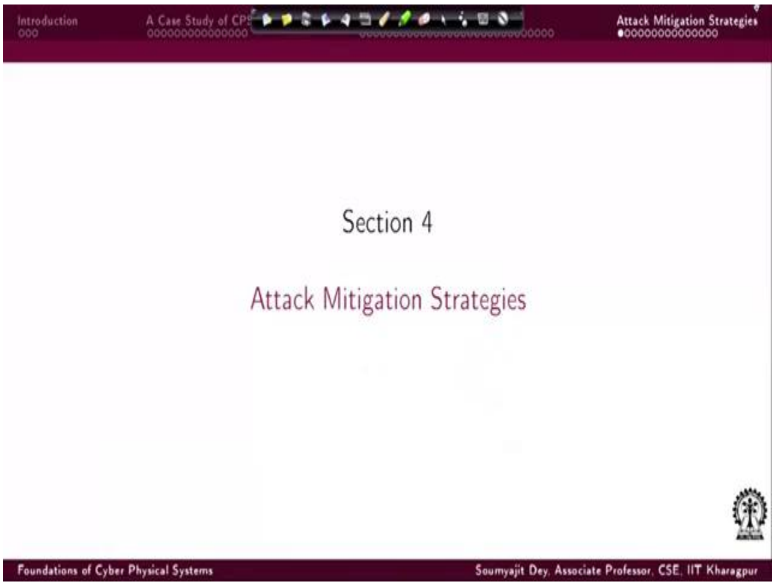
Foundations of Cyber Physical Systems

Prof. Soumyajit Dey
Department of Computer Science and Engineering
Indian Institute of Technology – Kharagpur

Lecture – 58
Attack Detection and Mitigation in CPS (Continued)

Welcome back to this lecture series on Foundations of Cyber Physical Systems. So, in the previous lecture, we have talked about different kinds of attack detection strategies, ah threshold based, variable threshold, RL based variable threshold, etcetera, etcetera. ah And also I mean detection based on CAN packets and they are ah aperiodic execution of the controller.

(Refer Slide Time: 00:53)

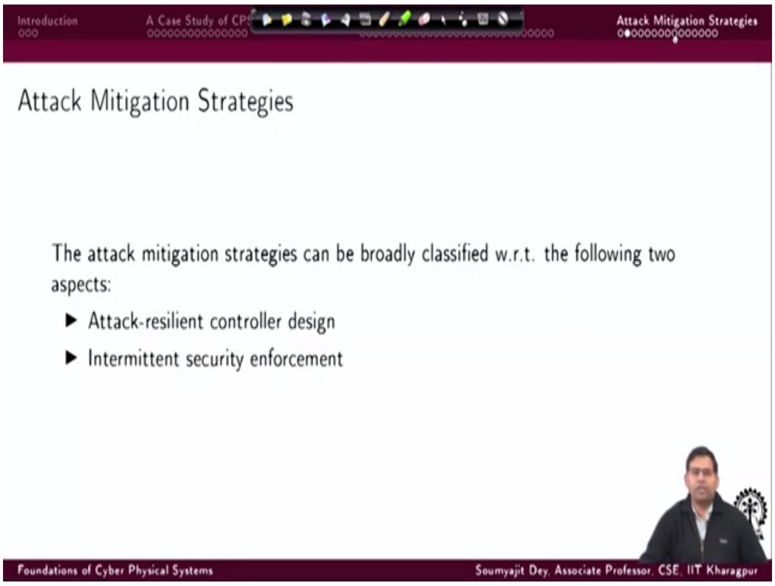

Introduction
A Case Study of CPS
Attack Mitigation Strategies
Foundations of Cyber Physical Systems
Soumyajit Dey, Associate Professor, CSE, IIT Kharagpur

Section 4

Attack Mitigation Strategies

So, in this lecture we will talk about some attack mitigation strategies ah which people have evolved and they are already in place.

(Refer Slide Time: 00:55)


Introduction
000
A Case Study of CPS
Attack Mitigation Strategies
Foundations of Cyber Physical Systems
Soumyajit Dey, Associate Professor, CSE, IIT Kharagpur

Attack Mitigation Strategies

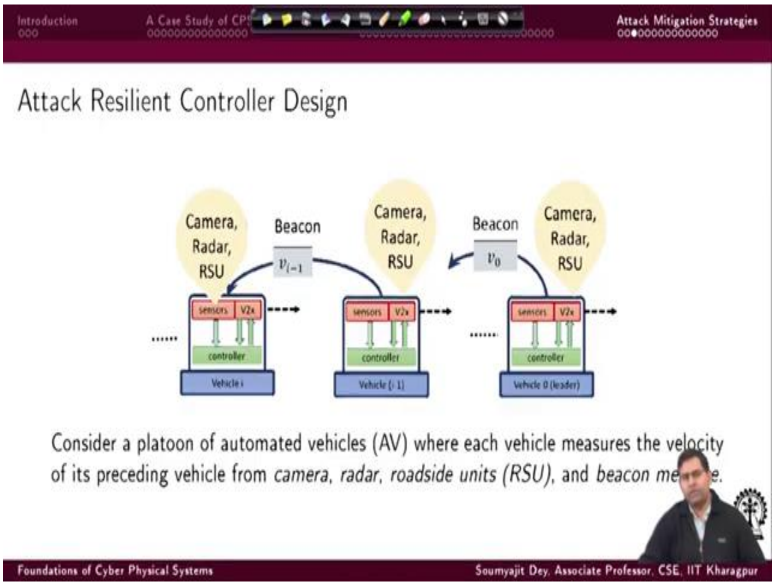
The attack mitigation strategies can be broadly classified w.r.t. the following two aspects:

- Attack-resilient controller design
- Intermittent security enforcement

So, ah when attack mitigation has to be done, there are several methods ah which can be broadly classified into two primary classes. ah One is attack resilient controller design. So, ah suppose you are able to design a controller for a plant in such a way that let us say at least for those attacks which are stealthy we know that the perturbations of the attacks that are injected would not be very big because eventually then they will be detected.

But for such small perturbations ah the controller is robust enough to guarantee that a system remains safe. So that can be one method. And the other method is what we call as intermittent security reinforcement.

(Refer Slide Time: 01:36)



Attack Resilient Controller Design

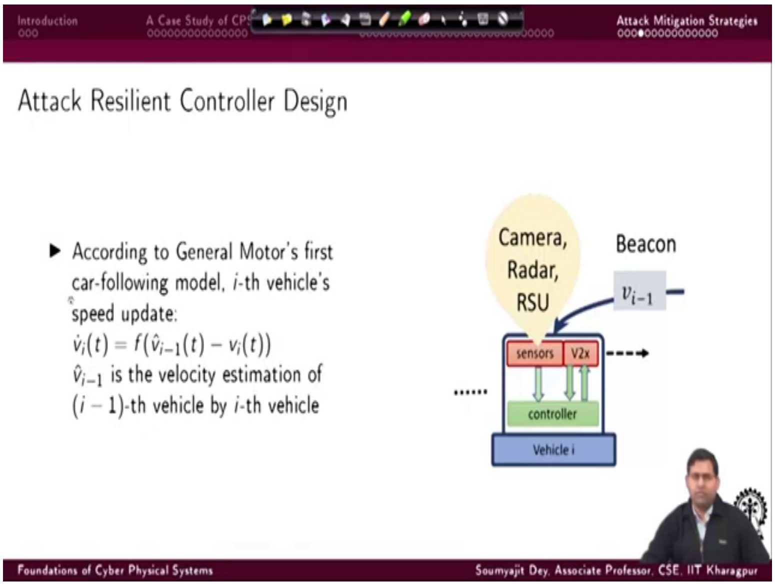
Consider a platoon of automated vehicles (AV) where each vehicle measures the velocity of its preceding vehicle from *camera*, *radar*, *roadside units (RSU)*, and *beacon message*.

So, ah let us take a small example of attack resilient controller design. So, ah we we consider a platoon of autonomous vehicles here, let us say. ah So that means you have multiple vehicles and every vehicle have got a significant number of sensors. So, it can have a camera sensor. It can have a radar sensor. It can have a roadside unit. OK And based on these sensors Ah so, let us say this is your vehicle v0, this vehicle 1, ah this is vehicle 2 like that.

In general, you have I minus 1 as vehicle and ith vehicle like that. ok So, ah every vehicle is sending a beacon message OK ah which is basically this sensor measurements. ok so So, each vehicle measures, the velocity of it is preceding vehicle Ah from this different sensor, measurements which are coming to it as part of this beacon message. So, ah let us say this vehicle ah lets a vehicle one here is going to receive v_0 beacon message from vehicle 0.

And v_0 is going to contain the camera measurement, radar measurement and the RSU, information. And is going to estimate the vehicles state in this case the velocity and maybe this may be the separation etcetera ah based on this beacon messages.

(Refer Slide Time: 02:57)



Introduction

A Case Study of CPS

Attack Mitigation Strategies

Attack Resilient Controller Design

- According to General Motor's first car-following model, i-th vehicle's speed update: $\dot{v}_i(t) = f(\hat{v}_{i-1}(t) - v_i(t))$ $\hat{v}_{i-1}$ is the velocity estimation of $(i-1)$-th vehicle by i-th vehicle

Foundations of Cyber Physical Systems

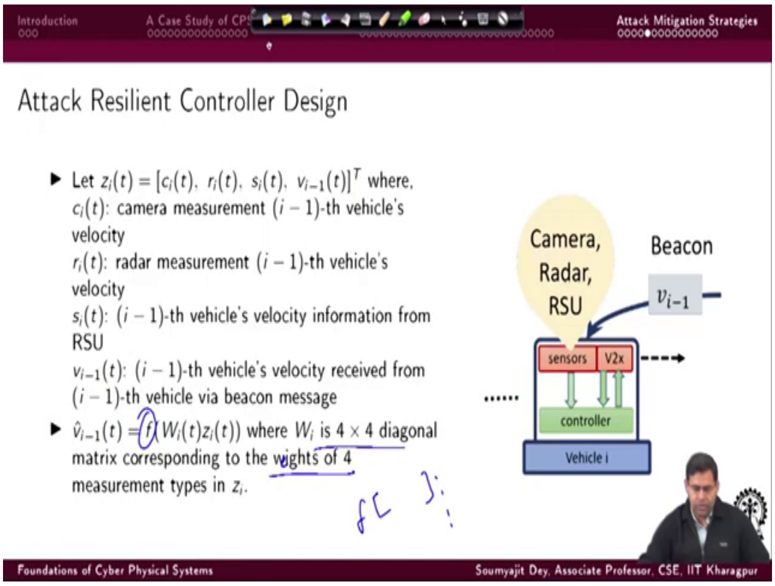
Soumyajit Dey, Associate Professor, CSE, IIT Kharagpur

So, ah yes using a standard car-following model ah model we have this ith vehicle speed update Ah given something like this that. ah So, what is essentially the platoon going to do? The platoon is basically going to follow. I mean every ith vehicle we will try to match it is speed with respect to the previous i minus 1th and vehicle. That is the platoon's objective. right So, ah following that idea ah what the the control law inside every vehicle must be something like that.

That will I have got the measurements from the previous vehicle's beacon based on that I will estimate the velocity of the previous vehicle. ok ah So, this v_{i-1} hat is the estimate of i minus 1 equals velocity, as is done by the ith vehicle based on this beacon message. OK And then from this it will calculate the difference. That means whether what is the error? And it will apply some control law which is a function of the error. We are not going to that law.

It can be any control law which is which can attain this target objective of minimizing the error. And based on that control law it will update these vehicles acceleration v_i dot right as simple as that. This is a simple car-following model.

(Refer Slide Time: 04:14)



Attack Resilient Controller Design

- Let $z_i(t) = [c_i(t),\ r_i(t),\ s_i(t),\ v_{i-1}(t)]^T$ where,
  $c_i(t)$: camera measurement $(i-1)$-th vehicle's velocity
  $r_i(t)$: radar measurement $(i-1)$-th vehicle's velocity
  $s_i(t)$: $(i-1)$-th vehicle's velocity information from RSU
  $v_{i-1}(t)$: $(i-1)$-th vehicle's velocity received from $(i-1)$-th vehicle via beacon message
- $\hat{v}_{i-1}(t) = f(W_i(t)z_i(t))$ where W_i is 4×4 diagonal matrix corresponding to the wights of 4 measurement types in z_i.

Foundations of Cyber Physical Systems
Soumyajit Dey, Associate Professor, CSE, IIT Kharagpur

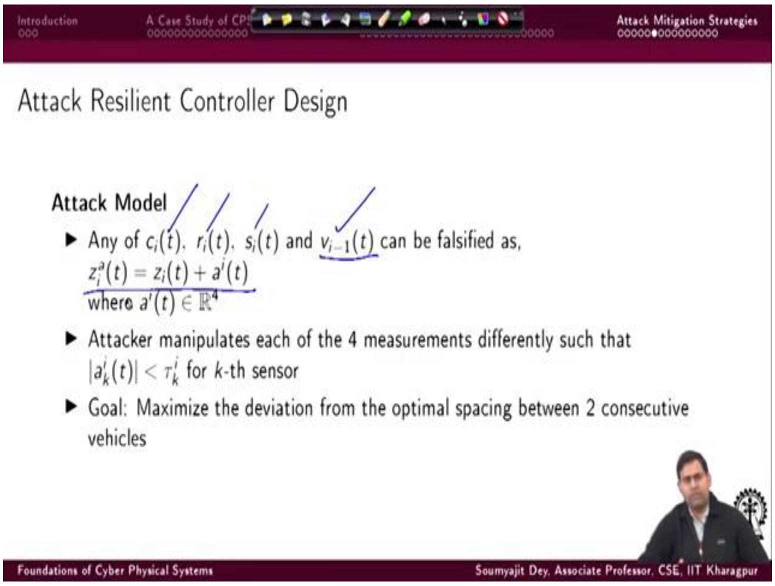
So, ah let us say, z_i denote Ah this sensor message vector where c_i is the camera measurement. ok ah So, basically, this ith vehicle is getting the i minus 1th vehicle’s beacon message which is coming like this. That is i minus 1th vehicle’s camera measurement at time t similarly, radar measurement at time t and similarly, velocity information as is received from the RSU at time t.

And based on this ah it will compute Ah this i minus 1th vehicle’s actual estimate ah of velocity. Ah And all this behind you is happening in the ith vehicle OK and let us say, ah there is some sensor fusion technique which is in use, ah which is like which is we are not discussing any specific technique. There are very well known available techniques. We are just abstracting it out using this function f.

And we are saying that well before applying this function, what is done is? Each of these measurement values would be weighted suitably ah by ah 4 cross 4 diagonal matrix which is corresponding to the ah weights of this force sensor measurement types. That means well I have got various sensor measurements. I have ah different levels of confidence on these sensor measurements. So, I will wait them out using this weight matrix.

And then I will use some functional form to estimate this actual value. ok It can be some estimator ah some algorithmic technique that is well known. So, there are several methods any of them can be used here.

(Refer Slide Time: 05:54)



Introduction
A Case Study of CPS
Attack Mitigation Strategies

Attack Resilient Controller Design

Attack Model

- Any of $c_i(t)$, $r_i(t)$, $s_i(t)$ and $v_{i-1}(t)$ can be falsified as, $z_i^a(t) = z_i(t) + a^i(t)$ where $a^i(t) \in \mathbb{R}^4$
- Attacker manipulates each of the 4 measurements differently such that $|a_k^i(t)| < \tau_k^i$ for k-th sensor
- Goal: Maximize the deviation from the optimal spacing between 2 consecutive vehicles

Foundations of Cyber Physical Systems
Soumyajit Dey, Associate Professor, CSE, IIT Kharagpur

So, in this case, ah let us understand what an attacker can do? So, we have understood what are the what is the system? We have a platoon of vehicles and the each vehicle has a controller. So, overall you have a distributed control implementation. And what the controllers together want to do is? They want to ensure that well all the vehicles ah kind of follow the platoon's velocity. I mean all the vehicles move with the velocity that is basically the platoon's velocity. Right

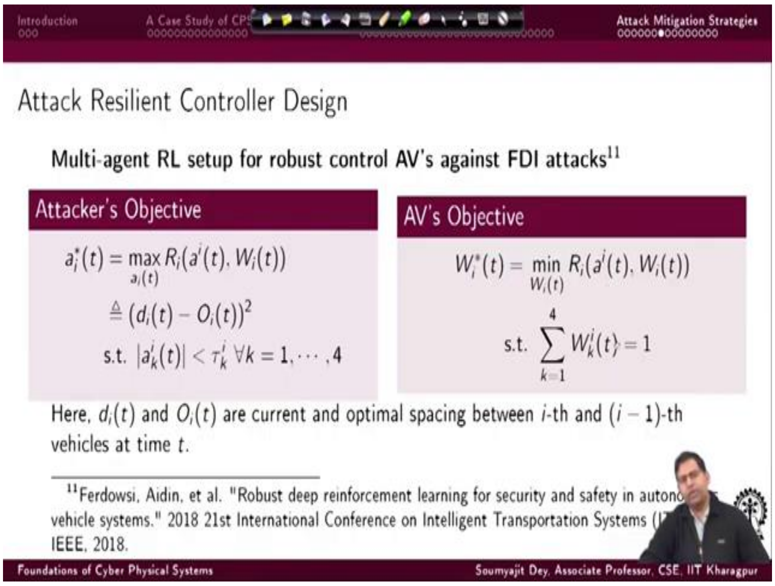
Now, the attack motor in this case is something like this Ah that you I mean the an attacker can target an ith vehicle. And it can It can actually falsify this camera data or the radar data or the RSU data and using that it can actually disrupt this estimate value. right So, overall ah this attack can be written like this in the vector form. So, this is the z which is the vector of all this information.

And it can and an attacker can actually manipulate all these measurements or maybe a subset of them. ok Now, the goal for the attacker would be to maximize the deviation from the optimal spacing. right ah Because ah there is an optimal spacing ah which is there between each vehicle.

So, what we want is that well? ah The every vehicle must move at a space at a speed which is almost equal to the platoon radar speed, while maintaining a minimum safe distance. Right

And in that case the attacker's goal must be that well this safe distance specification between two vehicles must be disturbed. right ah That is what an attacker will try to do.

(Refer Slide Time: 07:38)



Introduction A Case Study of CPS Attack Mitigation Strategies

Attack Resilient Controller Design

Multi-agent RL setup for robust control AV's against FDI attacks[11]

Attacker's Objective

$$a_i^*(t) = \max_{a_i(t)} R_i(a^i(t), W_i(t))$$
$$\triangleq (d_i(t) - O_i(t))^2$$
$$\text{s.t. } |a_k^i(t)| < \tau_k^i \; \forall k = 1, \cdots, 4$$

AV's Objective

$$W_i^*(t) = \min_{W_i(t)} R_i(a^i(t), W_i(t))$$
$$\text{s.t. } \sum_{k=1}^{4} W_k^i(t) = 1$$

Here, $d_i(t)$ and $O_i(t)$ are current and optimal spacing between i-th and $(i-1)$-th vehicles at time t.

[11] Ferdowsi, Aidin, et al. "Robust deep reinforcement learning for security and safety in autono vehicle systems." 2018 21st International Conference on Intelligent Transportation Systems (IEEE, 2018.

Foundations of Cyber Physical Systems Soumyajit Dey, Associate Professor, CSE, IIT Kharagpur

So, if you try to create a multi-agent, RL similarly, here or maybe any other optimization technique. Ah So, the attackers objective can be written like this. That well you maximize this function, ah where, ah basically it is that well you maximize this difference ah between ah the current and the optimal spacing. So, let us say the target spacing is y and the current spacing is d_i. So that is this difference here. It can be sometimes positive or sometimes negative.

So, let us take the square and the attackers objective at every time t is to maximize this by injecting a suitable attack a_k^i in the kth instant. In a way that this value does not manipulate z_i too much so that the estimate will kind of differ from the actual measurement ah by ah by beyond the threshold. ok So that is what we are trying to say that the attacker will try to be stealthy while it will try to ah maximize this difference.

Similarly, the autonomous vehicles objective is that well ah it will try to ah kind of the opposite thing. It will try to minimize this difference that means it should try to track the previous vehicle as nicely as possible in a matching velocity and also matching this displacement error. There is a safe separation that is targeted and it will try to be as close as that to that as possible.

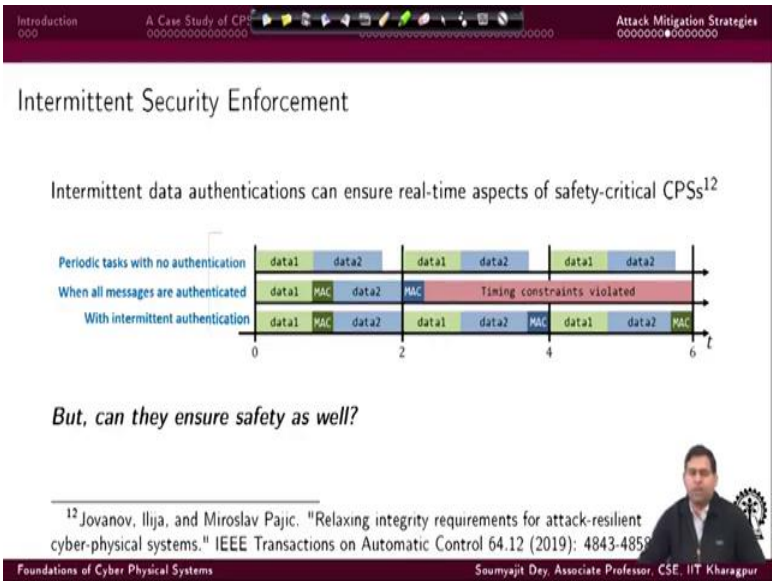
And this is the weight matrix different weights. So, the I mean those proportions have been summed up to one because that is how the weights have been distributed across the weight matrix. ah And so that acts as a constant over this optimization problem here. So, this is actually taken from this paper which was published in con this in the ITS conference on international I mean Intelligent Transportation Systems.

And so and they figured out well this can lead to interesting learnings that well it is really possible to attack vehicle platoons and by spoiling some of the measurements data. It can be possible to create vehicle collisions and stuff like that. But if you design this kind of a setup then well you can have a resilient system that has been designed where ah because you have two agents. Ah

You create an attacker agent and then you are trying to train an autonomous vehicle’s controller agent which is trying to thwart the attacks. So, the attacker agent will generate optimal attacks because it will try to generate attacks here. I would say optimal because it is an AI engineering. So, ah the attacker agents target is to maximize this separation, while satisfying this stealthy constraints, stealthness constraint.

And the vehicle in the process the vehicles agent, we learned that well how to thwart the attacker by nullifying the attacks and still maintaining the separation nicely. Ok So that is a way ah to create attack mitigating controllers, by a learning based technique. There are many other well known examples also. We thought this is a simple example and that should be nice for this introduction to this topic.

(Refer Slide Time: 10:50)



Introduction A Case Study of CPS Attack Mitigation Strategies

Intermittent Security Enforcement

Intermittent data authentications can ensure real-time aspects of safety-critical CPSs[12]

Periodic tasks with no authentication
When all messages are authenticated
With intermittent authentication
data1
data2
MAC
Timing constraints violated
0
2
4
6
t

But, can they ensure safety as well?

[12] Jovanov, Ilija, and Miroslav Pajic. "Relaxing integrity requirements for attack-resilient cyber-physical systems." IEEE Transactions on Automatic Control 64.12 (2019): 4843-4858

Foundations of Cyber Physical Systems Soumyajit Dey, Associate Professor, CSE, IIT Kharagpur

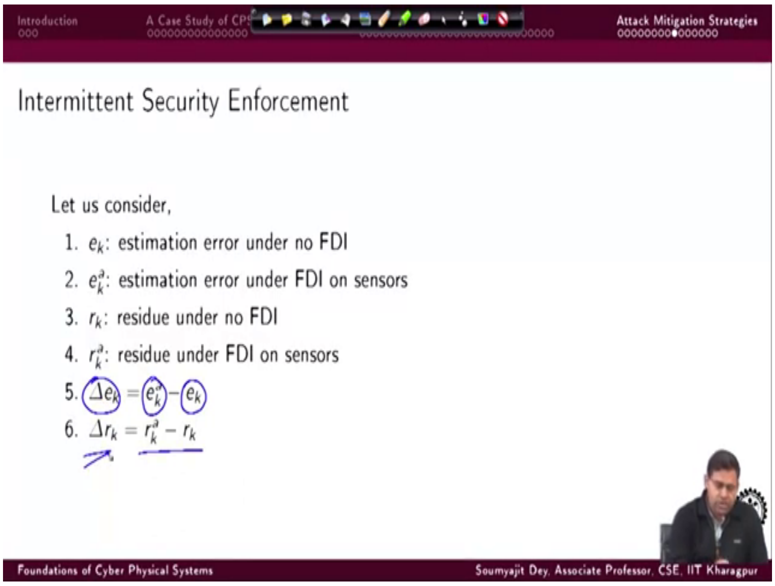
Now, let us come to the other topic which is what we call as intermittent security enforcement. So, let us understand what we mean. So, if you recall our previous computation of HMAC and AES, we said that well if we take a CAN packet and we use HMAC and AES. It will result in creating, I mean ah 6 times bus load for each for each kind of communication. The point is well then we cannot encrypt all the communications in a control loop because of bandwidth constraints.

But well what about having some intermediate situations? So, the thing is, if you can have a periodic task with no authentication a data 1 data 2 again in the next period, data 1 data 2 like that or you can have all messages being authenticated. So, data 1 followed by MAC then data 2 again the MAC. Right So, what is happening is data 1's timing constant will now get violated like we discussed.

So, what we do is? Now, will let us not ah encrypt all packets but let us encrypt some of them. So, data 1 we encrypt here then we send data 2. But in the next frame we do not send data 1 with encryption but we send data 2 with encryption. And we see that way in this way if we just reduce the encryption overhead by half ah things are kind of ah schedulable. I mean in this pictorial example but the trade-off is well I am not securing all packets.

But then the question comes that well if I do not see your all packets, how do I choose what should be the packets to secure so that still ah with this much small amount of security, even if there is an attack ah it will either be detected or if it is not detected then I have an ensure I have a guarantee that such smalls till the attacks would not make the system unsafe. How can I actually reduce that?

(Refer Slide Time: 12:43)


Introduction
000
A Case Study of CPS
Attack Mitigation Strategies
Foundations of Cyber Physical Systems
Soumyajit Dey, Associate Professor, CSE, IIT Kharagpur

Intermittent Security Enforcement

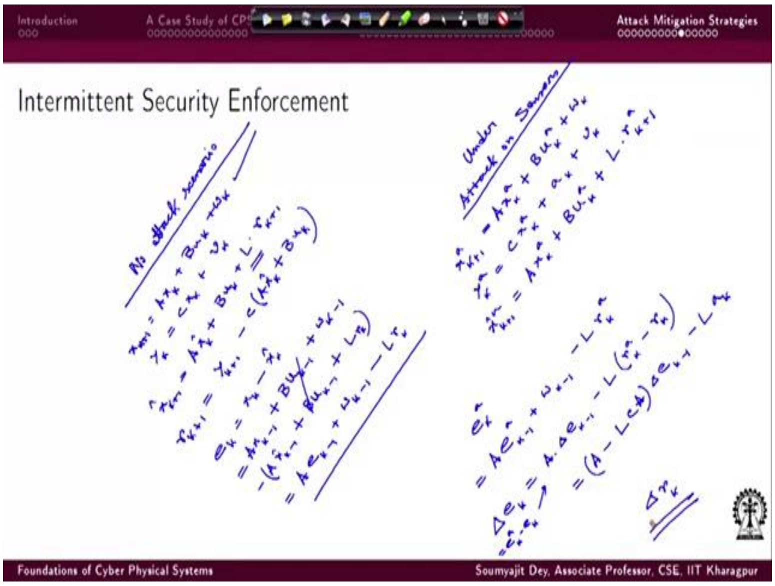
Let us consider,

1. e_k: estimation error under no FDI
2. e_k^a: estimation error under FDI on sensors
3. r_k: residue under no FDI
4. r_k^a: residue under FDI on sensors
5. $\Delta e_k = e_k^a - e_k$
6. $\Delta r_k = \underline{r_k^a - r_k}$

So, let us try to understand this problem. So, let us let us consider this ah simple for formulation here. So, let e_k^b ah estimation error under no false data injection e_k^a b and estimation error under false data injection on the sensors. Ok And you have a residue value r_k when there is no false data injection. And similarly, r_k^a when there is positive injection on the sensors.

So, based on this, ah we can say that well we have a differential error that is in case there is an error you have an estimation error, in case there is, in case there is an attack there is an estimation error. In case there is no attack you have another value of estimation error. So, this difference tells you that due to the attack, what is the increment in the estimation error? And similarly, due to the attack what is the change in the residue value? We will see that why we are actually interested in these quantities.

(Refer Slide Time: 13:44)



Introduction

A Case Study of CPS

Attack Mitigation Strategies

Intermittent Security Enforcement

No attack scenario
$x_{k+1} = Ax_k + Bu_k + w_k$
$y_k = Cx_k + v_k$
$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + L \cdot r_{k+1}$
$r_{k+1} = y_{k+1} - C(A\hat{x}_k + Bu_k)$
Under Attack on Sensors
$x^a_{k+1} = Ax^a_k + Bu^a_k + w_k$
$y^a_k = Cx^a_k + a_k + v_k$
$\hat{x}^a_{k+1} = A\hat{x}^a_k + Bu^a_k + L \cdot r^a_{k+1}$
$\Delta e_k = e^a_k - e_k$
$= (A - LCA)\Delta e_{k-1} - La_k$
Δr_k

Foundations of Cyber Physical Systems

Soumyajit Dey, Associate Professor, CSE, IIT Kharagpur

So, first let us consider the no attack scenario. So, this is a standard equation, like we discussed many times earlier. So, standard control update with a process noise. This standard measurement with measurement noise. The estimate, L is the Kalman gain. And then you can get the residue which is the actual measurement minus C times your estimate. So that is how you model residue right because well. Yeah

And from this what we can say is the error in the state estimation. And we can now, create an error dynamics here. ah If you remember, I mean our previous style of derivation because this will get crossed out. So, this is what you have as the estimation error and this is the no attack scenario. Now, what about the attack scenario? So, these were my standard equations in the attack scenario.

And from this again we can create the error equation in the attack scenario which will again be something like this only. So, I am just writing this one directly. So, from this, if you see what is the change in the estimation error? So, is basically this quantity minus this quantity it will eventually come as something like. So, this is like your e_k^a minus e_k. And next we will go about ah creating the ah this just like we created this value of delta e_k.

Which is telling me that well what is the difference in the estimation error between the attack and no attack scenario. Similarly, we also need to compute what is the difference in residue between the attack and no attack scenario. So, maybe we will continue from that in the next lecture. And we will end today's lecture with this derivation. Thank you.