Lv; Securing the Internet of Drones against Cyber-Physical Attacks; IEEE Internet of Things Magazine; December 2021

THANK YOU