# Foundations of Cyber Physical Systems Prof. Soumyajit Dey Department of Computer Science and Engineering Indian Institute of Technology – Kharagpur

# Lecture – 52 Attack Detection and Mitigation in CPS

Hello and welcome back to this lecture series on Foundations of Cyber Physical Systems. So, today we will be starting on with the ah last week's topic ah which is ah well I mean based on the security part.

## (Refer Slide Time: 00:39)

oduction	A Case Study of CPS Attack Attack Detection Techniques addoocoocoocoocoocoocoocoocoocoocoocoocoo	Att Ó	Attack Mitigation Str. 0	
ourse Or	ganization			
Тор	ic	Week	Hours	
CPS	6 : Motivational examples and compute platforms	1	2.5	
Rea	I time sensing and communication for CPS	2	2.5	
Rea	I time task scheduling for CPS	3	2.5	
Dyn	amical system modeling, stability, controller design	4	2.5	
Dela	ay-aware Design; Platform effect on Stability/Performan	ce 5	2.5	
Hyb	rid Automata based modeling of CPS	6	2.5	
Rea	chability analysis	7	2.5	
Lya	punov Stability, Barrier Functions	8	2.5	
Qua	dratic Program based safe Controller Design	9	2.5	
Neu	ral Network (NN) Based controllers in CPS	10	2.5	
Stat	te Estimation using Kalman Filters (KF)	11	2.5	
Att	ack Detection and Mitigation in CPS	12	2.5	

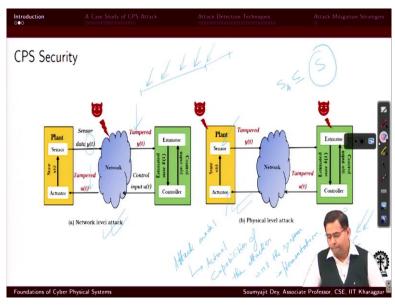
So, what we will be talking about is, how various kinds of CPS attacks happen. I mean, of course, there are a lot of in your classes of attacks. We will be, of course, touching up on some of them and we will also be discussing about several mitigation strategies that have come. I mean that have come out of research by people around the world. And so, we will be discussing how cyber physical systems their safety and performance, can be compromised using different kinds of attacks.

And also, what are the possible mitigation strategies that people are likely to adapt. ah in so that such attacks can be thwarted against. So, fine. Ah

(Refer Slide Time: 01:23)



(Refer Slide Time: 01:24)



So, just ah a basic introduction to CPU security here. So, we are showing you here, a very high level view of ah cyber physical implementation of a control system. So, let us say you have a plant, ah with some continuous dynamics and the plant has some attached sensors. OK And using these sensors, some data measurements are being obtained periodically. And so, at time t this data measurements will be sampled.

And let us say they are being traveling through a network and on the other side of the network. Ah So, it can be any network, it can be a wireless network, it can be a real time, vehicular network, intra-vehicular or inter-vehicular network, depending on what is the objective of the plant and how it is implemented whether it is a standalone or a distributed implementation, etcetera, etcetera. And then on the receiver side we have a controller which will be using this measurements y and it will be trying to estimate the state x hat. Ah So, we are now familiar with these symbols, based on our studies of Kalman filter and estimations. Right. So, let us say there is some estimator from which we are obtaining this estimate. And then there is this controller ah which is going to decide on a control signal  $u_t$ .

And that is again getting transmitted through the network to a plant side actuator ah which will be updating this control commands at the end by this period ah to the plant side. And then based on this updated control commands, the state will again evolve for the next ah next period. ah And then again, ah the updated state will again be sampled by the sensor here, ah in the at the start of the next period and that is how the cyclic behaviour will go on.

Now, what we are now considering here is. Let us say this network is under data. Now, well attacks can happen at various places. ah It can happen at the actuator or it can happen right at the sensor or it can happen on the network. ah So, the compromise on the signals can happen at various places, depending on the cyber physical system under question. Right Now, let us say, ah this sensor, data  $y_t$  is getting tampered due to this attack to some y prime t etcetera. Right

Now, this will definitely lead to a wrong estimation and wrong estimation will lead to a wrong control input. And not only that the attacker, if they have access to the control signals transmission, they can kind of tamper with that control signals also. right So, all these things very, I mean very much depend on the attack model. By attack model, we mean what are the real capabilities of the attacker to which the system is kind of exposed.

So, attacker may have access to certain network packets or, let us say, all network packets. So, if I just speak of an attack model here. So, this essentially depends on what are the actual capabilities of the attacker here. ah So, let us say like I was saying that the attacker can tamper any data here or maybe that attacker can tamper only some specific packets which are not encrypted or stuff like that.

Moreover, let us say the attacker can start attacking and there is a window. There is a window inside which those attack values will be ah tampering, the sensor signal. And maybe if the attacker is staying too long in the system, then of course the designers are not fools. Right So, they may have some detection systems. They may have some diagnostics running over the network. So, ah the attacker can also be detected and suitable actions may be taken against the attacker. Right

So, based on those actual system implementation details, the attacker will have some actual capabilities ah that for how much time they can sniff and for how much time they can inject data packets. I mean which are the variables on which the attacks can really happen. So, this is more like a vector. Right So, it may have the various components, so, the attack may be happening on some of the sensors data.

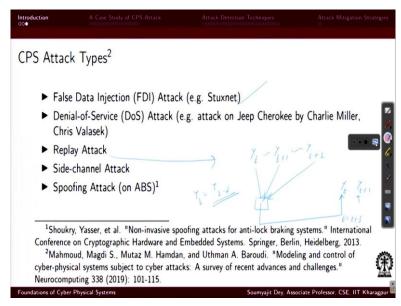
And some of the sensors may be may not be beyond the attackers capabilities of attack. right So, let us say you have a sensor suit S maybe the sensors which are under attack may be a subset of that because ah it may be invisible to for the attacker to attack this entire sensor. Similarly, such sub restrictions can also happen on the control side. ok ah So that is something about the attack's capabilities. Right

Now, also so, these are I mean if the attack is happening only the network side then it is a network level attack. If the attacker attack is happening ah directly on the hardware, let us say the attacker is able to modify the sensors readings by attacking the sensor only. he is He or she is perturbing the sensor readings right there OK or replacing some of the sensors with some faults or maybe actually attacking here or maybe while ah this.

So, finally, how is the controller implemented is a program. right It is a program p which is, let us say, running on some microcontroller. Right So, the attacker may be tampering with the memory of the microcontroller. Some of the bits on the microcontroller memory may be under attack the bits navigating flip or so. If you are reading specifically on hardware security, you will figure out that well there exists lot of practical ways in which of processor executing some code can be attacked. Right So, attacks can happen at the controller processor also. right But of course that also means that the attacker has access to this device which may or may not be true. This is kind of the more likely situation provided user distributed implementation where the controller and the plant are sitting for behind and the network is not protected well enough. But these are also use cases where things need attention.

So, we are not going into that detail like where and how the attack is happening. But we are like abstracting it out in terms of the mathematics here we are saying that if there is an attack, it will lead to a change in  $y_t$ . And if there is an attack on the controls it will lead to a change in  $u_t$  here.

## (Refer Slide Time: 08:34)



So, there has been several classes of attacks which have been reported in the cyber physical system security literature. For example, there was this widely known Stuxnet, you can read about it. ah So, ah that that is basically a similar kind of false data injection or FDI attack. So, basically they are kind of ah disturbing certain measurement variables. right So, there has been the Ukrainian power grid attack where some power system was attacked.

And the smart grid associated with this with the power system was compromised in a way that several brownouts and blackouts happen across the grid. So, this there has been attacks, ah due to which the nuclear reactors needed to be shutdown. right So, similar real so, this is what I am trying to point out is ah these are real attacks that have actually happened on critical infrastructures.

So, protecting such critical infrastructures is a very important issue and they can be protected using well. ah I mean cryptographic techniques other I mean well known existing network security techniques. But since this is the cyber physical system, if those techniques are also aware of the underlying dynamics, then maybe the protections would be more effective. So that is that that is something to think about.

For example, there has been these attacks which were reported by ah Charlie Miller and Chris Valasek is known as the attack on a Jeep Cherokee. So, the point is well they these experimenters took this specific vehicle for the attack. But at that level, at the time when this kind of attacks were reported, ah almost every automotive manufacturers, vehicles would have been vulnerable to the attacks that the denial of service types attacks that were implemented by this attacker.

So, it is not only about one vehicle, it was it was about almost every vehicle available. Its just that they chose one specific vehicle for the attack. Moreover, I mean there are several other attack classes for example, there are replay attacks. Now, this is an interesting class of attack, ah just to understand what this really is. um So, let us say, ah you are transmitting some measurement  $y_t$ . And then in the next cycle you will transmit  $y_{t+1}$  and  $y_{t+2}$ .

So that is the actual measurements. What the attacker is doing is, they are kind of observing ah and and storing this data in a buffer, so, the he is like a man in the middle. And then he is not really transmitting this data at the time points t, t plus 1, t plus 2 but rather than that he is transmitting this data. Let us say somewhere here. So, let us say at t equal to t plus 3 he is transmitting  $y_t$  then in the next cycle he is transmitting  $y_{t+1}$ .

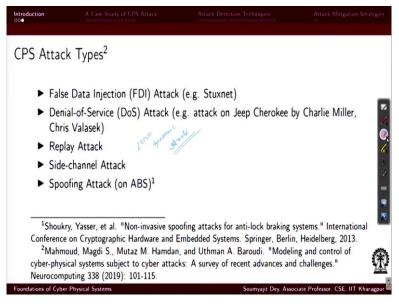
So, essentially I mean, if I think, of a mathematical function. What the attacker is doing is at  $y_t$  time? ah He is assigning the value  $y_{t-d}$  with the delay. right Now, the point is, this kind of attacks are very easy to implement, as you can see, is just a record and replay. But the thing is this may be good enough to destabilize the control system. Because we are actually using stale measurements to compute some update signals.

So, for a first ah first changing control system this may be dangerous. And the implementation is very easy, as you can see, is just recording and replaying mode of what this may be very difficult to catch. Why? Because typically in a control system with a high periodicity, ah these values may be almost nearby. right Because the dynamics is changing slowly over a very small amount of period. Right

So, even if I am putting in a delay ah it may not be observable to some monitoring software because the values are differing by a very small amount right in the in the in the in the running situation. However, with respect to the systems dynamics specifically when the system is operating around, I mean near to some boundary condition. These things may be very difficult for the system. I mean the system may not survive the safety criteria.

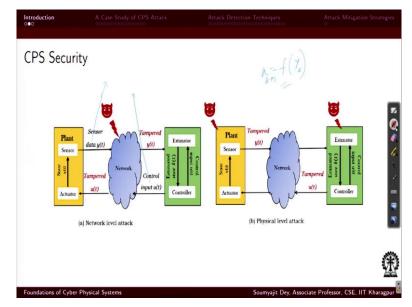
So that is an issue with replay attack and there are also several other interesting attack classes. So, if you are interested, you can study about something called zero dynamic attack.

### (Refer Slide Time: 13:07)



Now, of course, there are several other attacks that can happen on the controller side processor. For example, side-channel attacks ah or side channel attacks are typically used to reveal information about the system by observing the input, output channels OK and correlating variables. So, using such attacks people can actually figure out what is the underlying computation that is going on.

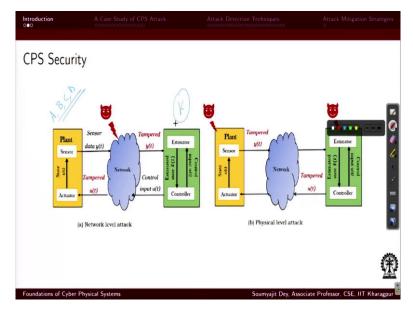
What are the parameters of the computation that is going on etcetera, etcetera. Moreover, there are also reports of well-known techniques like learning based attacks, so, just to give a context. (**Refer Slide Time: 13:45**)



So, suppose, ah you are you are unable to do anything right now but you can monitor the data. So that is another one attack model that is stage one at better. You just do not do anything but you just observe, ah as a as a as just as an observer that well this is the measurement. And then if I again update this is the control signal that is going and based on that this is the measurement. I am that is coming out.

So, essentially I can learn a function that will ah for this measurement. ah What is the output in the next step? Ok For this measurement y k what is the output in the control out, in control input or output of the function in the next step. So, if I learn a function, it is basically about learning the plant dynamics.

(Refer Slide Time: 14:36)



So, ah as you can see that initially, an attacker may not have much knowledge about what are the ah A,B,C,D matrices about the plant. right ah Because those are the matrices that that are kind of ah modelling, the plants dynamics, data may not have a clue about that. So that is also a part of the attacker model. Whether the attacker knows what are the plant matrices and whether the attacker know what is the control law. Ok

Now, the thing is, ah that those things may or may not be known to the attacker but if they are able to observe these measurements and control in mode inputs. They can use techniques like machine learning or other data driven techniques through which ah they can actually figure out well ah this is the plant matrix set and this is the controller. So, they can figure out those parameters.

#### (Refer Slide Time: 15:26)



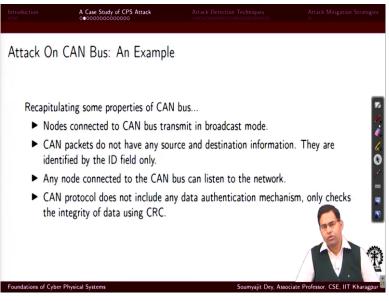
So that is exactly the point about attack model, what the attacker really knows and based on that knowledge what the attacker can really also learn. So that is why the important parameters of the system ah and whenever we are talking about an attack, we need to understand what is the attack model? What are the ah what are the things that the attacker can observe and for how much time.

And more importantly, using that knowledge how much more can the attacker derive about the system. And then using this original knowledge and the derived knowledge what kind of attacks can the attacker really do on the system. For example, there was this attack reported, ah which is a spoofing attack on ABS. So, ah what these people claim that well it may be possible to kind of inject disturbances into the sensor readings of ABS. OK

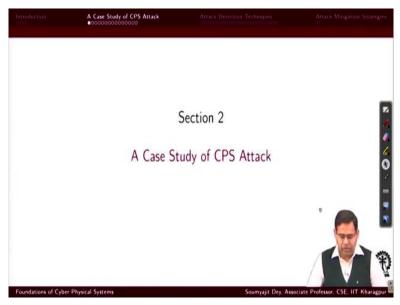
Using a suitable hardware that is kind of attached to the wheel well so, it is like they are creating an electromagnetic interference there using that they are kind of perturbing the Hall Effect sensor that is, measuring the wheel speed. ah But then the ABS system does not know about that. So that means whatever wheel speed measurement the ABS system is receiving is not the real wheel speed.

Now, you can understand how catastrophic that can be if the measurement I receive about the wheel speed is different from the actual wheel speed. The way the vehicular braking actions can be totally wrong. right Of course, there are questions from the automotive industry that whether such an attack is practically realizable. But what these people did is they kind of created such a toy example and showed that they can build a hardware through which it is possible to disturb the effects the readings of a Hall Effect sensor. Now, I think even that much is scary enough.

(Refer Slide Time: 17:13)



(Refer Slide Time: 17:16)



So, we will do a specific case study on cyber physical systems attacks. In this case, what we will be discussing here? Like I said that there are so many possibilities. It is a very important, ah discipline to to gain expertise on and do research by itself. And of course, all of that we cannot talk about inside a small lecture inside a course actually which is a general course on CPS. So, we will take some exciting examples here.

For example, we will talk about how ah signals on a CAN bus can be attacked. So, before going into that let us just recapitulate for some properties of CAN. So, as we know that CAN is a broadcast-based communication protocol. right So, nodes are connected to ah the CAN bus and they transmit in broadcast mode. That means every transmission by some control unit is practically readable by all other control units that are attached to the CAN bus.

And CAN packets do not have any source and destination information. right They have an ID which tells that what does this message really signify and any node if they require that ID, they can read it. Right So, ah such ah CAN packets ah I mean the CAN bus can listen to them any node that is addressed to the CAN bus condition to such a network and read all such ID's. Right

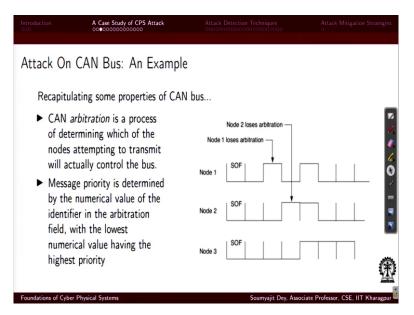
Now, in general, the CAN protocol will not have any data authentication mechanism. It is only checking the integrity of the data that whether the data has gone through any fault or some perturbation using an is into a check that is the CRC check. And it is known that CRC polynomials can be easily ah reversed and I mean and I mean that is reversible somebody can actually observe the packets and they can dig out the reverse kind of the CRC polynomial and rest on that. Ah

Even the node which does not have the polynomial, can actually figure out what it is and they can actually ah figure out well what are the actual message bits that were transmitted and whether they whether there is many perturbation or not. right So, overall anyway ah whatever you are transmitting ah on CAN is the pure data. The CRC computation is for the integrity check, ah and and that I mean that does not suffice.

Because anyway, you are not ah kind of. Ah I mean like I have been saying that well the CRC polynomial can be reverse engineer. So, when you do not know some some malicious node, if they get attached to the CAN bus, they can easily find out what is the CRC polynomial. But in spite of all that it does not really matter because whatever data you are sending there, there is a raw data. right

I mean there is no kind of transformation in the data Ah to actually hide, it is significance. So, all you have is the CRC integrated check which is useful only to figure out whether there has been a perturbation on the original data. right So, well if that can be broken then people can go people can get away with that that ECU also, I mean that this is not a very secure mode of communication anyway.

#### (Refer Slide Time: 20:17)

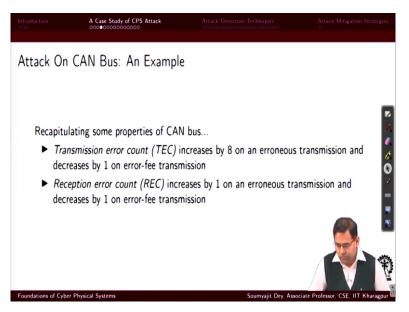


Let us have a relook on the arbitration process that is followed in CAN. So, ah if you remember that lets let us have a relook to this diagram that multiple nodes are trying to ah communicate and they are competing for access to the bus. So, here everybody is synchronized with a start of frame bit. And then let us say all of them initially, transfers are 0, so, everybody also sees the 0 on the bus.

Because the bus behaves like a wired and and then let us say, node 1 tries to transmit a one and then it will lose the arbitration because ah the bus is an AND. So, it will transmit node 1 but it will see a 0. So, it knows that somebody of higher priority is there. Then let us say here node 2 he is transmitting a 1 but the others are transmitting a 0. So, node 2 will lose the arbitration. So, node 3 will win and then whatever message it wants to transmit.

Let us say here is just an example it is transmitting all 1 the point is after this point, whatever node 3 will transmit will be actually being seen by everybody. And the other nodes by the by the CAN protocol they will just back off. right So that is the kind of arbitration process that can follows and if you remember in this process works because ah the higher priority is assigned to the value with the lower ID.

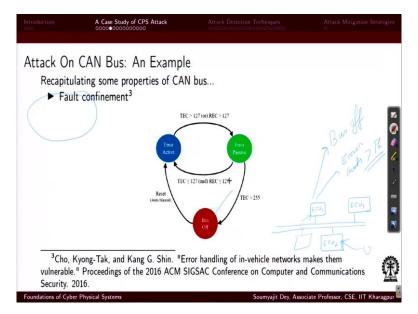
That means the lower ID will will be having smaller more number of zeros and it will be winning the arbitration more often. ok ah (**Refer Slide Time: 21:37**)



So now, in CAN bus if a CAN transmission fails, there is an error count that is maintained by the CAN controller for each of the nodes. ok So, ah actually there are two counts that they are being that are being kept. So, this is known as the transmission error count and the reception error count. So, in case of a transmission error, this transmission error count value will increase by 8. And in case there is an error free transmission, the transmission error count, the TEC value will decrease by 1.

Similarly, there is this reception error count that in case of an erroneous reception I mean which of course has happened due to an erroneous transmission this REC value will increase by 1. And again, in whenever there is an error fee transmission and reception this REC value the reception error count will again decrease by 1. So, a plus 8 minus 1 and REC is plus 1 minus 1.

(Refer Slide Time: 22:38)



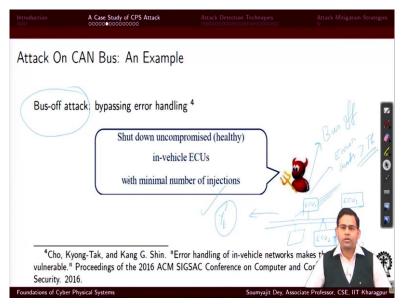
So, this is a fault confinement property of CAN bus. So, what happens is the CAN, the The all the nodes will be in each of each any possible of these states. So, it can be an active state known as the error active state. right So, there are some t some transmission errors but those values are kind of less than 127 and the reception error count is also less than 127. If any of them goes beyond 127 then the cost it still attached to the bus ah but ah it is in an error passive state, where it should not be transmitting again. Right

And then in case this transmission error count will go beyond 255. Then ah this ECU will go to a bus off state that means the current controller will take it to a bus off state. That means for a period of time it will not be transmitting until it is again online after that period. So, after bus off all these counts will be reset and after that specified bus of period of the protocol, it will again be back to the interactive error active state.

So, this is how the CAN system is built and the idea here has been that well there may be many faults happening and in case of faults, if there are too many faults that are happening then maybe right now it is not good idea for this ECU to continue. So, let us give it some time and for that small interval of time we will take this thing offline from the bus. It will not allow, we will not allow it to transmit anything or receive anything.

And we will expect that after this small interval it will, if I reset and again start it which is a part of the protocol itself. ah Then ah that transient period, where many faults were occurring on the system will not be happening and we expect that the system will be working fine and it will be back in the reductive state. So, so that was the idea behind this bus of attack.

### (Refer Slide Time: 24:29)



Now, ah this protocol so that was the idea behind this fault confinement protocol that is part of what of CAN. But this is also now a vulnerability based on this work ah which reports that well how this protocol can be actually exploited to kind of attack ECU and force them to a bypass to a to a bus of state. So, what it means that well ah there is a healthy ECU and the and what an attacker does is, ah the attacker is sitting in another ECU.

So, let us just try to understand why that can happen. So, let us say this is your vehicular architecture this is your CAN bus. And let us say I am targeting this ECU, so, what can happen is, ah the attacker is a piece of malicious code which can be injected into the system in various possible ways. Let us say you took the vehicle to a ah kind of malicious service center who inserted a code in this ECU.

Or maybe they attached a small processor here to the system without your knowledge, or maybe they because today is the age of (25:59) connected vehicles, somehow they bypass the vehicles authentication systems. And they downloaded this malicious software and many of these ECUs are actually multi code. Right So, they have their existing software which may be running in some of the course.

And let us say one of the course, one of the other codes have been um I mean made to run this malicious software or maybe that malicious software has downloaded and it is not active right now. So, you know your diagnostics in vehicle diagnostics does not have a clue. Let us say,

after some specific interval of time may be a year or something the this is programmed to suddenly go online and then do some malicious work.

And what it what it does is as follows. That it knows that what is it is target ECU. And they will implement this bus-off attack which will do something about this specific protocol. And what it really does is that it it actually exploits this protocols vulnerability and it takes suitable actions using which you take suitable actions using which for this ECU at this error counts. This will be greater than the thresholds and the ECU will be taken to a bus-off mode. This mode.

So that is not really happening the reason the can the CAN protocol was built like this is because possibly at that time, nobody thought that there can be an attacker who will inject wrong things. All these things were built with the idea of giving redundancy and fault confinement. The idea was that well whatever imperfections can happen can happen due to ah hardware environment. So, I mean imperfections cannot be are not going to be improv, I mean intentionally added to the system. Ok

So, ah but now, what can happen is with so much of software and so much of reward to be gained people can use this redundancy mechanism that is there for fault confinement to actually create a security breach. And how that is done we will talk about. But just to give the idea what it does is this thing this malicious software will keep on injecting packets here in such a specific way.

That these ECUs, like I have been saying it is error count goes below beyond the threshold and once that happens ECUs is no more there. right Now, when this ECU is no more there when it was there, it was let us say, transmitting Ah some sensor readings y t of some specific variable, of some specific specific variable. Now, this period, when this ECU one is being made offline, the malicious software will actually take it is role.

And now, it will start transmitting  $y_t$  primes which are I mean values that it is sending based on some sensor measurement that it is observing and it is not really transmitting those actual measurements but it has part of those measurements which which some some wrong values and it is sending those wrong values. And this is becoming possible ah because the actual values which were supposed to come with the same ID's are not even coming.

Because the sender of those actual values is taken offline. Now, of course, this will not continue for enough time because if we see there was a bus-off period after which this is ECU 1 will again come back. But this malicious software can keep on repeating. It is a software unlike a human in the loop. So, it can keep on hitting it will keep on hitting ECU 1 in a suitable vulnerability and periodically forcing it to the bus off state.

And for the time when you see one is in the bus of state, this malicious software will keep on injecting this part out messages. right So, with this we will end this lecture and in the next lecture we will see that how this process of forcing ECU 1 to a bus of state can really be done. And by the way what are we really trying to show here is something like this that we are trying to talk about.

We said that well there can be attacks there can be things happening based on this model. Now, we are really trying to show that well ah it is not kind of only confined to a mathematical level. ah But it can really happen that mathematical ah situation where the  $y_t$  is getting replaced by a  $y_t$  prime or a  $u_t$  getting replaced by  $u_t$  prime, is definitely possible through suitable platform level attacks. So that is one example we will see. So, with this we will end this lecture. Thank you for your attention.