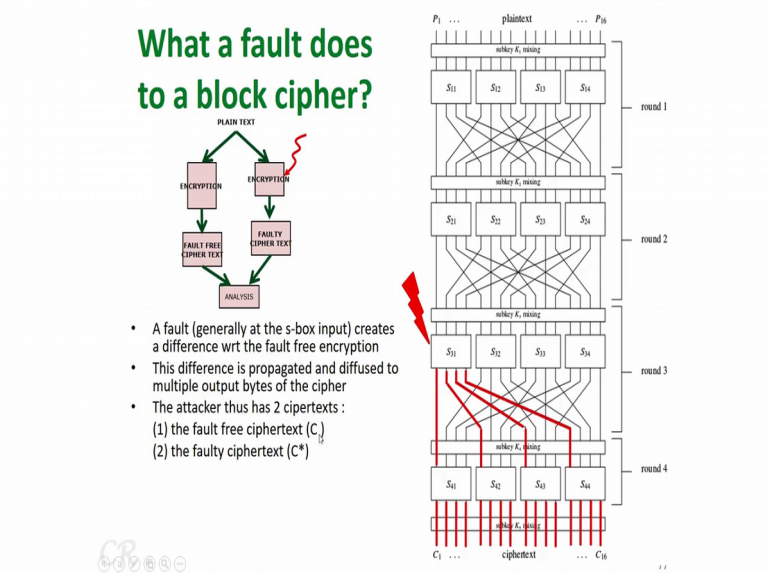
Information Security - 5 - Secure Systems Engineering
Professor Chester Rebeiro
Indian Institute of Technology, Madras
Fault Attacks on AES

Hello and welcome to this lecture in the course for secure systems engineering. In the previous lecture we had actually looked at an introduction to fault injection attacks and we had seen a very simple fall injection attack on the RSA public key cipher. In this video we would look at fault injection attacks on a popular block cipher known as AES. The assumption here is that the viewers know about this AES algorithm and the various operations that are involved with an AES encryption.

(Refer Slide Time: 0:53)



So let us look at how a fault attack on a block cipher actually works, so we assume that the attacker has a device which is doing an encryption like an AES encryption and in order to do this particular encryption the device has a key which is stored inside the device. Now the objective for the attacker is to extract this secret key from the device in order to do this the attacker would inject false as the cipher is actually doing an encryption or decryption. It is also assumed that the attacker is able to control what plane text gets encrypted and is also able to view the corresponding cipher text.

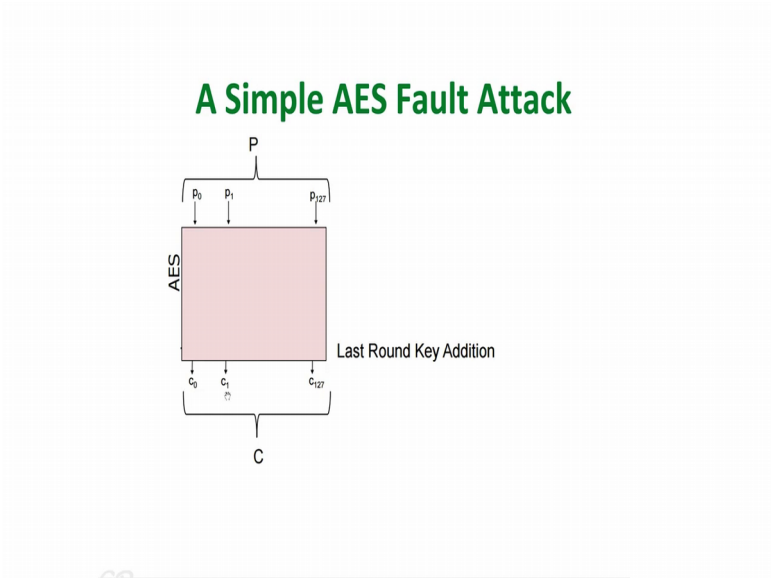
The basic attack as we have seen in the previous video is as following, attacker chooses a random plane texts passes it to the device and process the device to performance an encryption, the device would then pick the secret key which is stored inside the device perform an encryption on the plaintext using that particular secret key and obtain a cipher

text, we call the cipher text as the fault free cipher text. Now the attacker would use the same plane text and pass it to the device and force the device to do an encryption on that plain text, the device would encrypt that plain text using the same stored secret key and the plain text and during this encryption process the attacker now injects a fault.

Now the fault is injected and it would disturb the encryption process resulting in a faulty cipher text. The attacker then uses a fault free cipher text and the faulty cipher text to in some information about the secret key, so what happens when you actually inject a fault is that some computation during the process of encryption gets disturbed. Now if he actually think that a fault is injected say at this particular location in this particular block cipher, this fault modifies the output of this particular operation.

The error due to the fault then propagates to the remaining parts of the cipher, so these are red lines over here show how the fault propagates through the cipher structure and eventually effects all the bits of the cipher text thus the attacker has 2 cipher text the fault free cipher text and the faulty cipher text, so the faulty cipher text is denoted as C star and the quality and a fault free cipher text is denoted by C. Now the differences between these 2 are then used by the attacker to gain secret information.

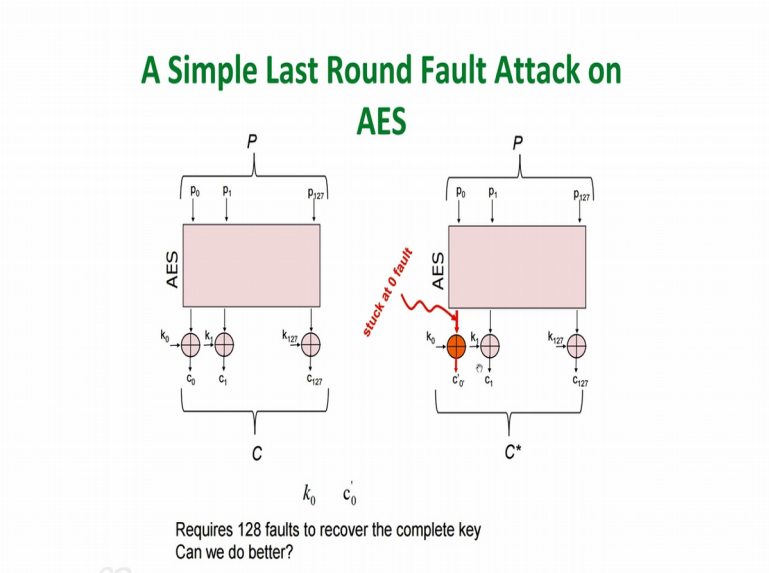
(Refer Slide Time: 3:55)



Let us take a very simple fault attack on the AES block cipher, so as many of you would know the AES is probably the most famous or more commonly used block cipher used these days. The AES takes 128 bit plane text as input which we label her as P and each bit is labelled P0 P1 to P127 it does various operations on this particular plane text, these

operations not only depend on the input but also on the secret key or the AES secret key and eventually after 10 rounds what we obtained is the cipher text C which comprises of bits C naught to C127.

(Refer Slide Time: 4:47)



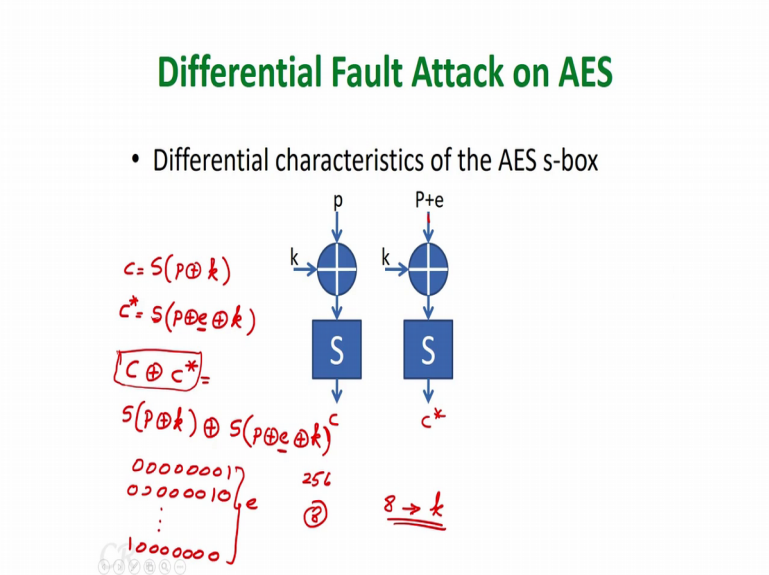
Now in order to demonstrate a simple attack on AES what an attacker needs to do is to inject a fault in exactly the last round of AES, so the attacker need not know what is happening during the entire AES but important for him is the last operation what is performed on AES. The last operation is as follows it has a key this is known as the 10^{th} round key and this 10^{th} around key is (())(5:16) with some intermediate state obtained from this during the encryption process to give you the cipher text, so thus what we will see in this attack is the attacker would be able to get one bit of this cipher text by injecting a fault.

Now the fault the attacker would inject is targeted to this specific location and the fault is very specific, this type of fault would force this particular line to be 0, so this is known as stuck at 0 fault and this is a very popular fault model especially from the (())(5:57) site testing perspective, so independent of what value was present whether it was 0 or 1 or so on, this fault would force this line to 0, as a result what we would see in the output is this value see hash 0 would always have the value of K naught, so therefore what we see is that secret key bit K naught is then passed to the output, so the attacker would just lead to look at these significant bit of cipher text to determining what the value of K naught is, so in this way the attacker has obtained one bit of the secret key.

In a similar approach by injecting false and all other bits the attacker would be able to obtain all the bits of the secret key. Now the problem with this attacker is extremely simple is the fact that you would require 128 bits to completely recover the entire AES secret key. Now 1st of all creating this bit fault is extremely difficult and furthermore we would require 128 such bit falls to completely extract the secret key of AES.

So in the past people have actually studied this problem and they have tried to find out solutions by which firstly we can reduce the number of faults that are required to obtain the secret key of the cipher and 2nd also relax default model. Instead of having a bit fault the researchers have tried to find out whether other kinds of faults such as random faults in bytes or multiple byte falls can be exploited to obtain the secret key, so let us see an attack which is slightly better than this one. What we will show next is that the attacker can reduce the number of faults required from 128 to 16 by just changing the location of where the fault is injected.

(Refer Slide Time: 8:18)



The central idea for this is to look at this structure, this structure comprises of some input which we denote P, P is not necessarily the plain text but it could be some intermediate state of the cipher. This gets (())(8:33) with the key and then there is s-box operation and then you have a cipher text output, now if the attacker is able to inject a fault at this location on P and change the value of the P to P plus e where e is the fault that is injected.

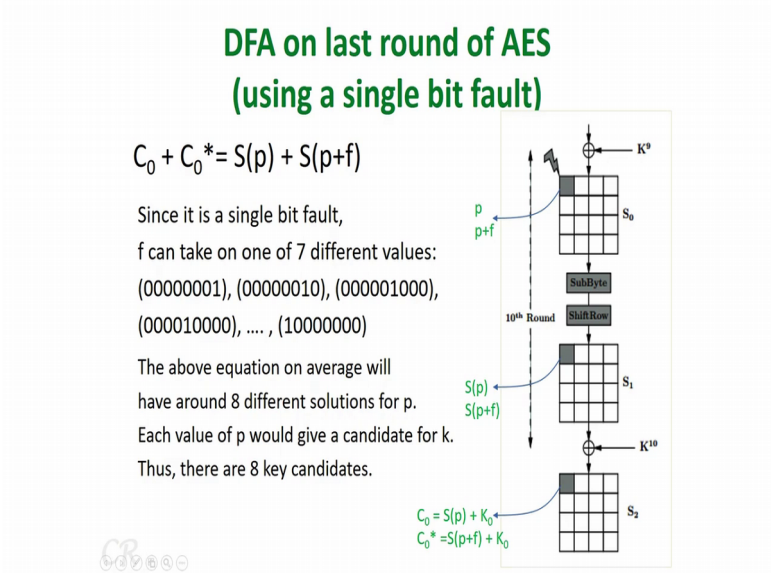
Now the output obtained would be erroneous, so in fact with this the attacker can get 2 equations one is the fault free equation, let us consider this output as C and this output due to

the fault getting induced at this point to be C star. What we definitely know is that since the fault is induced, C will definitely not be equal to C star, so we can represent this in an equation form as follows, C is equal to S of P XOR k similarly this will be equal to C star would be equal to S P XOR e XOR k.

Note that since e has a value which is not equal to 0 therefore we have an s-box operation based on this which would look completely different, so now let us look at what C XOR C star is and what we see is that we can represent this as S of p XOR k XOR S of p XOR e XOR k, since the attacker knows the faulty cipher text and fault free cipher text this particular component of the equation is known.

Now also what the attacker could then do is iterate to all possible values of k and identify which of these equations are actually satisfied, so let us say that e is a single bit fault and therefore e could have a value either 00000001 because it is in this case modifying the LSB bit or it could have values like 00000010 and so on to up to 10000000 thus there are 8 possible values for e, now for each of these possible values of e one can identify what is the solution for this equation and validate whether it is matching the LHS, so since this s-box or the s-function has 256 different values for AES the number of solutions for this particular equation reduces down to just 8. In other words what the attacker would require would end up over here is just 8 different values for this key k thus the attacker has reduced the key space for this from 256 different values of k to just 8 different values for k.

(Refer Slide Time: 12:18)



So let us see how this can be done in practice, now if we consider the last round of AES there are 3 operations one is substituted by followed by the shift row and then mix column, now if the attacker injects a fault at this particular location it would disturb exactly this byte, now this byte would actually be modified due to the substitute byte operation and substitute byte is essentially represented as S of p, so as a result of the fault the output of this substitute byte is S of p when there is no fault and S of p plus f when a fault is injected.

As this passes through and finally reaches the output this particular byte of the output is either S of p plus XOR or S of p plus f plus k, so based on this the attacker can build this equation C zero plus C zero star where C zero star is the faulty cipher text byte. C zero is the correct cipher text byte or as we call it the fault free cipher text byte and for each of the different values of the fault f the attacker would be able to iterate all possible values that satisfy this equation and then as the AES s-box is well known we would then be able to identify 8 values for p and therefore we would obtain 8 different candidate values for the secret key k, so there have been a lot more different attacks for the AES block cipher.

In fact the attacks have further reduced the number of false injected from 16 faults as we have seen over here to just a single fault. In fact the best known attack on AES just requires a single fault present in the 8^{th} round which can completely obtain the AES secret key, so will not go into details about these fault attacks but we will actually stop over here and leave it to the interested viewers to actually go through the relevant papers to look at the other more powerful fault attacks on AES. Thank you.