Deep Learning

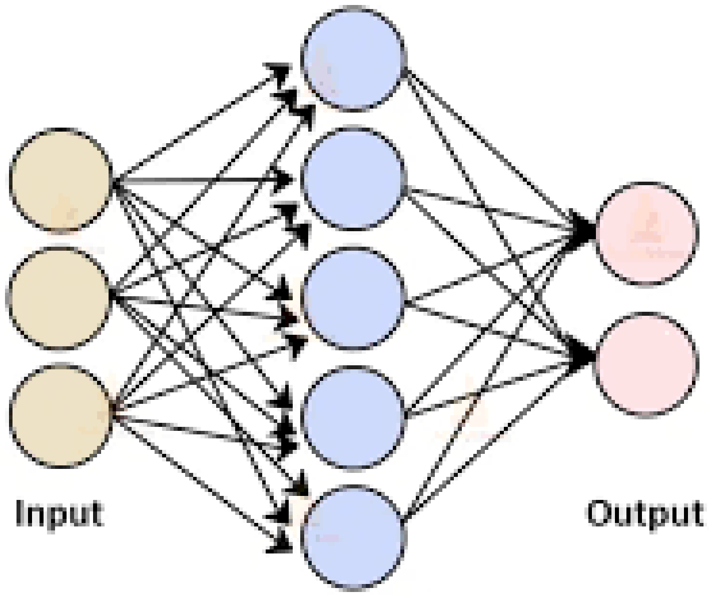
Deep learning is a method in artificial intelligence (AI) that teaches computers to process data in a way that is inspired by the human brain. Deep learning models can recognize complex patterns in pictures, text, sounds, and other data to produce accurate insights and predictions. You can use deep learning methods to automate tasks that typically require human intelligence, such as describing images or transcribing a sound file into text.

Uses of Deep Learning

- Artificial Neural Networks for Regression and Classification
- Convolutional Neural Networks for Computer Vision
- Recurrent Neural Networks for Time Series Analysis
- Self-Organizing Maps for Feature Extraction
- Deep Boltzmann Machines for Recommendation Systems
- Auto Encoders for Recommendation Systems

How does deep learning work?

- Deep learning algorithms are neural networks that are modeled after the human brain.
- For example, a human brain contains millions of interconnected neurons that work together to learn and process information. Similarly, deep learning neural networks, or artificial neural networks, are made of many layers of artificial neurons that work together inside the computer.
- Artificial neurons are software modules called nodes, which use mathematical calculations to process data. Artificial neural networks are deep learning algorithms that use these nodes to solve complex problems.



Components of a Deep Learning network

Input layer:

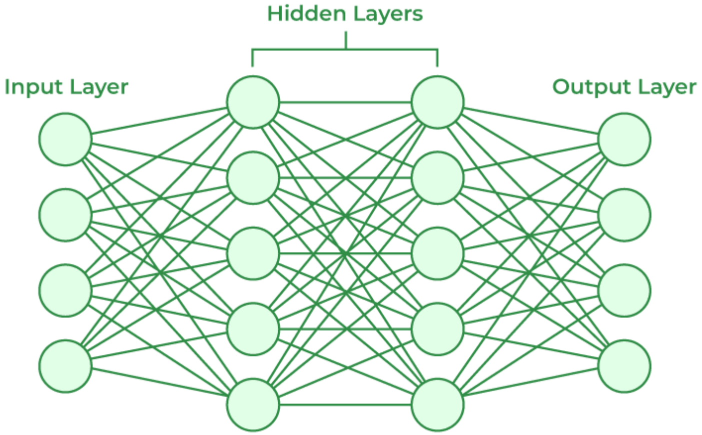
- An artificial neural network has several nodes that input data into it. These nodes make up the input layer of the system.

Hidden layer:

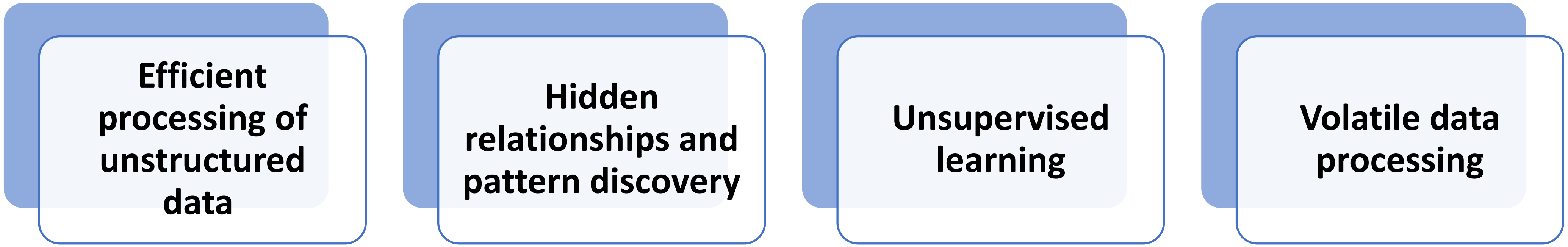
- The input layer processes and passes the data to layers further in the neural network. These hidden layers process information at different levels, adapting their behavior as they receive new information. Deep learning networks have hundreds of hidden layers that they can use to analyze a problem from several different angles.

Output layer:

- The output layer consists of the nodes that output the data. Deep learning models that output “Yes" or “No" answers have only two nodes in the output layer. On the other hand, those that output a wider range of answers have more nodes.



Benefits of Deep Learning



- Efficient processing of unstructured data
- Hidden relationships and pattern discovery
- Unsupervised learning
- Volatile data processing

Challenges of Deep Learning

Large quantities of high-quality data

- Deep learning algorithms give better results when you train them on large amounts of high-quality data. Outliers or mistakes in your input dataset can significantly affect the deep learning process
- To avoid such inaccuracies, you must clean and process large amounts of data before you can train deep learning models. The input data preprocessing requires large amounts of data storage capacity.

Large processing power

- Deep learning algorithms are compute-intensive and require infrastructure with sufficient compute capacity to properly function. Otherwise, they take a long time to process results.

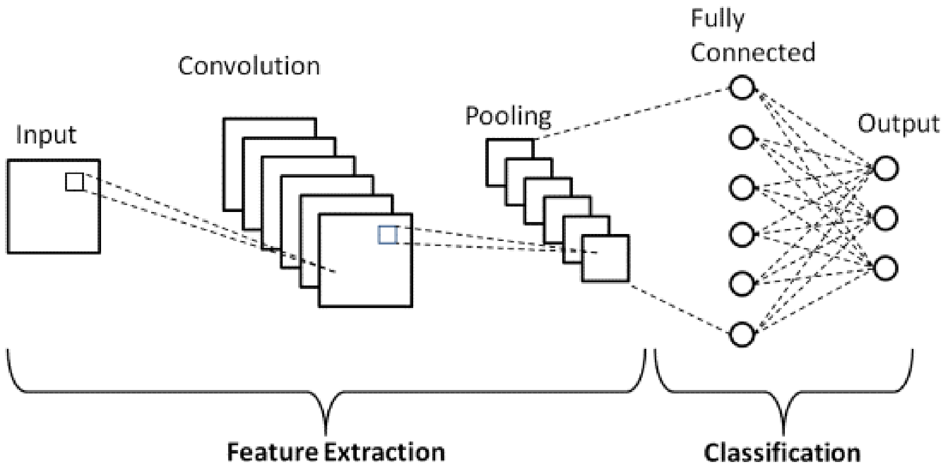
Convolutional Networks(LeCun, 1989)

- Convolutional neural networks or CNNs, are a specialized kind of neural network for processing data that has a known, grid-like topology.
- Convolutional networks are simply neural networks that use convolution in place of general matrix multiplication in at least one of their layers.
- Their name and structure are inspired by the human brain, mimicking the way that biological neurons signal to one another.

How do Convolutional Neural Networks work?

Main types of layers that make up the CNN which are:

- Convolutional layer
- Pooling layer
- Fully-connected (FC) layer
- Batch Normalization
- Dropout

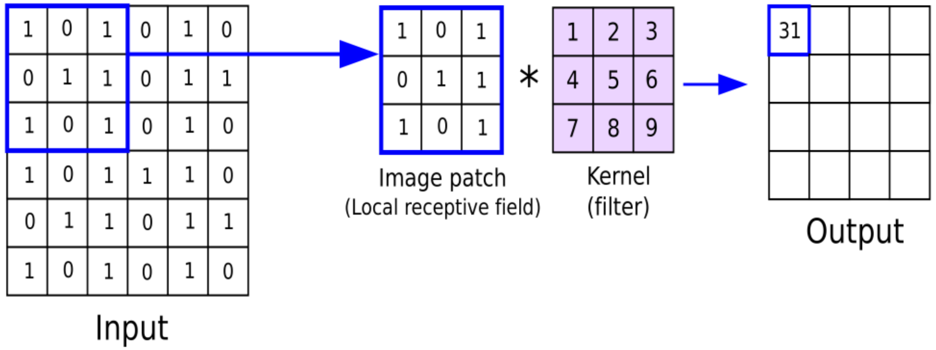


How do convolutional neural networks work?

- The convolutional layer is the first layer of a convolutional network.
- Convolutional layers can be followed by additional convolutional layers or pooling layers.
- The fully-connected layer is the final layer.
- With each layer, the CNN increases in its complexity.
- Earlier layers focus on simple features, such as colors and edges.
- As the image data progresses through the layers of the CNN, it starts to recognize larger elements or shapes of the object until it finally identifies the intended object.

Convolutional layer

- A color image, which is made up of a matrix of pixels in 3D will have three dimensions—a height, width, and depth.
- A **kernel** or a **filter** move across the receptive fields of the image, checking if the feature is present.
- The **feature detector** is a two-dimensional (2-D) array of weights
- The filter is then applied to an area of the image, and a dot product is calculated between the input pixels and the filter. This dot product is then fed into an output array.
- Afterwards, the filter shifts by a **stride**, repeating the process until the kernel has swept across the entire image.
- The final output from the series of dot products from the input and the filter is known as a **feature map**, **activation map**, or a **convolved feature**.



1	0	1	0	1	0
0	1	1	0	1	1
1	0	1	0	1	0
1	0	1	1	1	0
0	1	1	0	1	1
1	0	1	0	1	0

Input

1	0	1
0	1	1
1	0	1

Image patch
(Local receptive field)

∗

1	2	3
4	5	6
7	8	9

Kernel
(filter)

31			

Output

Convolutional layer

- The weights in the feature detector remain fixed as it moves across the image, which is also known as parameter sharing.
- Some parameters, like the weight values, adjust during training through the process of **backpropagation** and **gradient descent**.
- There are three hyperparameters which affect the volume size of the output that need to be set before the training of the neural network begins. These include:
  1. Number of filters
  2. Stride
  3. Zero-padding
- There are three types of **padding**:
  1. Valid padding
  2. Same padding
  3. Full padding
- After each convolution operation, a CNN applies an **Activation Function** to the feature map, introducing nonlinearity to the model.

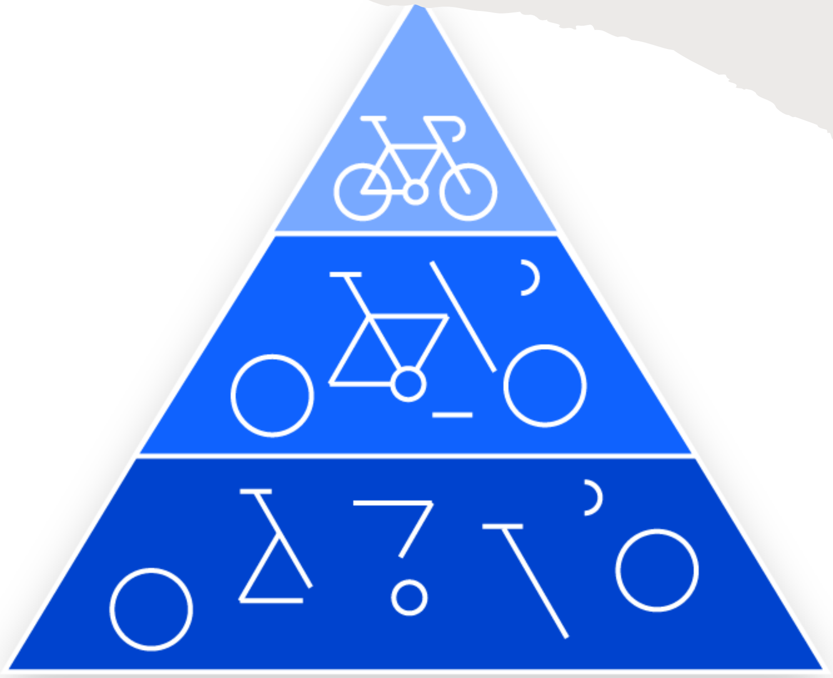
Activation Layers

- After each Convolutional layer in a CNN, we apply a nonlinear activation function, such as ReLU, ELU, or any of the other Leaky ReLU variants. We typically denote activation layers as RELU
- In network diagrams as since ReLU activations are most used, we may also simply state ACT
- Activation layers are not technically “layers” (since no parameters/weights are learned inside an activation layer)

Additional convolutional layer

- Another convolution layer can follow the initial convolution layer
- The structure of the CNN can become hierarchical as the later layers can see the pixels within the receptive fields of prior layers.

Example:



- Let's assume that we're trying to determine if an image contains a bicycle.
- You can think of the bicycle as a sum of parts. It is comprised of a frame, handlebars, wheels, pedals, et cetera.
- Each individual part of the bicycle makes up a lower-level pattern in the neural net, and the combination of its parts represents a higher-level pattern, creating a feature hierarchy within the CNN.
- Ultimately, the convolutional layer converts the image into numerical values, allowing the neural network to interpret and extract relevant patterns.

Pooling layer

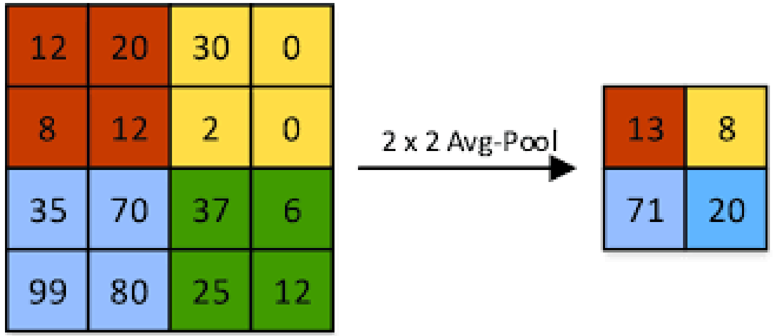
- Pooling layers, also known as down sampling, conducts dimensionality reduction, reducing the number of parameters in the input.
- The pooling operation sweeps a filter across the entire input, but the difference is that this filter does not have any weights.
- Instead, the kernel applies an aggregation function to the values within the receptive field, populating the output array.
- They help to reduce complexity, improve efficiency, and limit risk of overfitting.
- There are two main types of pooling:

Max pooling

Average pooling

Max Pooling and Average Pooling

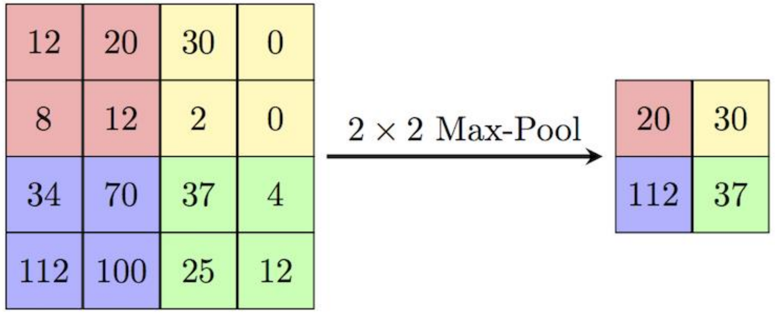
- **Max pooling:** As the filter moves across the input, it selects the pixel with the maximum value to send to the output array. As an aside, this approach tends to be used more often compared to average pooling.
- **Average pooling:** As the filter moves across the input, it calculates the average value within the receptive field to send to the output array.



12	20	30	0
8	12	2	0
35	70	37	6
99	80	25	12

2 x 2 Avg-Pool

13	8
71	20



12	20	30	0
8	12	2	0
34	70	37	4
112	100	25	12

2 × 2 Max-Pool

20	30
112	37

Fully-connected layer

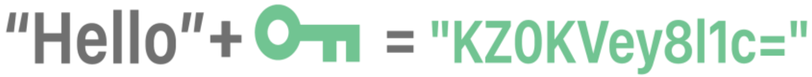
- This layer performs the task of classification based on the features extracted through the previous layers and their different filters.
- While convolutional and pooling layers tend to use ReLu functions, FC layers usually leverage a softmax activation function to classify inputs appropriately, producing a probability from 0 to 1.

Encryption

- Encryption is a way of scrambling data so that only authorized parties can understand the information.
- In technical terms, it is the process of converting human-readable plaintext to incomprehensible text, also known as ciphertext.
- Encryption requires the use of a **Cryptographic key**: a set of mathematical values that both the sender and the recipient of an encrypted message agree on.

What is a cryptographic key?

- In cryptography, a key is a string of characters used within an encryption algorithm for altering data so that it appears random. Like a physical key, it locks (encrypts) data so that only someone with the right key can unlock (decrypt) it.
- The original data is known as the *plaintext*, and the data after the key encrypts it is known as the *ciphertext*.



How does encryption work?

- Encryption is a mathematical process that alters data using an encryption algorithm and a key.
- Although encrypted data appears random, encryption proceeds in a logical, predictable way. allowing a party that receives the encrypted data and possesses the right key to decrypt the data, turning it back into plaintext.
- Truly secure encryption will use keys complex enough that a third party is highly unlikely to decrypt or break the ciphertext by brute force.
- Data can be encrypted "at rest," when it is stored, or "in transit," while it is being transmitted somewhere else.

Types of Encryption

- The two main kinds of encryption are **symmetric encryption** and **asymmetric encryption**. Asymmetric encryption is also known as public key encryption.
- **Symmetric Encryption:** In symmetric encryption, there is only one key, and all communicating parties use the same (secret) key for both encryption and decryption.
- **Symmetric Encryption:** In asymmetric, or public key, encryption, there are two keys: one key is used for encryption, and a different key is used for decryption.

Importance of Encryption:

Privacy: Encryption ensures that no one can read communications or data at rest except the intended recipient or the rightful data owner.

Security: Encryption helps prevent data breaches, whether the data is in transit or at rest.

Data integrity: Encryption also helps prevent malicious behavior such as on-path attacks.

Regulations: For all these reasons, many industry and government regulations require companies that handle user data to keep that data encrypted.

Common Encryption Algorithms

Commonly used symmetric encryption algorithms include:

- AES
- 3-DES
- SNOW

Commonly used asymmetric encryption algorithms include:

- RSA
- Elliptic curve cryptography

Block Cipher

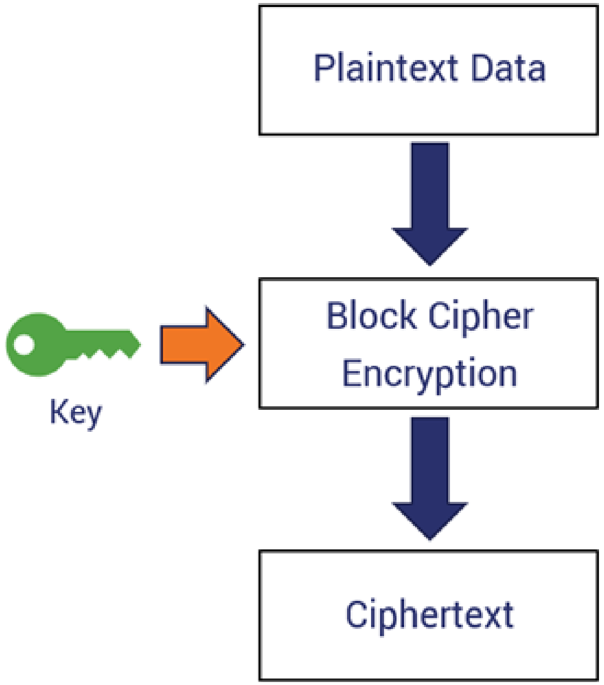
- A block cipher is a method of encrypting data in blocks to produce ciphertext using a cryptographic key and algorithm.
- The block cipher processes fixed-size blocks simultaneously, as opposed to a stream cipher, which encrypts data one bit at a time.
- Most modern block ciphers are designed to encrypt data in fixed-size blocks of either 64 or 128 bits.

Applications of Block Ciphers

- Data Encryption
- File and Disk Encryption
- Virtual Private Networks (VPN)
- Secure Sockets Layer (SSL) and Transport Layer Security (TLS)
- Digital Signatures

How does a block cipher work?

- A block cipher uses a symmetric key and algorithm to encrypt and decrypt a block of data.
- A block cipher requires an **Initialization Vector** (IV) that is added to the input plaintext to increase the key space of the cipher and make it more difficult to use brute force to break the key.
- The IV is derived from a random number generator, which is combined with text in the first block and the key to ensure all subsequent blocks result in ciphertext that does not match that of the first encryption block.
- The **block size** of a block cipher refers to the number of bits that are processed together.

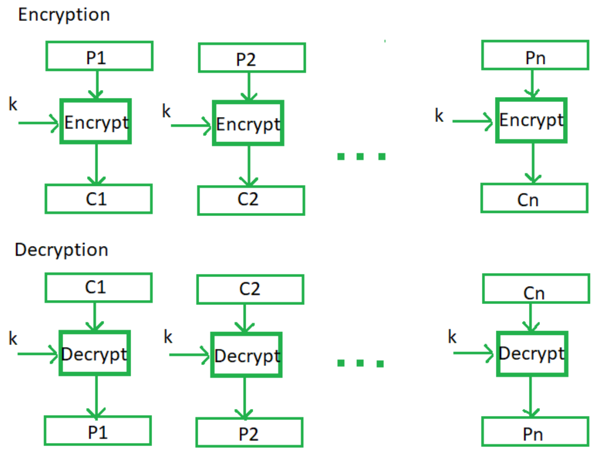


Block Cipher modes of Operation

- **Block cipher** is an encryption algorithm that takes a fixed size of input say b bits and produces a ciphertext of b bits again.
- If the input is larger than b bits it can be divided further.
- For different applications and uses, there are several modes of operations for a block cipher.
- There are several modes of operations for a block cipher:
  - Electronic Code Book (ECB)
  - Cipher Block Chaining (CBC)
  - Cipher Feedback Mode (CFB)
  - Output Feedback Mode (OFM)
  - Counter Mode

Electronic Code Book (ECB)

- Electronic code book is the easiest block cipher mode of functioning.
- It is easier because of direct encryption of each block of input plaintext and output is in form of blocks of encrypted ciphertext.
- Generally, if a message is larger than *b* bits in size, it can be broken down into a bunch of blocks and the procedure is repeated.



Advantages and Disadvantages of ECB

Advantages of using ECB:

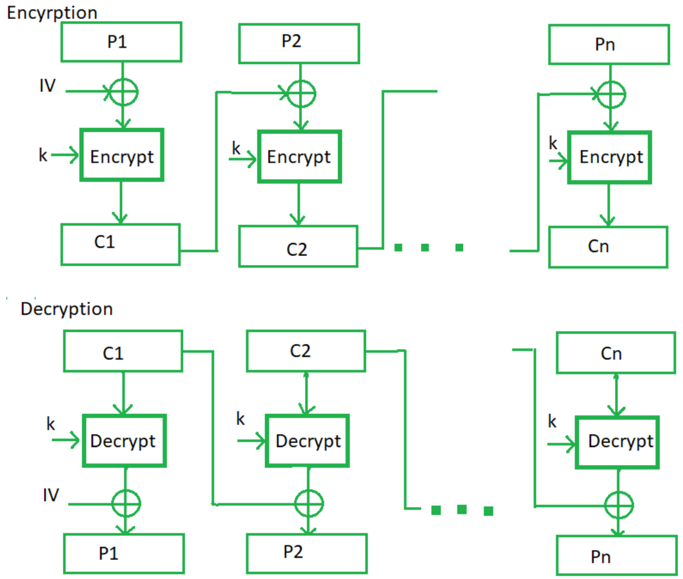
- Parallel encryption of blocks of bits is possible, thus it is a faster way of encryption.
- Simple way of the block cipher.

Disadvantages of using ECB:

- Prone to cryptanalysis since there is a direct relationship between plaintext and ciphertext.

Cipher Block Chaining (CBC)

- CBC is an advancement made on ECB since ECB compromises some security requirements.
- In CBC, the previous cipher block is given as input to the next encryption algorithm after XOR with the original plaintext block.
- Here, a cipher block is produced by encrypting an XOR output of the previous cipher block and present plaintext block.



Advantages and Disadvantages of CBC

Advantages of CBC:

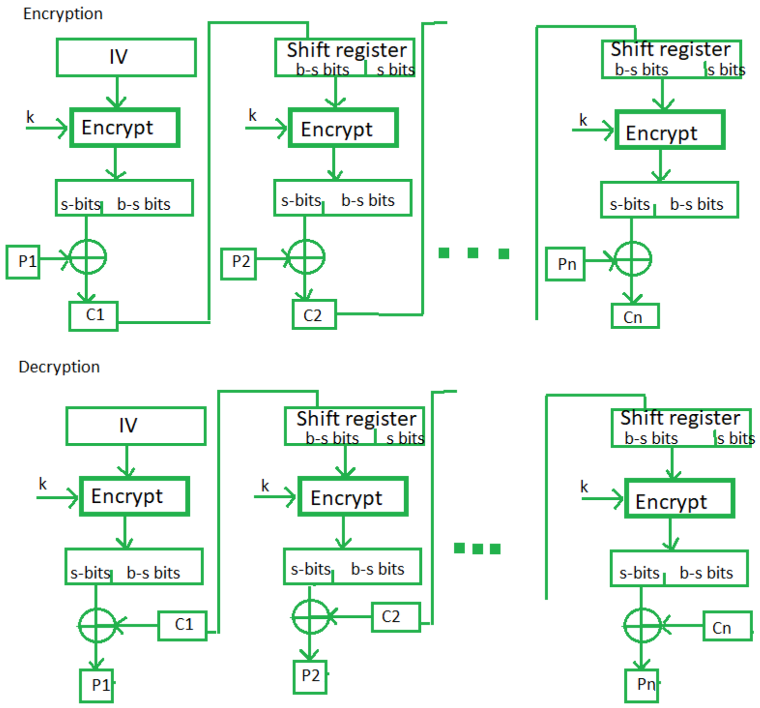
- CBC works well for input greater than *b* bits.
- CBC is a good authentication mechanism.
- Better resistive nature towards cryptanalysis than ECB.

Disadvantages of CBC:

- Parallel encryption is not possible since every encryption requires a previous cipher.

Cipher Feedback Mode (CFB)

- In this mode the cipher is given as feedback to the next block of encryption with some new specifications:
- First, an initial vector IV is used for first encryption and output bits are divided as a set of *s* and *b*-*s* bits.
- The left-hand side *s* bits are selected along with plaintext bits to which an XOR operation is applied.
- The result is given as input to a shift register having b-s bits to lhs, s bits to rhs and the process continues.



Advantages and Disadvantages of CFB

Advantages of CFB:

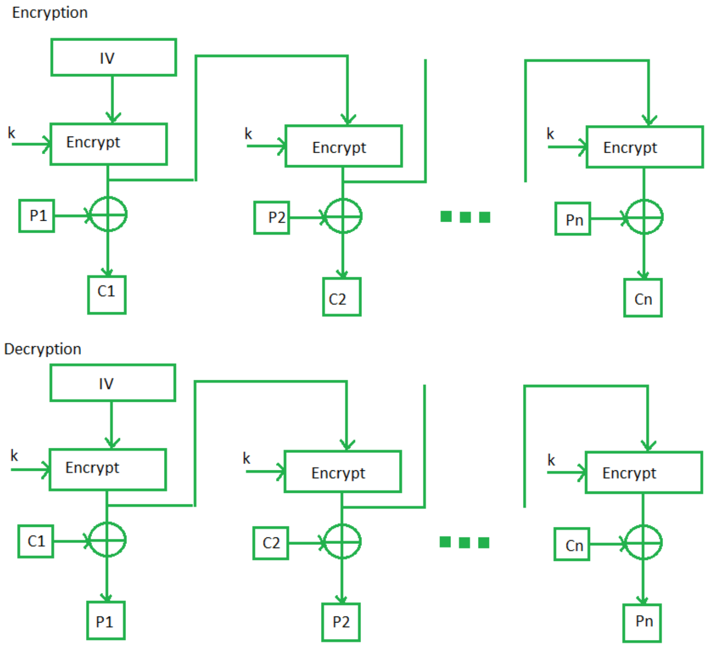
- Since, there is some data loss due to the use of shift register, thus it is difficult for applying cryptanalysis.

Disadvantages of using CFB:

- The drawbacks of CFB are the same as those of CBC mode. Both block losses and concurrent encryption of several blocks are not supported by the encryption. Decryption, however, is parallelizable and loss-tolerant.

Output Feedback Mode (OFB)

- The output feedback mode follows nearly the same process as the Cipher Feedback mode except that it sends the encrypted output as feedback instead of the actual cipher which is XOR output.
- In this output feedback mode, all bits of the block are sent instead of sending selected *s* bits.
- The Output Feedback mode of block cipher holds great resistance towards bit transmission errors. It also decreases the dependency or relationship of the cipher on the plaintext.



Advantages and Disadvantages of OFB

Advantages of OFB:

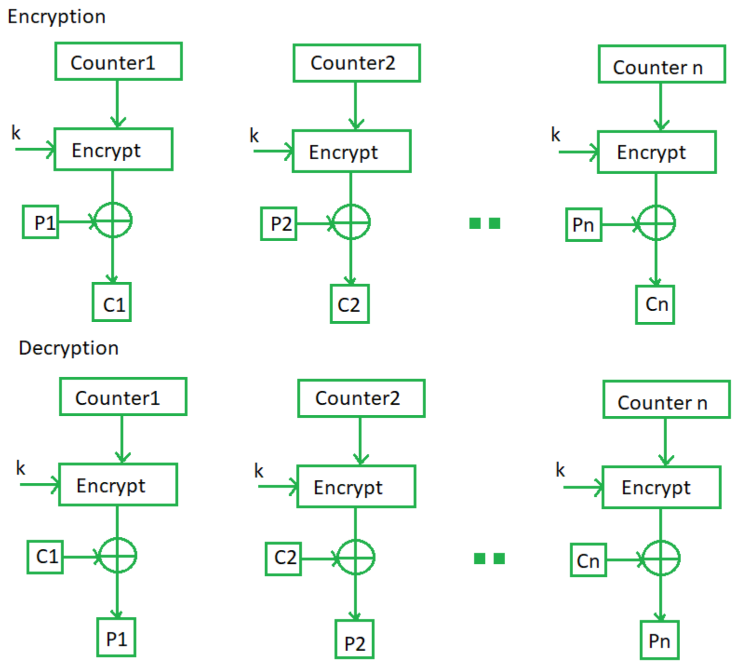
- In the case of CFB, a single bit error in a block is propagated to all subsequent blocks. This problem is solved by OFB as it is free from bit errors in the plaintext block.

Disadvantages of OFB:

- The drawback of OFB is that, because to its operational modes, it is more susceptible to a message stream modification attack than CFB.

Counter Mode (CTR)

- The Counter Mode or CTR is a simple counter-based block cipher implementation. Every time a counter-initiated value is encrypted and given as input to XOR with plaintext which results in ciphertext block.
- The CTR mode is independent of feedback use and thus can be implemented in parallel.



Advantages and Disadvantages of CTR

Advantages of CTR:

- Since there is a different counter value for each block, the direct plaintext and ciphertext relationship is avoided. This means that the same plain text can map to different ciphertext.
- Parallel execution of encryption is possible as outputs from previous stages are not chained as in the case of CBC.

Disadvantages of CTR:

- The fact that CTR mode requires a synchronous counter at both the transmitter and the receiver is a severe drawback. The recovery of plaintext is erroneous when synchronization is lost.

Data encryption standard (DES)

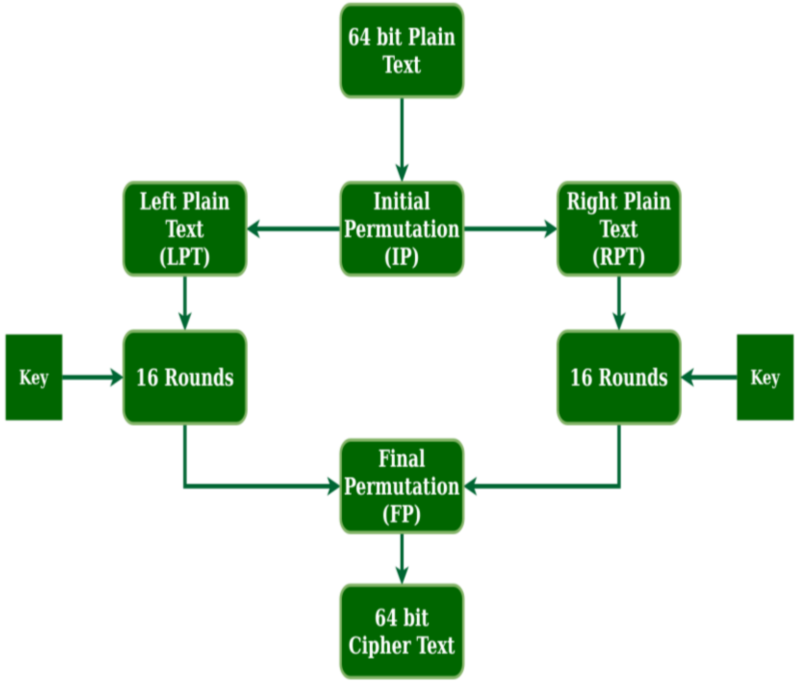
- Data Encryption Standard (DES) is a block cipher with a **56-bit key** length that has played a significant role in data security**.**
- DES is a block cipher and encrypts data in blocks of size of **64 bits** each, which means 64 bits of plain text go as the input to DES, which produces 64 bits of ciphertext. The same algorithm and key are used for encryption and decryption, with minor differences.
- The initial key consists of 64 bits. However, before the DES process even starts, every 8th bit of the key is discarded to produce a 56-bit key. That is bit positions 8, 16, 24, 32, 40, 48, 56, and 64 are discarded.

Data encryption standard (DES)

- DES is based on the two fundamental attributes of **cryptography**: **substitution** (also called confusion) and **transposition** (also called diffusion).
- DES consists of 16 steps, each of which is called a **round**. Each round performs the steps of substitution and transposition.

Broad-level steps in DES

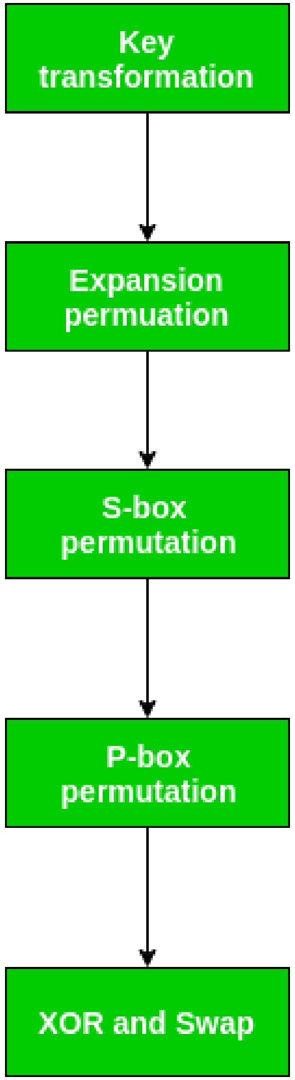
- The 64-bit plain text block is handed over to an initial Permutation (IP) function.
- The initial permutation is performed on plain text.
- Next, the initial permutation (IP) produces two halves of the permuted block; saying Left Plain Text (LPT) and Right Plain Text (RPT).
- Now each LPT and RPT go through 16 rounds of the encryption process.
- In the end, LPT and RPT are rejoined, and a Final Permutation (FP) is performed on the combined block
- The result of this process produces 64-bit ciphertext.



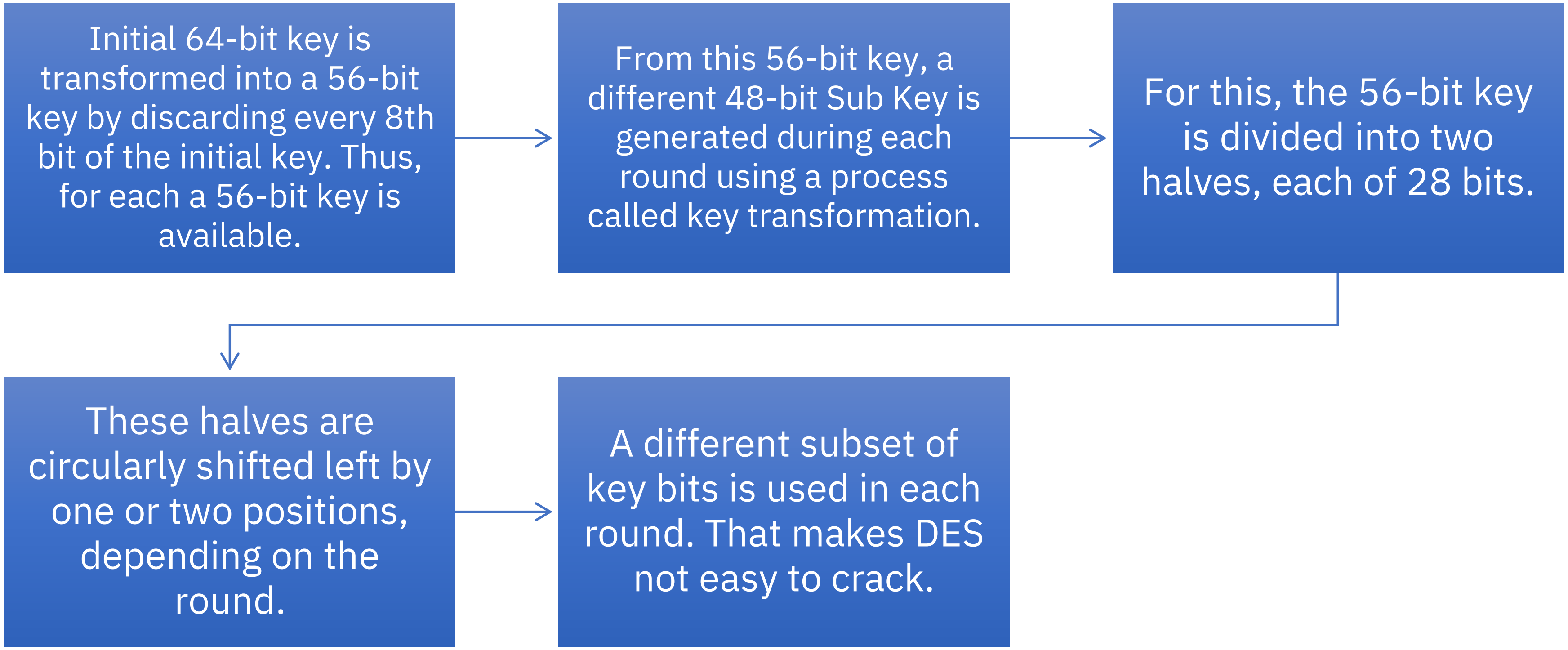
Broad-level steps in DES

Initial Permutation (IP):

- Initial permutation (IP) happens only once, and it happens before the first round.
- It suggests how the transposition in IP should proceed. This is nothing but jugglery of bit positions of the original plain text block.
- As we have noted after IP is done, the resulting 64-bit permuted text block is divided into two half blocks. Each half-block consists of 32 bits, and each of the 16 rounds, in turn.



Step 1: Key transformation



Initial 64-bit key is transformed into a 56-bit key by discarding every 8th bit of the initial key. Thus, for each a 56-bit key is available.

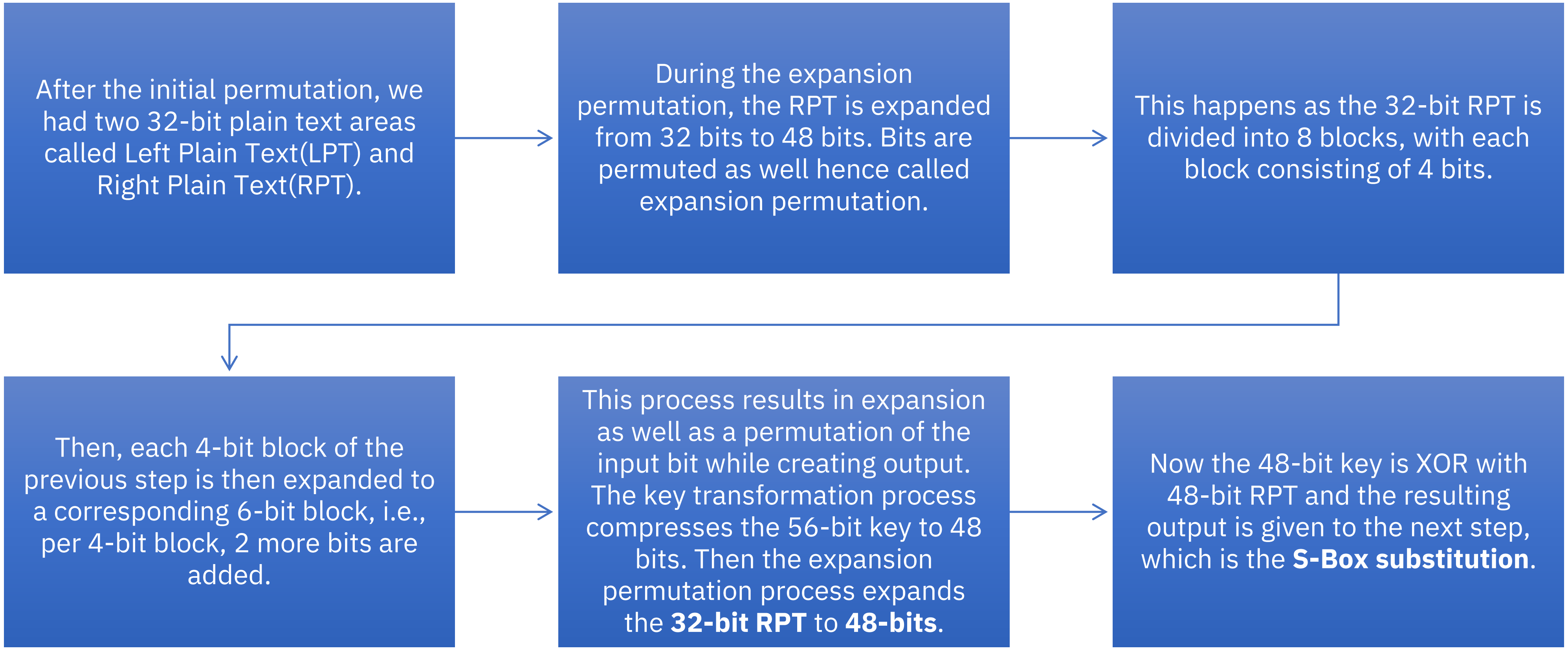
From this 56-bit key, a different 48-bit Sub Key is generated during each round using a process called key transformation.

For this, the 56-bit key is divided into two halves, each of 28 bits.

These halves are circularly shifted left by one or two positions, depending on the round.

A different subset of key bits is used in each round. That makes DES not easy to crack.

Step 2: Expansion Permutation



After the initial permutation, we had two 32-bit plain text areas called Left Plain Text(LPT) and Right Plain Text(RPT).

During the expansion permutation, the RPT is expanded from 32 bits to 48 bits. Bits are permuted as well hence called expansion permutation.

This happens as the 32-bit RPT is divided into 8 blocks, with each block consisting of 4 bits.

Then, each 4-bit block of the previous step is then expanded to a corresponding 6-bit block, i.e., per 4-bit block, 2 more bits are added.

This process results in expansion as well as a permutation of the input bit while creating output. The key transformation process compresses the 56-bit key to 48 bits. Then the expansion permutation process expands the **32-bit RPT** to **48-bits**.

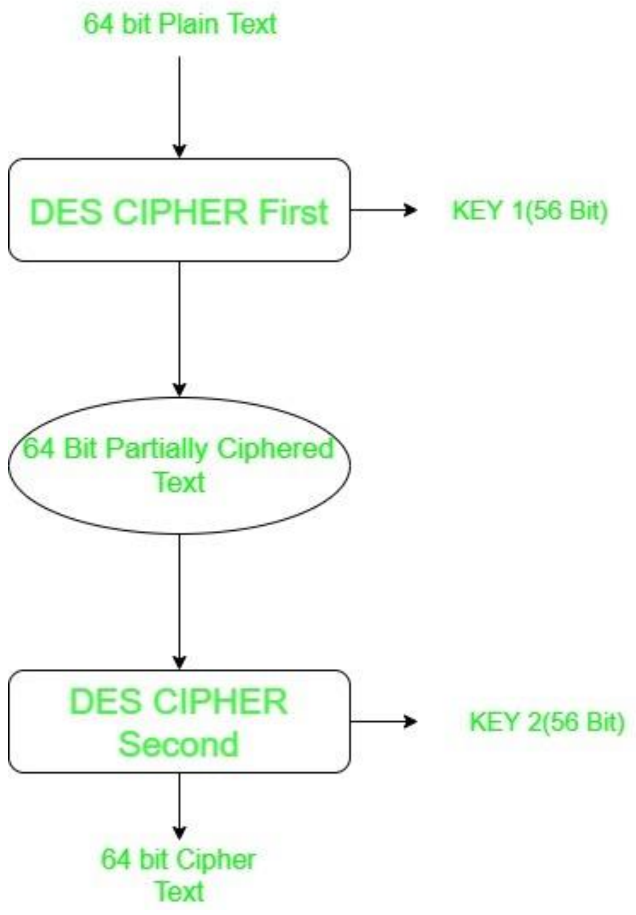
Now the 48-bit key is XOR with 48-bit RPT and the resulting output is given to the next step, which is the **S-Box substitution**.

Double DES and Triple DES

- The Data encryption standard (DES) uses 56 bit key to encrypt any plain text which can be easily be cracked by using modern technologies.
- To prevent this from happening double DES and triple DES were introduced which are much more secured than the original DES because it uses 112 and 168-bit keys respectively.
- They offer much more security than DES.

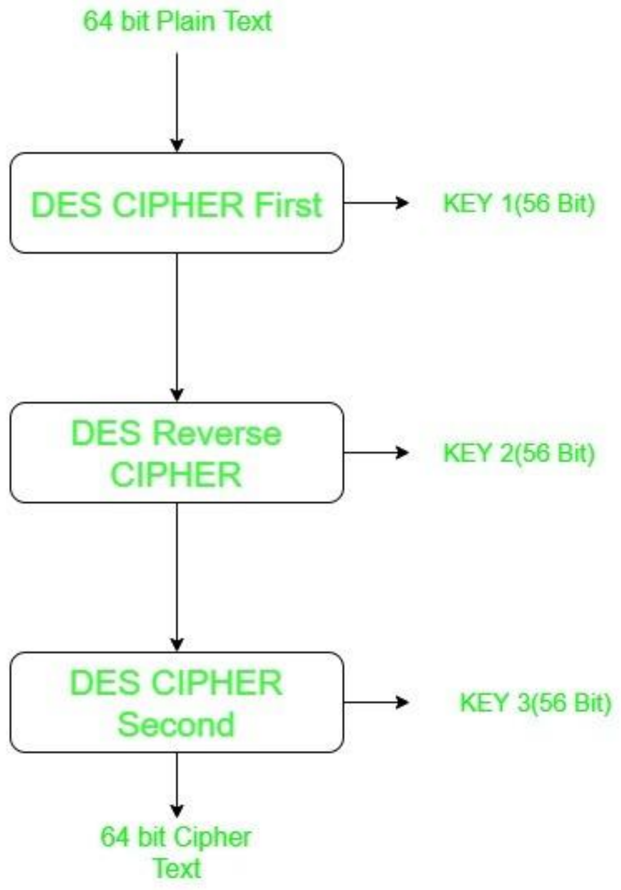
Double DES

- Double DES is an encryption technique which uses two instance of DES on same plain text.
- In both instances it uses different keys to encrypt the plain text. Both keys are required at the time of decryption.
- The 64-bit plain text goes into first DES instance which then converted into a 64-bit middle text using the first key and then it goes to second DES instance which gives 64-bit cipher text by using second key.
- However double DES uses 112 bit key but gives security level of 2^56 not 2^112 and this is because of meet-in-the middle attack which can be used to break through double DES.



Triple DES

- Triple DES is an encryption technique which uses three instance of DES on same plain text.
- It uses their different types of key choosing technique in first all used keys are different and in second two keys are same and one is different and in third all keys are same.
- Triple DES is also vulnerable to meet-in-the middle attack because of which it give total security level of 2^112 instead of using 168 bit of key.
- The block collision attack can also be done because of short block size and using same key to encrypt large size of text.



Advanced Encryption Standard (AES)

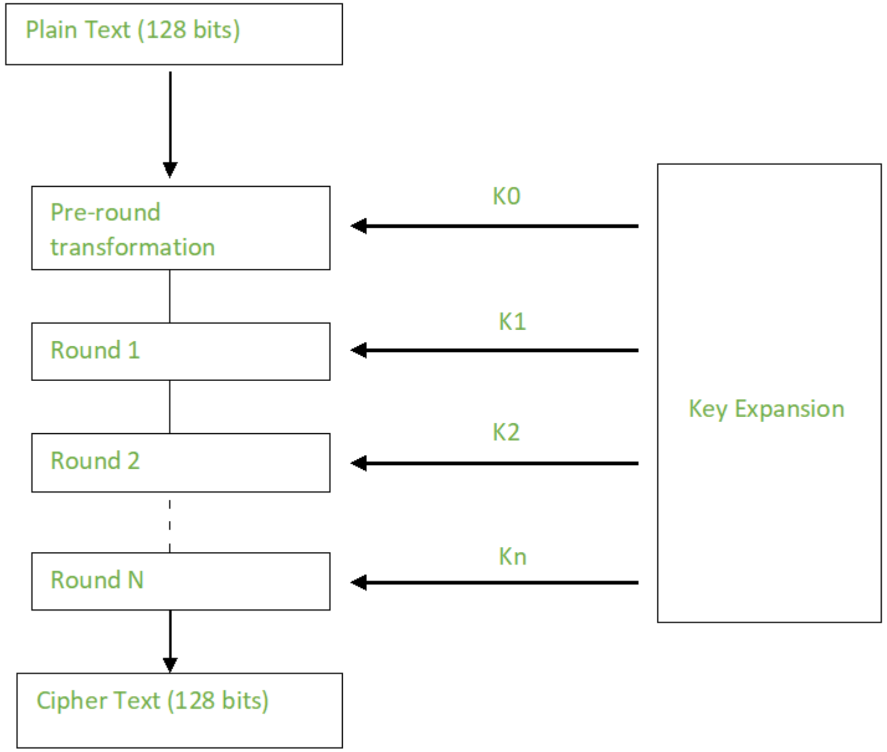
- Advanced Encryption Standard (AES) is a specification for the encryption of electronic data established by the U.S National Institute of Standards and Technology (NIST) in 2001.
- AES relies on substitution-permutation network principle which means it is performed using a series of linked operations which involves replacing and shuffling of the input data.
- The key size can be 128/192/256 bits.
- Encrypts data in blocks of 128 bits each.
- AES is widely used today as it is a much stronger than DES and triple DES despite being harder to implement.

Working of the cipher

- AES performs operations on bytes of data rather than in bits. Since the block size is 128 bits, the cipher processes 128 bits (or 16 bytes) of the input data at a time.
- The number of rounds depends on the key length as follows :
  - 128 bit key – 10 rounds
  - 192 bit key – 12 rounds
  - 256 bit key – 14 rounds

Creation of Round keys

- A Key Schedule algorithm is used to calculate all the round keys from the key.
- So, the initial key is used to create many different round keys which will be used in the corresponding round of the encryption.



Encryption In AES

- AES considers each block as a 16-byte (4-byte x 4-byte = 128) grid in a column major arrangement.
- Each round comprises of 4 steps :
  - SubBytes
  - ShiftRows
  - MixColumns
  - Add Round Key

Encryption In AES

1. **SubBytes :**
   - In this step each byte is substituted by another byte.
   - It is performed using a lookup table also called the S-box.
   - This substitution is done in a way that a byte is never substituted by itself and not substituted by another byte which is a compliment of the current byte.
2. **ShiftRows :**
   - This step each row is shifted a particular number of times.

Encryption In AES

3. MixColumns :

- This step is basically a matrix multiplication. Each column is multiplied with a specific matrix and thus the position of each byte in the column is changed as a result.
- **This step is skipped in the last round.**

4. Add Round Keys:

- Now the resultant output of the previous stage is XOR-ed with the corresponding round key. Here, the 16 bytes is not considered as a grid but just as 128 bits of data.
- After all these rounds 128 bits of encrypted data is given back as output.
- This process is repeated until all the data to be encrypted undergoes this process.

Decryption In AES

- The stages in the rounds can be easily undone as these stages have an opposite to it which when performed reverts the changes.
- Each 128 blocks goes through the 10,12 or 14 rounds depending on the key size.
- The stages of each round in decryption is as follows :
    - Add round key
    - Inverse MixColumns
    - ShiftRows
    - Inverse SubByte

Decryption In AES

Inverse MixColumns :

- This step is like the MixColumns step in encryption but differs in the matrix used to carry out the operation.

Inverse SubBytes :

- Inverse S-box is used as a lookup table and using which the bytes are substituted during decryption.

Applications Of AES

- Wireless security
- Database Encryption
- Secure communications
- Data storage
- Virtual Private Networks (VPNs)
- Secure Storage of Passwords
- File and Disk Encryption