Ethical Hacking
Prof. Indranil Sengupta
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

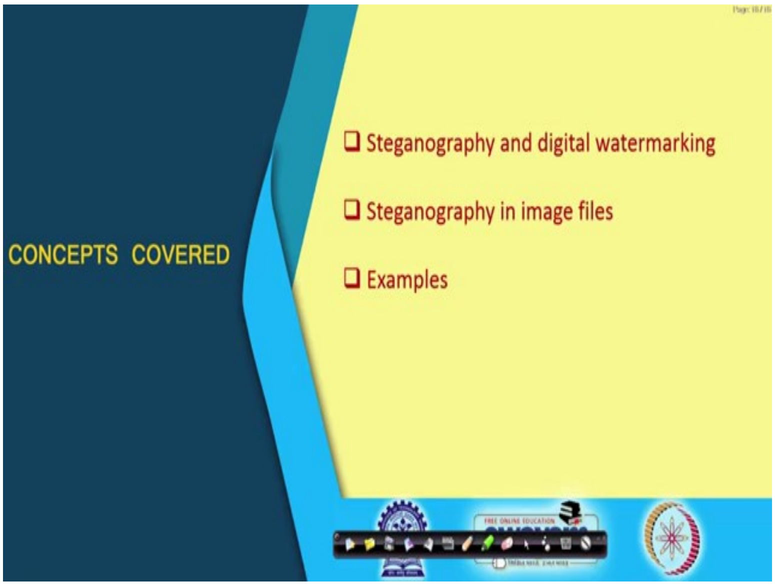
Lecture - 36
Steganography

In this lecture we shall be talking about something called Steganography, ok. So, the title of the lecture is Steganography. Now, let us first try to understand what we are trying to discuss here. Now, we talked about networks; we talked about some security issues and security protocols that are typically used over a network to secure our data to achieve confidentiality, authenticity, integrity and various others. Now, here we are talking about our own data, our own communication; we are trying to secure them, but to imagine we are now living in a world where so many different kinds of communication are going on over the network, over the public network. Well, we are securing our own network, alright.

But, there are some other people which may not be that innocent, may not be that good who are also communicating among themselves with some malicious intent. So, we should be on the alert; we should also be, try to find out whatever is going on; what the others are sending and receiving over the network. This is just from the point of view of securing our, you can say infrastructure and in a much larger context securing our nation, ok.

So, if you can detect some communication that is going on, then you can take some action, alright. But, steganography is a technique which is meant to hide some information so that you cannot detect them so easily. Let us look at it.

(Refer Slide Time: 02:11)

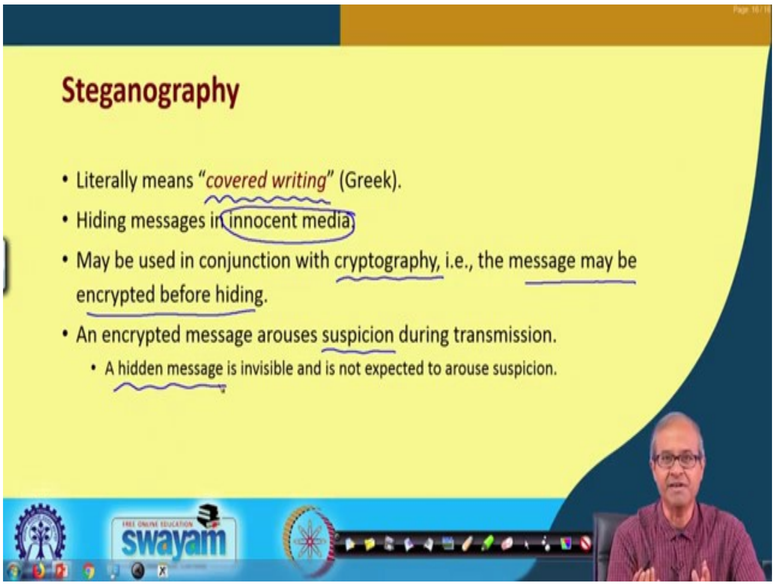


CONCEPTS COVERED

- ❑ Steganography and digital watermarking
- ❑ Steganography in image files
- ❑ Examples

Now, in this lecture we shall first be talking about steganography and digital watermarking. What these are? There, these two are very similar things, but maybe they are, they used in slightly different context and steganography can be used in many different kinds of files. We shall specifically be looking at some examples with image files and look at some examples.

(Refer Slide Time: 02:39)



Steganography

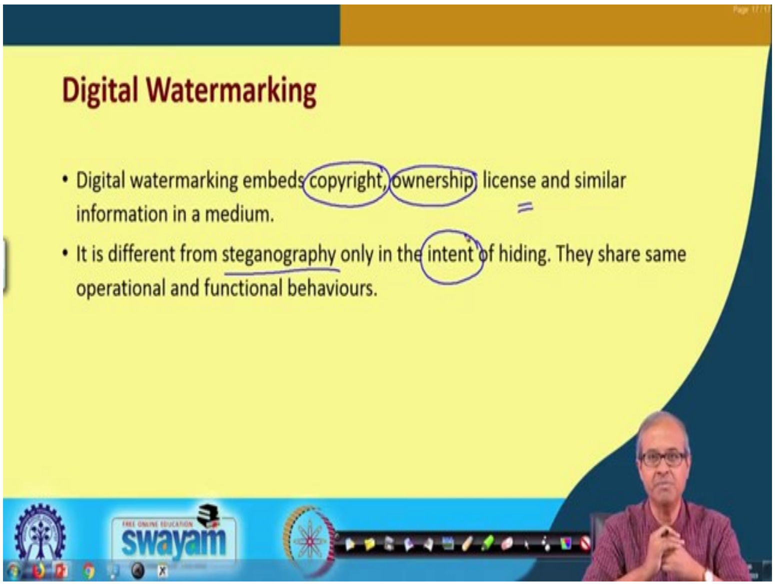
- Literally means "*covered writing*" (Greek).
- Hiding messages in innocent media.
- May be used in conjunction with cryptography, i.e., the message may be encrypted before hiding.
- An encrypted message arouses suspicion during transmission.
  - A hidden message is invisible and is not expected to arouse suspicion.

First let us look at what steganography means? Steganography as for the definition if you look at the dictionary meaning, it means covered writing. It has come from a Greek word. Covered writing means you have written something, but you are hiding it; no one else is able to see it. This is the basic idea. Hiding messages in innocent media. Innocent media means generally you will not have any idea at all that there is some, there is a hidden message that is being carried. But somehow that hidden message is hiding inside an apparently innocent media. It can be a very nice picture; you are viewing a very nice picture, but inside that picture somewhere some information has been hidden, ok.

This is steganography, ok. Steganography may or may not be used in conjunction with cryptography. So, if the person who is hiding the message is doubly, you can say aware of the fact that he also wants to do some encryption before hiding so that if someone at all breaks it, if someone finds out that the steganography and extracts the information, but still, because this encryption that person will not be able to decode it. So, you may possibly encrypt the message before hiding. That is a more sophisticated way of data hiding.

Now, the thing is that in a normal channel if you send an encrypted message and if I capture it, that if I am not able to read it; that means, it must be encrypted. So, the first thing is that there be some suspicion; that means, what are they sending among each other that they have to encrypt. Are they good people or they are not so good people? So, I do not know, ok. So, any message flowing through in an encrypted form may raise suspicion, ok. So, the other alternative to be, is to send in such a way that no one will suspect at all. The message will be hidden. There will be nothing visible which can attract any suspicion, ok. This is the idea.

(Refer Slide Time: 05:13)



Digital Watermarking

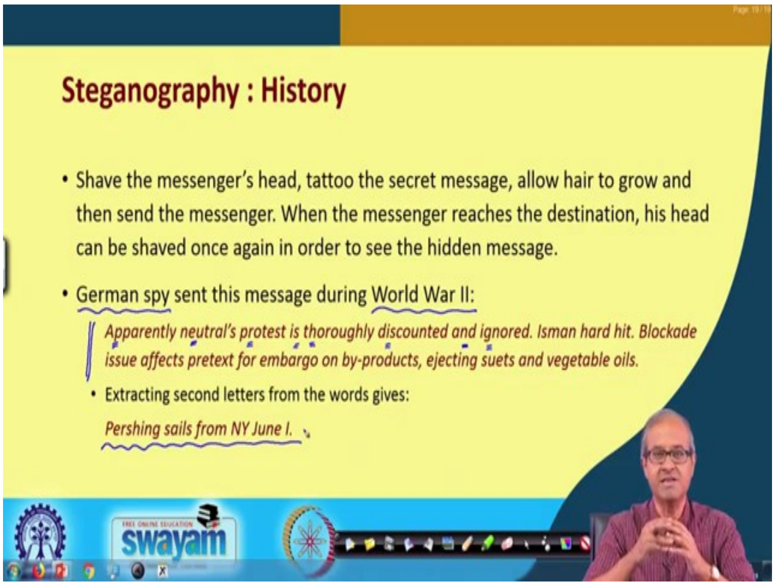
- Digital watermarking embeds copyright, ownership, license and similar information in a medium.
- It is different from steganography only in the intent of hiding. They share same operational and functional behaviours.

Digital watermarking, as I said is something which is very similar to steganography; here also we are hiding something, but here there is a particular application. Why we are hiding? Here let us say, we can have some copyright information like for example, some company produces some music, produces some music CDs, some movies on some media, CD, DVDs, Blu-rays and whatever various medias are there, online media. Now, there can be some copyright or ownership information. There can be licenses and other information that may be hidden into that media which the person who is downloading and listening will not be able to see.

But, if a person makes unauthorized copies, then the law enforcement agencies can check that and try to find out who was the original owner and whether it is a legitimate copy or not, ok. This is different from steganography. Only with the intent, normally we use the term steganography with the intent that something wrong is trying to happen.

Someone is trying to send something without the others knowing; that means, they do not want to share it with others, maybe some very secret information or very you can say not so good something malicious is going on, but in digital watermarking we are trying to protect the copyright of some information which are generated by some authorized parties, but the technologies are same, similar, ok.

(Refer Slide Time: 07:09)



Steganography : History

- Shave the messenger's head, tattoo the secret message, allow hair to grow and then send the messenger. When the messenger reaches the destination, his head can be shaved once again in order to see the hidden message.
- German spy sent this message during World War II:

  Apparently neutral's protest is thoroughly discounted and ignored. Isman hard hit. Blockade issue affects pretext for embargo on by-products, ejecting suets and vegetable oils.

  - Extracting second letters from the words gives:

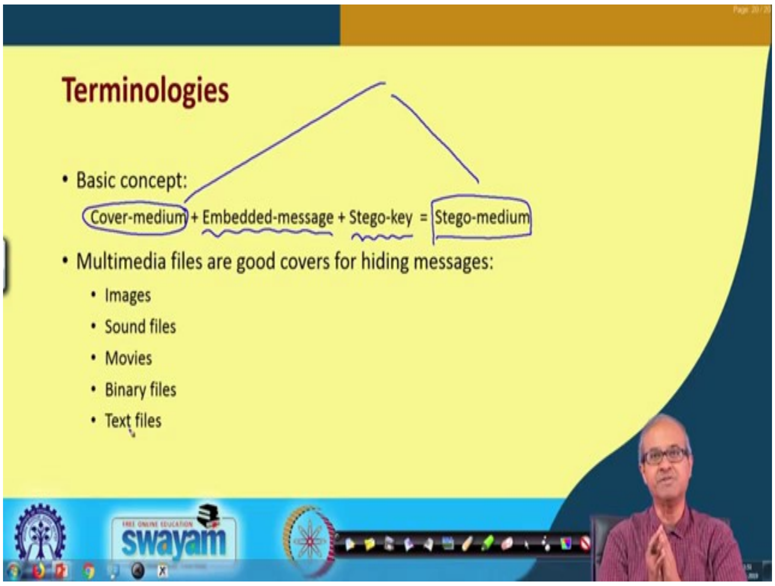
    Pershing sails from NY June I.

Let us look at the history of steganography and look at some of the very classical examples. Well, the very first example of steganography was very interesting. So, what happened? Some very secret information needed to be sent from one place to another, long distance. So, what was done, a messenger who was carrying the message; so, he was not carrying the message on a piece of paper. So, if the messenger gets captured, then the paper will also be seen. So, what was done? The head of the messenger was shaved and some tattoo or some information with permanent ink was marked on the head. Then some days were given so that the hair can again grow and then the messenger was sent.

So, apparently even if the messenger is caught, no one will suspect that he is carrying some secret information on his head. So, when he reaches the intended destination, his head will be shaved again and the hidden message can be read out. So, this was actually used in ancient times, ok and during World War II, some very simple type of this kind of steganography was used for some German spy to send a message from one post to another. So, the idea was very simple. He was carrying a printed message on a piece of paper, but the content of the message was apparently very, you can say straightforward, nothing suspicious.

Let us say "*apparently neutral's protest is thoroughly discounted and ignored*". Isman hard hit. Blockade well, you cannot make anything out of it, but the sender and receiver knows how the information was hidden. If we extract the second letter of each word like I extract this $p\ e\ r\ s\ h\ i\ n\ g$ and so on then you extract something like this. $Pershing$ is, maybe the name of a ship it sails from New York June 1. So, this is an information which you are sending to someone else so that, that person may be able to attack or do something based on that, ok. So, this was one of the very old form of steganography, ok.

(Refer Slide Time: 09:59)



Terminologies

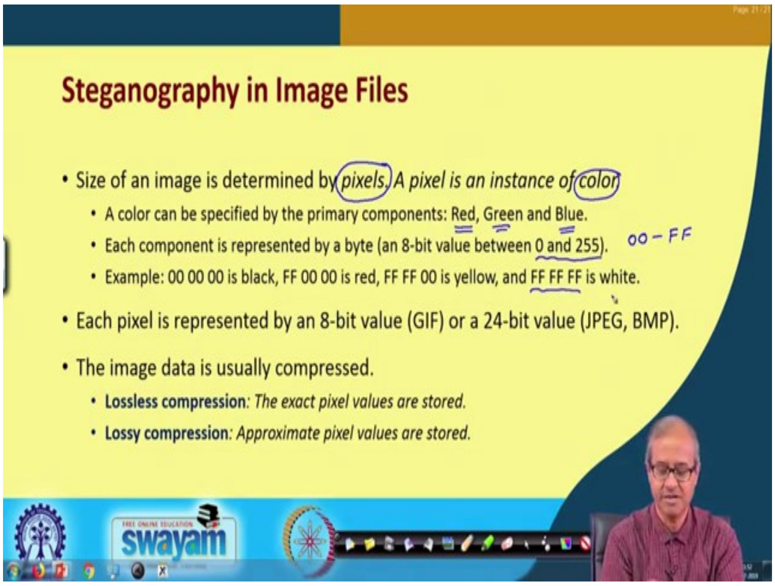
- Basic concept:
  - Cover-medium + Embedded-message + Stego-key = Stego-medium
- Multimedia files are good covers for hiding messages:
  - Images
  - Sound files
  - Movies
  - Binary files
  - Text files

We use some terminologies in connection with steganography. We talk about a cover medium that where we are hiding; if it is a text that is our cover medium; if it is an image that is our cover medium; if it is an audio clip or a music file that is our cover medium. So, where we are hiding, ok. Then the embedded message and if you are using encryption in addition some kind of, then there has to be some key also. Let us call it steganographic key or stego-key and all 3 combined together will generate this stego-medium.

Now, the objective is this. Cover medium and stego-medium should look very similar. A person who does not know that this is carrying some hidden message will not be able to distinguish just by looking at this stego-medium. That person will feel that he is actually looking at the original file only. Multimedia files are typically used to hide messages like images, sound, movies, binary files, text files, but the other ways also. Like you see, you have seen the IP packets, network packets, there are some fields which are unused. Those unused bits can also be used to carry some message or information.

So, there are many ways, there are many fields in data, in files which are not used; you can also use those fields to carry some data. Like some executable files. You compile a program in C or C++ or Fortran, whatever. You generate machine code. Now, in the machine code, in the header, there are lots of fields which are not used. You can also use those fields to carry some hidden information.

(Refer Slide Time: 12:07)



Steganography in Image Files

- Size of an image is determined by pixels. *A pixel is an instance of color.*
  - A color can be specified by the primary components: Red, Green and Blue.
  - Each component is represented by a byte (an 8-bit value between 0 and 255). 00 – FF
  - Example: 00 00 00 is black, FF 00 00 is red, FF FF 00 is yellow, and FF FF FF is white.
- Each pixel is represented by an 8-bit value (GIF) or a 24-bit value (JPEG, BMP).
- The image data is usually compressed.
  - **Lossless compression**: *The exact pixel values are stored.*
  - **Lossy compression**: *Approximate pixel values are stored.*

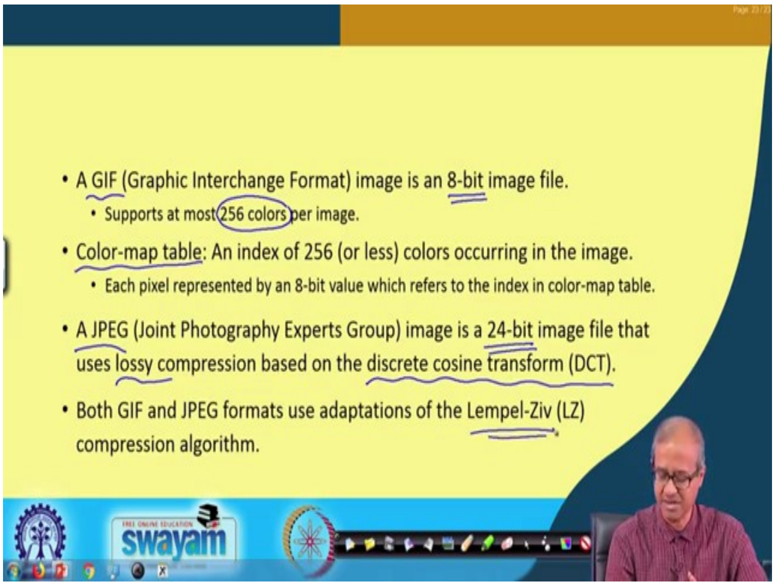
Now, specifically as I said we shall be talking about steganography in image files. Now, a few terminologies, the size of an image, we talk about an image in terms of picture elements or pixels. How many dots are there? We imagine that a picture is made out of dots; let us say, we say 1024×1024. There are 1024 dots horizontally and 1024 dots vertically. That will define the size of my image and each picture element or pixel will be having a color. That is, have an image is formed and a color is typically specified by some combination of the fundamental colors red, green, blue. You know these three are fundamental colors RGB, as combination of Red, Green and Blue.

Typically, each of this red, green and blue can be represented in 8-bits or a byte. In 8 bits the value can between 0 and 255. Now, in hexadecimal, it will be 0 to FF. Let us say black; the color black where red, green, blue all these components will be 0. So, it can be encoded as 00 00 00, red, green, blue; let us say white, white means red, green, blue all are at the maximum intensity. So, to the FF FF FF that is white; if you talk about pure red, only the red component is FF. Green and blue at 0, ok. If you talk of yellow, red and green combined, makes yellow. Blue is 0. So, in this way you can generate any possible colors right.

Now, it depends on what kind of image format you use. For example, in the GIF, GIF format which is very popular. Here each pixel is typically represented by an 8-bit value. But, in JPEG or BMP formats, we use 24 bit value for each pixel and this image data, well, there are image formats which are uncompressed, but normally they are in compressed. GIF, JPEG these are all compressed. Compressed form they are stored. Now, when you are doing compression, there are two terminologies, lossless compression, lossy compression.

Lossless means you are compressing alright, but if required you can uncompress it and get back the original image. There is no loss of information anywhere. So, the exact pixel values can be generated somehow, but in lossy compression, JPEG is an example of lossy, lossy compression where you are trying to reduce the size of your image file and you cannot get back the original. You are reducing the quality of your image in some sense. So, only approximate pixel values are stored, ok.

(Refer Slide Time: 15:37)



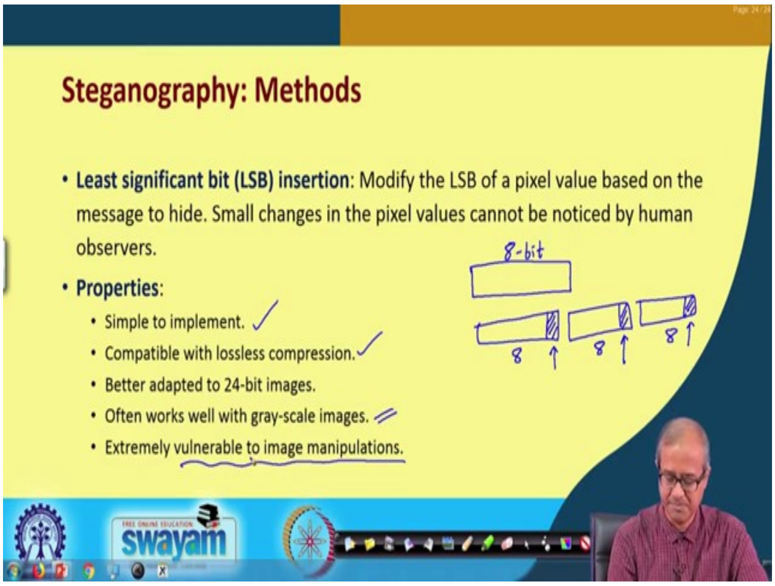
- A GIF (Graphic Interchange Format) image is an 8-bit image file.
  - Supports at most 256 colors per image.
- Color-map table: An index of 256 (or less) colors occurring in the image.
  - Each pixel represented by an 8-bit value which refers to the index in color-map table.
- A JPEG (Joint Photography Experts Group) image is a 24-bit image file that uses lossy compression based on the discrete cosine transform (DCT).
- Both GIF and JPEG formats use adaptations of the Lempel-Ziv (LZ) compression algorithm.

So, you should understand these distinctions. Now, as I said, in a GIF image format which is Graphic Interchange Format where each color is encoded in 8 bits. In each in 8-bits you can represent 256 colors. Now, you may say that 256 colors is not sufficient.

There is a concept of a color map table. In GIF you can change from 1 color map table to another. Each color map table will support 256 colors, ok. So, the color map table, this 8-bit value which is stored, it will refer to one of the entries in the color map table. From there you can extract color from the color map table and that color can be encoded in 24-bits also, ok.

JPEG as it said, JPEG uses 24-bit color format and that uses lossy compression, because, internally it uses discrete cosine transform to mix an transformations, compression and then it generates the final JPEG file and during that you can also adjust the quality of the final image and they use the compression algorithm LZ, Lempel-Ziv compression algorithm, ok. These are not very important to remember.

(Refer Slide Time: 17:07)



Steganography: Methods

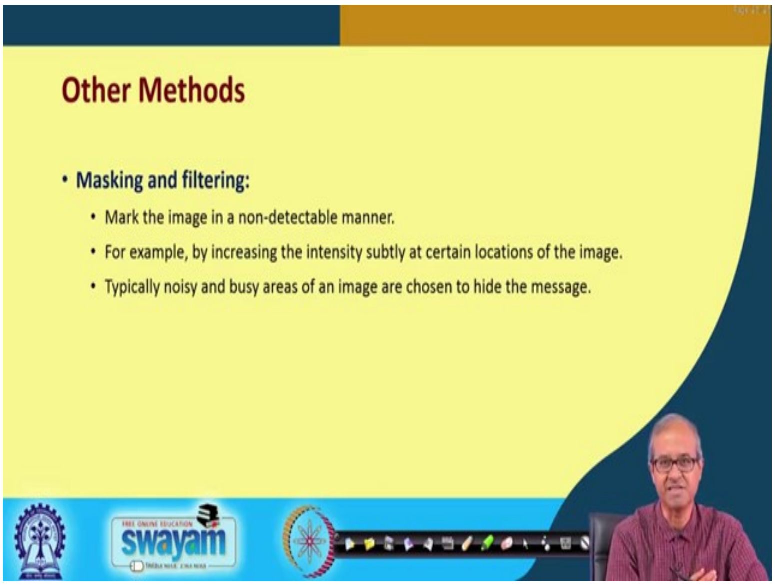
- **Least significant bit (LSB) insertion**: Modify the LSB of a pixel value based on the message to hide. Small changes in the pixel values cannot be noticed by human observers.
- **Properties**:
  - Simple to implement. ✓
  - Compatible with lossless compression. ✓
  - Better adapted to 24-bit images.
  - Often works well with gray-scale images.
  - Extremely vulnerable to image manipulations.

Now, talking about steganography which you are more interested in here. One of the very simple methods using which you can hide information in an image, well, not necessarily an image, in any media file in fact, is called Least significant bit insertion or modification. The idea is as follows; you see in GIF image format I said you use 8 bits to store a color, in JPEG you use 24-bits; red, green, RGB: Red, Green, Blue, 8-bits, 8-bits, 8-bits. Then LSB it says, let us say in JPEG, it says you can modify the least significant bit of these colors, you see the color can be from 0 to 255, right, each of the components RGB.

So, if you change the LSB; that means, I am either increasing it by 1 or decreasing it by 1 then a very small difference in the color, in the actual image; you will not be able to distinguish, if you see it with your eye. So, what if I put in some secret information in these places? Suppose I have a message. I want to send. It requires 1000 bits. So, I use 1000 of these pixel bytes. I put these 1000 bits in the least significant bit positions. Someone, when he or she views the image, you will not find any difference. The image looks exactly the same, but in the LSB positions some other information is hidden. Well, you can use either 1 LSB or 2 LSBs, if you want. There will be a little bit degradation in quality of the image, but you can hide like this, right.

Now, properties that this is very simple to implement. Lossless compression, this is most suitable, because, if you do a compression, lossy compression, then the LSBs positions might get lost; you cannot use LSB. Steganography with JPEG images for example, right. JPEG is a lossy compression, 24-bit images, it is easier because you have means, 1 bit. You take up for each of the colors. It will be very small changes in the final color value. Grayscale images where you are not talking of colors, but black and white and shades of black gray, these images work and as I told you, these are vulnerable to image manipulations; when you transform an image, when you do compression, decompression, lossy compression. These LSBs will all get lost. So, this hidden method might get easily lost, if you do some kind of media manipulation.

(Refer Slide Time: 20:17)

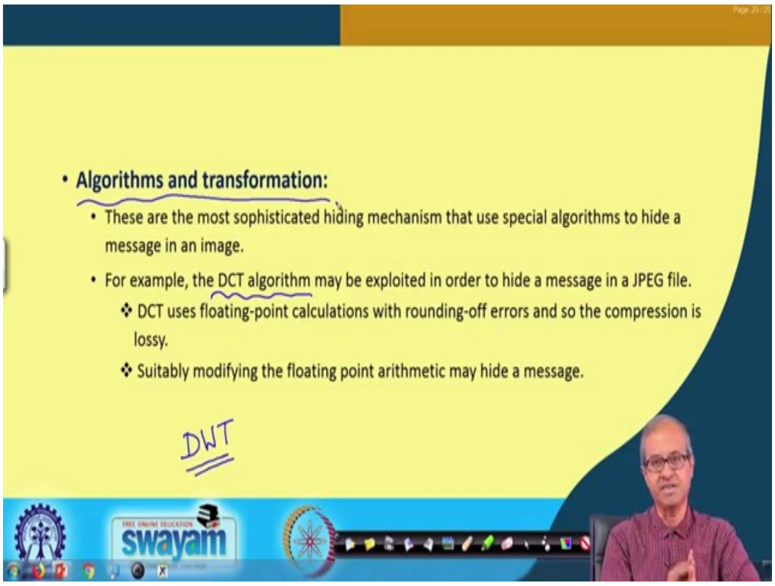


Other Methods

- Masking and filtering:
  - Mark the image in a non-detectable manner.
  - For example, by increasing the intensity subtly at certain locations of the image.
  - Typically noisy and busy areas of an image are chosen to hide the message.

Well, there are some other methods also masking and filtering. Like, you can mark the image in a way so that it is very difficult to detect. Like, you do not modify all the pixel in the image. Rather, only certain locations in the image you modify the image. It change the intensity a little bit. Usually, the busy areas of image, there may be some part which is the same color. Like, the sky, sky is totally blue, do not modify that part, because even a small modification can be easily visible, but there is a house, there is a man, there is a flower, lot of things are there; if you make some small changes, there it will not be easily visible. So, here we talk about that.

(Refer Slide Time: 21:03)



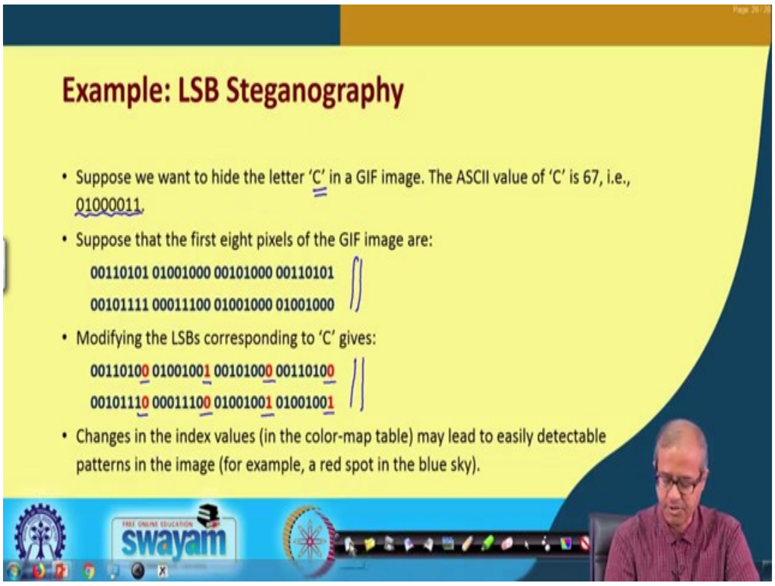
- **Algorithms and transformation:**
  - These are the most sophisticated hiding mechanism that use special algorithms to hide a message in an image.
  - For example, the DCT algorithm may be exploited in order to hide a message in a JPEG file.
    - ❖ DCT uses floating-point calculations with rounding-off errors and so the compression is lossy.
    - ❖ Suitably modifying the floating point arithmetic may hide a message.

DWT

Now, there are methods which are more sophisticated, where you use some algorithms and transformations before you can hide. So, you do not directly hide the data in the image or in the sound file or in the video file, in the LSB directly, but you carry out some kind of a transformation. Like discrete cosine transformation is one popular method using which you can hide a message in a JPEG file, but not directly in a JPEG file; you have to use DCT.

Then there can be Discrete Wavelet Transform: DWT is a very popular method that is used in the context of steganography. You can use it to hide information in different types of media, right. So, these algorithms and transmissions are more sophisticated techniques where even in lossy compression methods, you can store some information, hide some information and there will be some degree of robustness against transformation. Like earlier I told you LSB transformation or means LSB steganography, the data can be easily destroyed.

(Refer Slide Time: 22:31)



Example: LSB Steganography

- Suppose we want to hide the letter 'C' in a GIF image. The ASCII value of 'C' is 67, i.e., 01000011.
- Suppose that the first eight pixels of the GIF image are:

```
00110101 01001000 00101000 00110101
00101111 00011100 01001000 01001000
```

- Modifying the LSBs corresponding to 'C' gives:

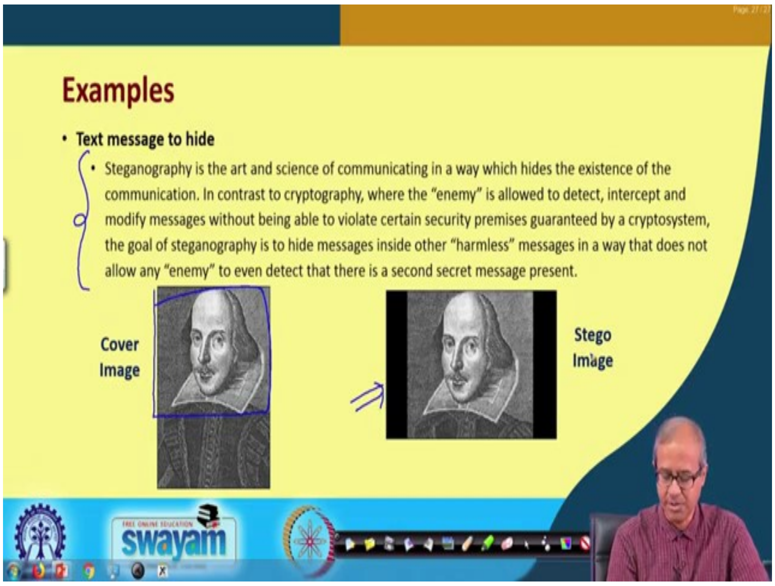
```
00110100 01001001 00101000 00110100
00101110 00011100 01001001 01001001
```

- Changes in the index values (in the color-map table) may lead to easily detectable patterns in the image (for example, a red spot in the blue sky).

But, here even if you do compression and decompression the data may not be easily destroyable, ok. These are more robust. Let us take a very small example of LSB steganography. Let us say, we want to hide a single letter C, just one letter. We take an example, in a GIF image and C in terms of ASCII code, in decimal is 67, in binary it is this is 01000011; 43 in hex, 67 in decimal. It is alright. So, let us suppose the GIF image will be a large image. Here we have to hide 8 bits.

So, let us look at 8 bytes in the GIF image, the first 8 bytes. Let us suppose, are like this. What do we do? We modify the least significant bits to hide this. 01000011, the bits are marked in red. Here you see 01000011. This is how we are modifying the LSB to hide this letter C. So, in this way you can hide the other letters to store the entire message you want to carry in the image, fine.

(Refer Slide Time: 23:45)



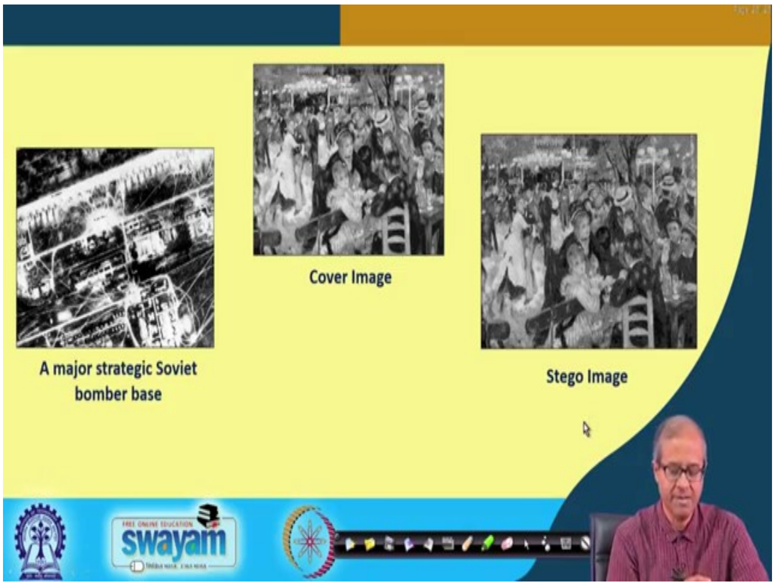
Examples

- Text message to hide
  - Steganography is the art and science of communicating in a way which hides the existence of the communication. In contrast to cryptography, where the "enemy" is allowed to detect, intercept and modify messages without being able to violate certain security premises guaranteed by a cryptosystem, the goal of steganography is to hide messages inside other "harmless" messages in a way that does not allow any "enemy" to even detect that there is a second secret message present.

Let us take some examples here. Now, will you see, this is one image. You can see, this is a picture which is the covered image and inside that cover image this entire text is hidden and a modified image is also generated. You see; of course, you can generate the whole image also but here only a part of this is generated now. There is no problem.

You say, apparently you do not see much difference, while maybe the quality has been degraded a little bit, but unless you also have the original image with you side by side, you cannot really know that whether there was in a degradation or not. But if someone gives you this picture, because well this is a fine picture, ok. I can recognize this person, but this image is, this is the stego-image. This is hiding some information inside it. You will not be able to detect that easily.

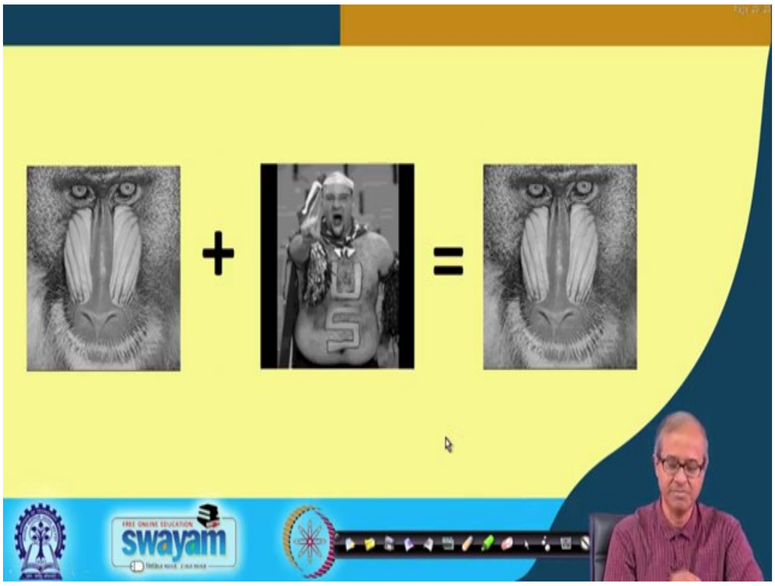
(Refer Slide Time: 24:55)



This is another example which is which was publicized. It is available in the public domain. Here this is a famous piece of art. This was considered as the cover image. Inside this picture, some other picture was hidden. This was the map of a strategic Soviet bomber base. This was a satellite picture which was hidden in it. Any such picture means not ultimately a stream of 0s and 1s that was hidden inside this larger image.

And, after hiding you see this stego-image look like this. Well you do not see much difference. You can still identify that picture. Maybe you will feel that well this picture was not of a that good quality, may be little degraded, but fine, the same image. You can identify the image, but, you will not understand that such a critical information is hidden inside this image, ok. These are some examples.

(Refer Slide Time: 26:05)



Now, another example I am showing. Let us say this is a picture of a baboon, the face and this is the picture of a person you are hiding inside this and this is the final image. Just by looking at this picture you will not be able to understand anything that such a big image is being hidden here inside. So, this is what steganography is. It is rather easy to hide an image. But, suppose you are a person who is on the lookout of whether such secret message is being transmitted by some parties which are suspect, whom you feel they may be exchanging some information which are detrimental to your security.

Then you should be able to identify or try to identify whether such things are going on. This is a field of research called steg-analysis, steg-analysis, steganographic analysis, but it is very difficult. If the person who is hiding an information, is very intelligent. It will be quite difficult for anyone to understand whether some hidden information is being carried with a media file.

So, with this we come to the end of this lecture. In the next lecture we shall be looking at some other means of security which is also used quite popularly nowadays.

Thank you.