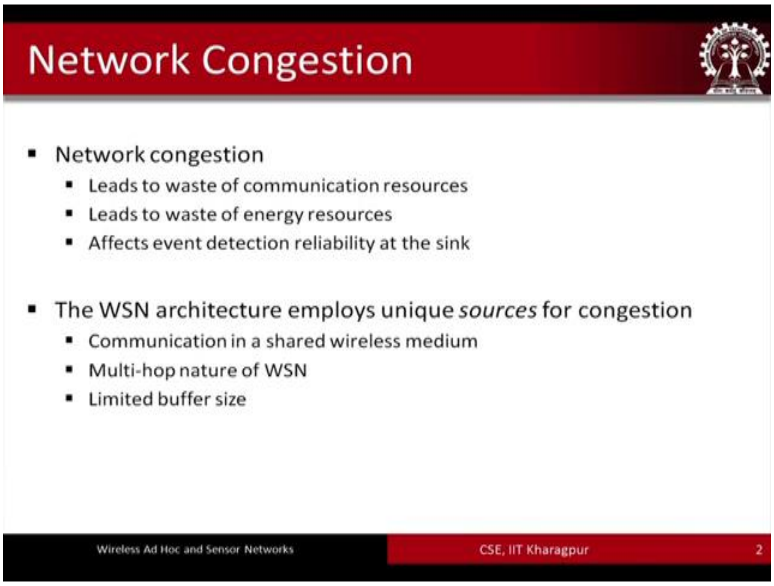
Wireless Ad Hoc and Sensor Networks

Prof. Sudip Misra
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 31
Congestion and Flow Control in Wireless Sensor Networks-Part-I

Congestion and flow control in sensor networks. So, this topic as well we have divided into two parts. So, overall what we are going to cover in both of these parts is a discussion about how congestion is a serious problem in sensor networks and what are the sensor network's specific issues that have to be taken into account while dealing with congestion in these networks, and then we will look into some of the solutions that have been proposed so far by different researchers to handle congestion and flow control in sensor networks.

(Refer Slide Time: 00:58)

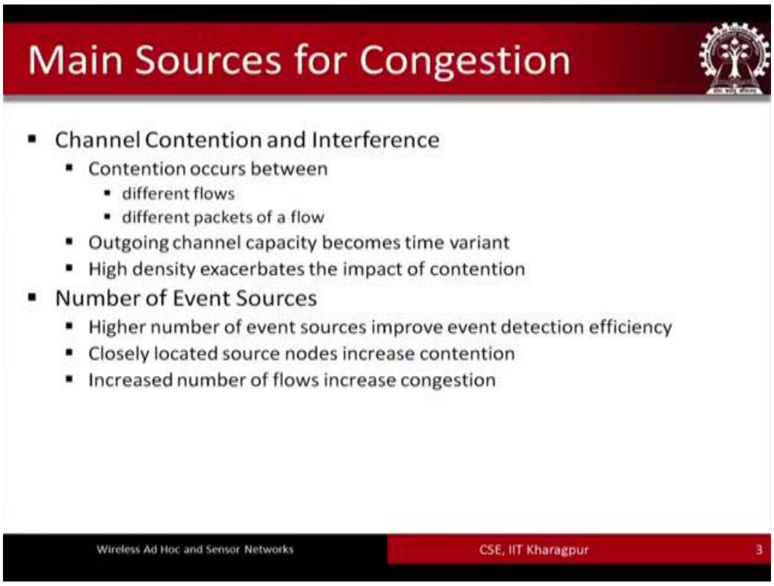


So, looking into congestion, network congestion basically is very undesirable, so whether it is sensor network or any kind of network. Congestion basically is undesirable because you know when we were talking about congestion that leads to wastage of communication resources, energy resources and so on. And overall, the reliability of data delivery due to congestion is affected seriously. This is due to the fact that congestion basically leads to typically packet drops. And, packet drop basically means that is something that is undesirable and is going to, you know is going to decrease the overall reliability of packet delivery to the sink or the intended destination node.

Now, taking WSN specifically in sensor network there are some unique sources of congestion. Congestion due to the fact that it is a wireless sensor network; it is a wireless network, multiple nodes, sharing a common medium. And, at the same time we have all the issues that we have discussed about multi hop communication and the problems leading to congestion due to that fact. And, it is something that we have discussed at a time of Ad Hoc networks congestion control. And, the third is that in these networks we have very limited buffer size. The nodes have very limited buffer size; only a few kilobytes of buffer in each of these different nodes.

So, if there is limited buffer, then only very small amount of buffer space can be exploited. That means that if you have too much of information that has to be buffered that is not possible, and that will lead to packet drops and so on. So, you know, typically you know what happens is when you have a congestion kind of scenario, you try to buffer it; buffer, you know, whatever is available. So, if buffer space itself is limited, then basically you know there is no other way, but to drop the packets. And that is undesirable for not only sensor networks, but for any network as we know.

(Refer Slide Time: 03:02)

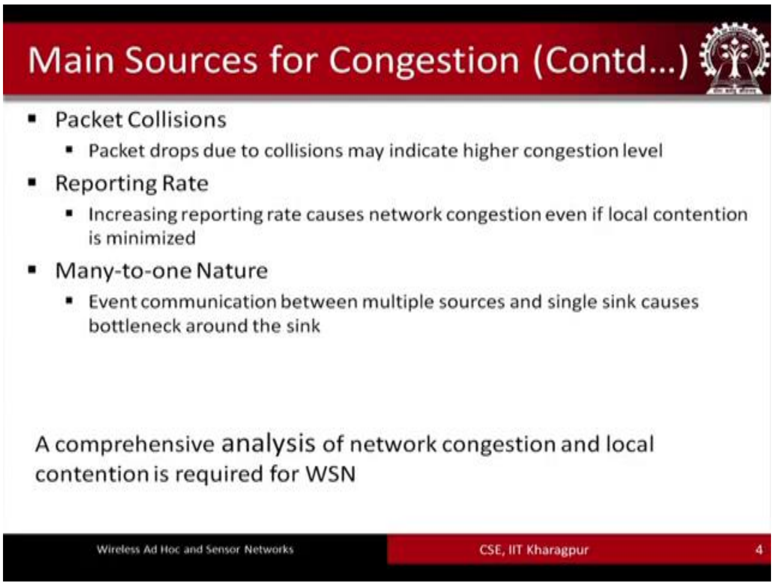


So, what are the main sources of congestion? So, one thing is channel contention and interference. And, this contention can occur in different ways. Contention can occur between the different flows or even different packets of a single flow. So, contention can occur. Second thing is outgoing channel capacity that becomes time variant. And, this is another source of, another source of congestion in sensor networks high density of sensor networks different nodes you know placed closely with one another. And that also has an indirect effect on the contention to the medium, and consequently congestion.

Second thing is the number of event sources. So, what we have? In sensor networks, we have large number of event sources, unlike in most of the other types of networks, not even Ad Hoc networks. In sensor networks, what we have is we have a event which is detected by multiple sources; multiple source nodes. And that is intentionally done. That is done to improve the overall detection efficiency. And so, but that basically has an implication on the congestion.

So, basically if you have large number of source nodes which are detecting an event that will lead to congestion at the receiver nodes. Second is closely located source nodes will increase the contention. Third is increase number of flows, will also increase congestion. So, these are the different sources of congestion in sensor networks.

(Refer Slide Time: 04:50)



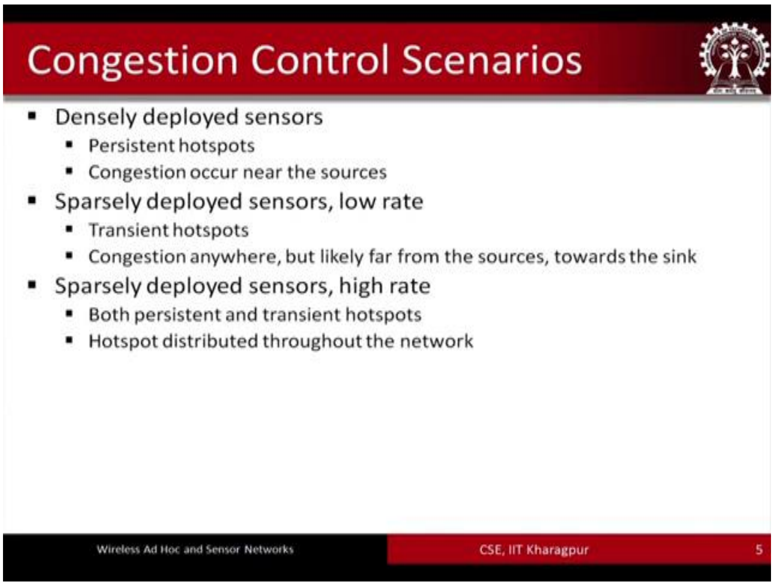
There are few others; packet collisions. Packet drops due to collision may indicate higher level of congestion. Reporting rate; increasing the reporting rate causes network congestion, even if the local contention is minimized and, the many-to-one nature of sensor networks.

So, sensor network, the four structuring sensor networks, if we look at it is somewhat like an inverted funnel like structure. It is an inverted funnel like structure. So, which means that the source nodes once they have detected an event, they will send to their list of neighbor which is towards the sink node, and multiple source nodes which have detected the event.

At the same time, they are all trying to pump in the sales data, at the same time towards their next of neighbors. So, the next of neighbor, if you look at from the next of neighbor's perspective, different nodes, all have sensed data at the same time, once an event has occurred. And, we have a tapering kind of scenario towards the sink of the different flows that are coming from these different source nodes. So, ultimately what we have is a funnel like structure towards the sink. From a sink, if you look at it is an inverted funnel like structure.

So, there, this particular specific type of structure, this inverted funnel like structure or the many-to-one kind of structure, these basically invites issues with congestion. Congestion is likely to occur at the intermediate nodes, which lie between the source node and the sink node. So, what is required is to analyze the different issues with congestion, contention, etcetera that are required for wireless sensor network.

(Refer Slide Time: 06:49)

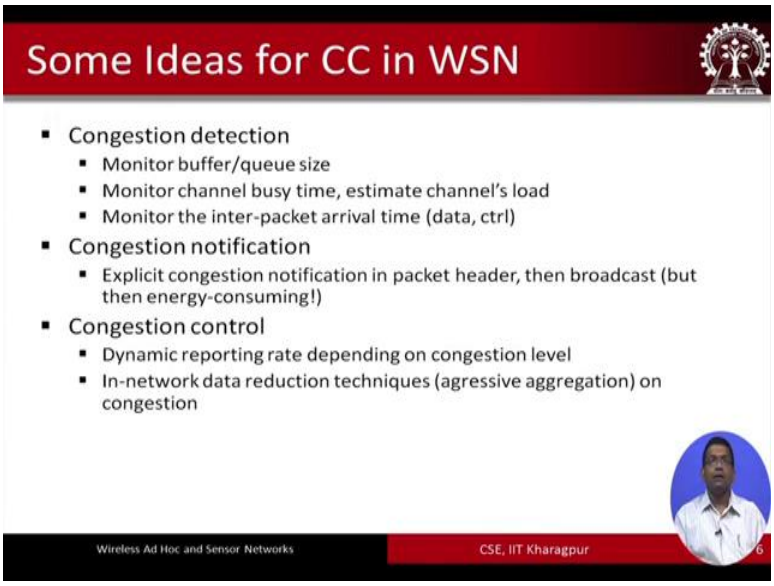


So, if we look at any sensor network these are some of the scenarios that we can have. We have a scenario where all the sensor nodes are densely deployed in the region. We have; we could have a scenario where the sensor nodes are sparsely deployed. Either densely deployed or sparsely deployed. Now the sparsely deployed, we can have in two subdivisions. You know, sparsely deployed nodes pumping data at the slow rate, at reduced rate and sparsely deployed sensors where the data are pumped in to the network at high rate. So, we have three scenarios now. Densely deployed sensors, sparsely deployed sensors at low rate and sparsely deployed sensors at high rate.

So, densely deployed sensors will lead to; it has been analyzed and it has been observed that generally densely deployed sensors in a network will lead to persistent hot spots. Hot spots are the nodes or the regions where congestion is likely to occur leading to packet drops. And, persistent implies that hotspot over time is going to remain over time. So, it is not going to change. So, those nodes which are hot spots will remain as hotspots over longer durations of time. In this sparsely deployed sensors with low rate of data feeding, this will lead to transient hotspots. And, congestion is likely to occur anywhere but, are likely to be far away the sources. I mean, the spots for congestion are likely to be far away from the sources closer to the sink.

Sparsely deployed sensors with high rate will lead to the formation of both persistent and transient hotspots. And, these hot spots that are formed are going to be distributed throughout the network, not just localized in a particular part of the region.

(Refer Slide Time: 08:55)

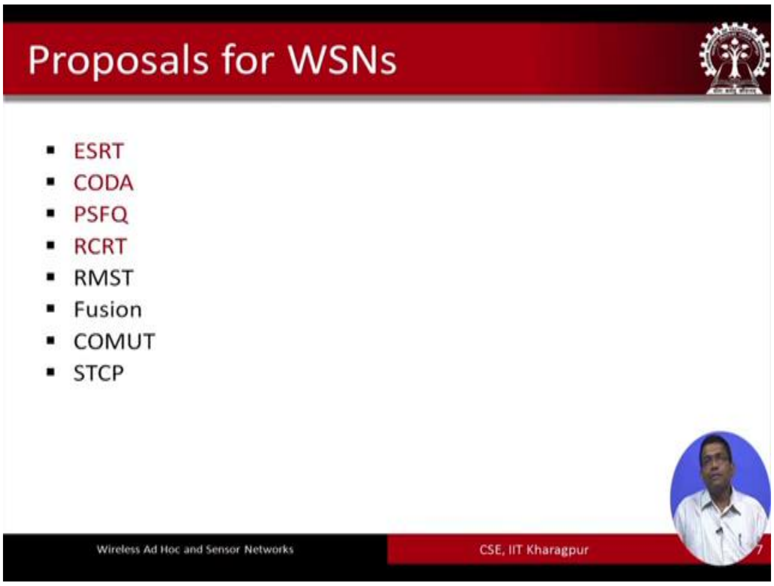


So, there are some issues with congestion control when we talk about it in the context of WSNs. So, when we talk about congestion control? So, even before we talk about congestion control, the first basic thing is to detect congestion. So, after the detection is done that this particular congestion has to be notified. And, the third is then you control congestion. So, these are some of the steps towards congestion control. First, you have to detect congestion. So, for detecting congestion there are couple of different ways in which congestion can be detected. Monitoring the buffer or the queue size at the different nodes; you know how the buffer is getting full. Second is monitoring the channel busy time and estimating the channel's load. Third is monitoring the inter packet arrival time with respect to both data packets as well as control packets. So, this is how you know different ways by which congestion can be detected.

So, the third one is congest, sorry, the second step is congestion notification. Once congestion is detected, it has to be notified. It has to be typically notified to the source node. So, the node which is detected, congestion has to explicitly notify in the form of some kind of sitting a bit or something like that in the packet header that congestion is occurring. So, either you notify to the source node or it can be broadcasted and some of the nodes in the neighborhood, and maybe all the nodes in the network will eventually come to know that this is the place where congestion is occurring.

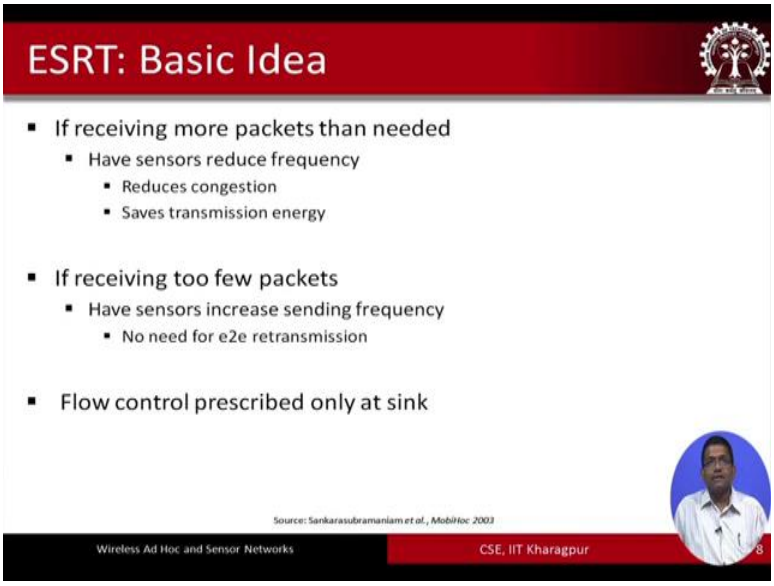
Third is congestion control, where this basically involves dynamic reporting of the rate depending on the congestion level. And, so basically you know if congestion is coming closer or it is happening, then somehow that to control congestion, the rate at which the source node should send the data that has to be reported, that has to be controlled by the node which is affected by congestion. So, in network data reduction techniques on congestion are required in order to control congestion.

(Refer Slide Time: 11:09)



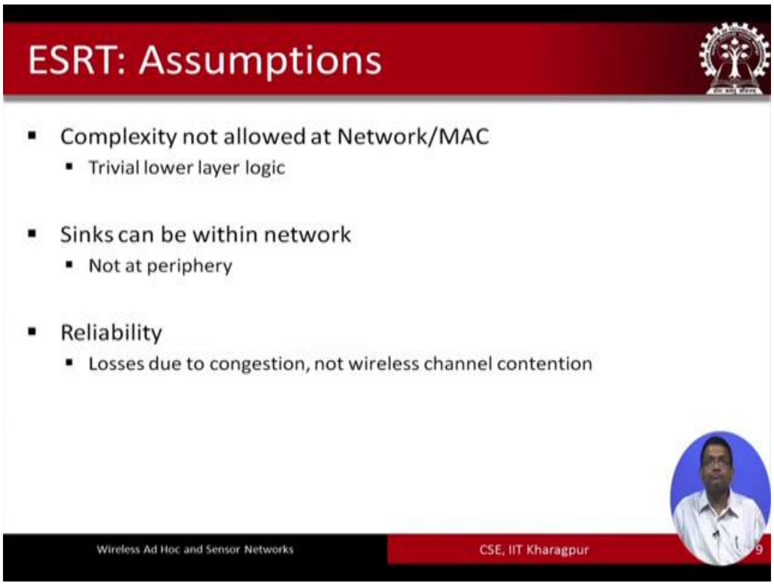
So, different protocols, I told you already have been proposed for wireless sensor networks. Here are list of different protocols that have been proposed like this. Actually, this is not the complete list. But, there are many other protocols that have been proposed to, you know to basically control congestion. The ones that are marked in red color are the ones that I am, that we are going to focus on in our lecture. So, ESRT, CODA, PSFQ and RCRT are the ones that we are going to discuss in this particular lecture and the next one, the next part of it.

(Refer Slide Time: 11:49)



So, let us start with the ESRT. So ESRT, there are few basics. We have to understand those first. So, first thing is that ESRT says that if more packets are received than are needed, then have the sensors to reduce the frequency which sensors the source sensors. So, source sensor nodes reduce the frequency of sending the packets. This will lead to reduce congestion and will also save transmission energy. If receiving too few packets, this is also not good. If fewer number of packets, then what can be handled are received by the intermediate nodes or the intended destination node or the sink node. Then, have the sensors; that means, the source sensor nodes increase the sending frequency. And, there is no need for end-to-end retransmission of packets because there is no packet drop that is going to happen. So, there is no need for in end-to-end retransmission of packets. And, ESRT says that the flow control should be done at the sink node. So, sink should control the flow of packets from the source through the intermediate nodes to the sink.

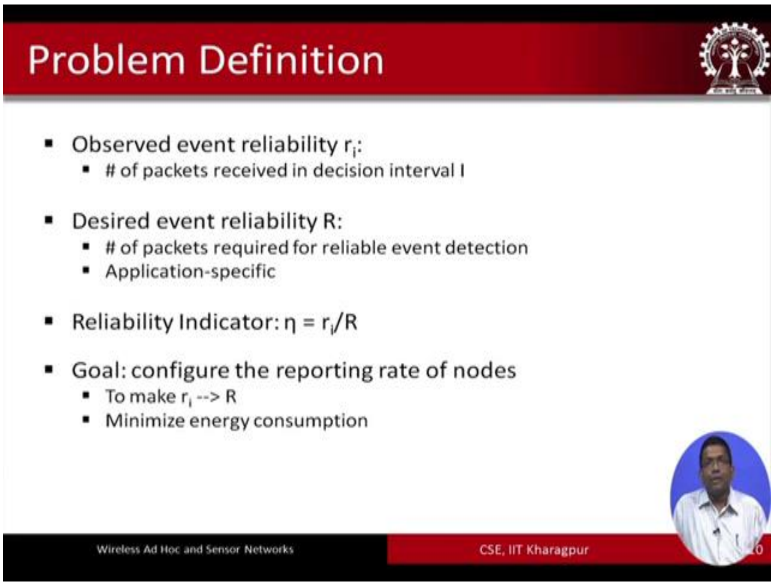
(Refer Slide Time: 13:00)



Here are few assumptions that ESRT makes. So, first thing is that there is not much complexity that is allowed at the network layer or the MAC layer; that means, you know when we are talking about transport layer, the bottom layers are network layer and MAC layer. So, trivial lower layer logic; that means network layer logic and MAC layer logic are implemented. Then, the second is the sinks can be within the network, not necessarily at the periphery of the network. Sinks can be existing within the network itself. And, the third is the reliability.

So, that means that the losses that are occurring, packet losses that are occurring are only due to congestion, and not due to any other reason such as wireless channel congestion and so on. So, this is how the problem of congestion has been defined in a ESRT protocol.

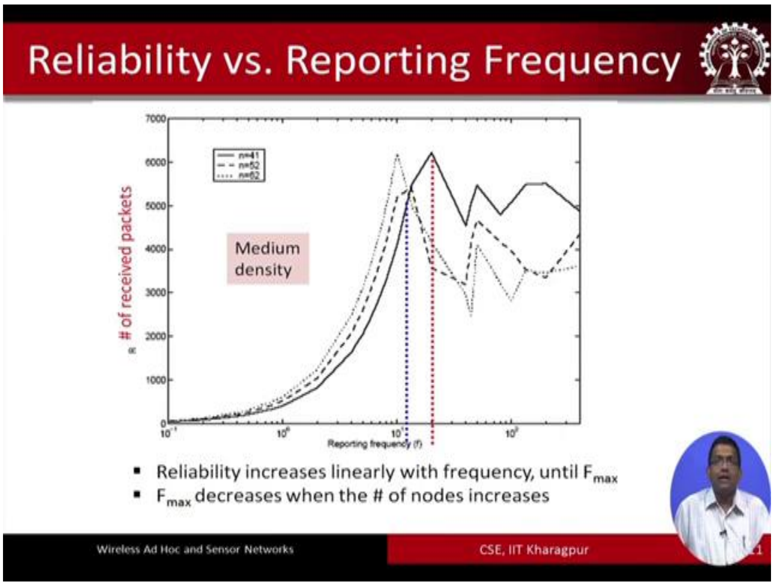
(Refer Slide Time: 13:49)



So, here are a few parameters that are defined. First, the observed event reliability is denoted as r i. Observed event reliability r i, which basically is the number of packets that are received, in a particular decision interval. The second is capital R, which is the desired event reliability which is defined as the number of packets that are required for reliable event detection. And, this particular parameter is application specific. So, with taking the ratio of the observed event reliability and the desired event reliability, what we get? This particular ratio is denoted as the, is defined as the reliability indicator.

So, the reliability indicator is the ratio between the observed event reliability and the desired event reliability. So, the overall goal is to configure the reporting rate of the nodes in order to make this particular ratio r i divided by R tending to 1; that means, r i tending to the value R, and thereby minimizing energy consumption.

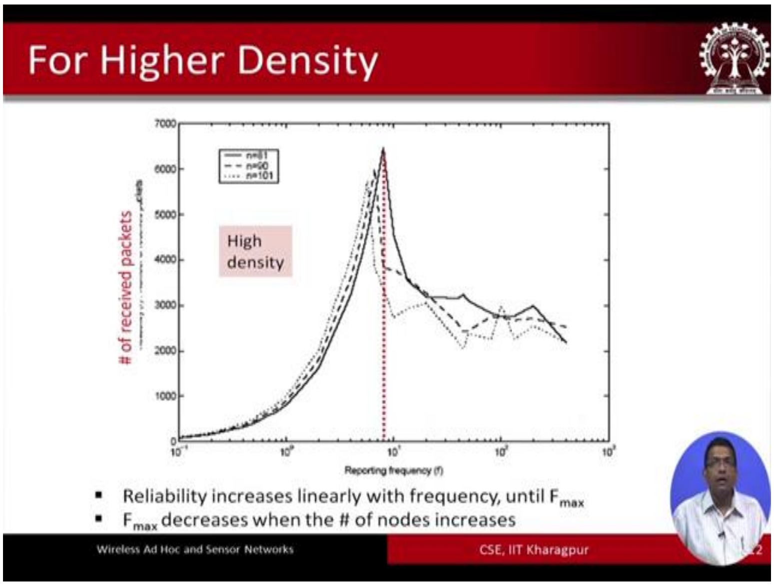
(Refer Slide Time: 15:19)



So, looking at a few plots. Let us look at and try to understand this plot first. So, this is a plot of the number of received packets versus the reporting frequency. And, there are three plots as we can see over here. This particular, the continuous one here is for n equal to 41; n equal to 41 means the number of nodes is 41. Then, we have n equal to 52. The next one is n equal to 52. And, a third one; that means, this one is for n equal to 62, 41 nodes, then 52 nodes in the network and 62 nodes in the network. So, this basically represents a scenario of medium density of network. So, this is what has been with a, you know, with a fixed size. Within a fixed size, you know different nodes being plotted in such a way that we get a medium density network. And, we observed as we can see over here that as the number of nodes increases this particular peak that is obtained, also shifts leftwards. Leftwards means that as we can see that the reporting frequency gets, you know we have reduced reporting frequency and the reduced number of received packets.

So, consequently from this, basically we can infer that the reliability basically increases linearly with the frequency. And thereafter, once the peak is obtained, you know after that, the number of received packets decreases. So, first we see that the number of packets increases. If we consider each of them, we, this is what we observe. The number of packets received increases first to a certain value and then decreases thereafter. So, this is what we see for all the three configurations, three different scenarios of n equal to 41, 52 and 62. So, that was for the medium density network.

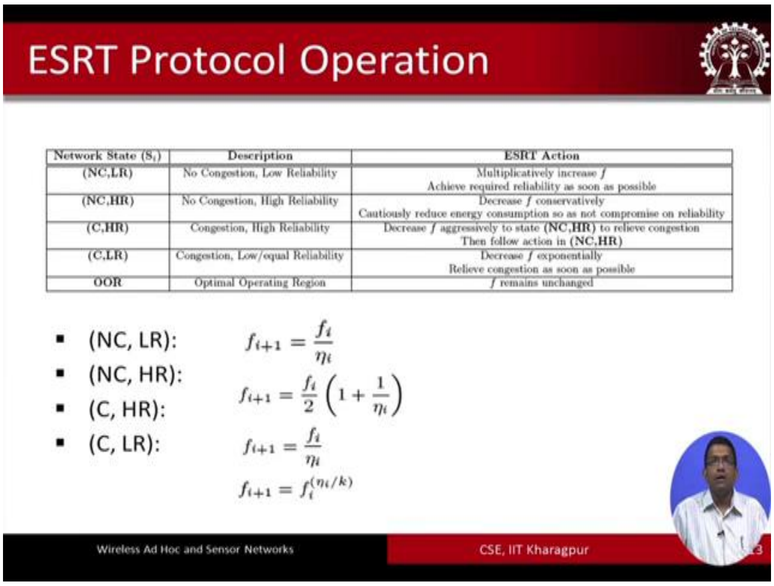
(Refer Slide Time: 17:45)



Now, let us look at this particular plot. So, this is which is for a highly dense network; high density network. So, it is showing n equal to 81, 90 and 101. So, this is the first one, the middle one. The innermost one is for n equal to 81. Then going outwards, we have n equal to 90 and 101. So, this is a scenario. In a similar kind of pattern, we can see over here as well. So, initially with the increase in the reporting frequency, the number of received packets also increases to a certain value. And thereafter, it drops quite drastically.

So, the reliability basically increases linearly with the frequency until a certain value f max. And, this f max decreases when the number of nodes increases. So, f max, this is the f max value. It decreases when the number of nodes increases or builds the report. Maximum reporting frequency decreases when the number of nodes increases. And, both of these results for the medium density as well as the high density network. These are quite intuitive. And, we do not need to spend time analyzing these. So, these observations that we have are quite evident and understandable.

(Refer Slide Time: 18:58)

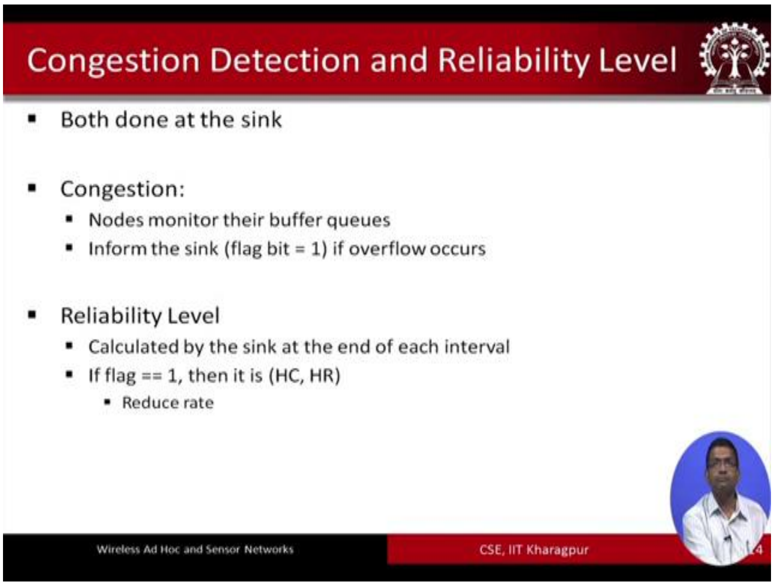


So, the ESRT protocol operation is like this that it starts. So, the network basically is considered to be in different states. So, initially we have this particular; you know, so this particular network, state the first one; NC, LR. Signifying that we have a network with no congestion and low reliability then, we have the congestion low congestion, high reliability, then we have congestion high reliability, congestion low or equal reliability and optimal operating region OOR. So, network is the optimal operating region. So, which means that you know it is performing optimally over all.

So, let us now see what actual ESRT prescribes corresponding to these different states of the network. So, if the sensor network is in a no congestion, low reliability state, then ESRT says that you multiplicatively increase, increase if; that means, the reporting frequency. And, this is achievable, achieved and achieve the required reliability as soon as possible. And then, you know achieve the required reliability as soon as possible, if it is a state of no congestion and high reliability. ESRT prescribes these actions decrease if conservatively and cautiously reduce the energy consumption so as not to compromise on reliability. Then, we have the case of congestion and high reliability, where it prescribes, ESRT prescribes that you decrease the f value aggressively to the state NC, HR; that means, no congestion, high reliability to relief congestion.

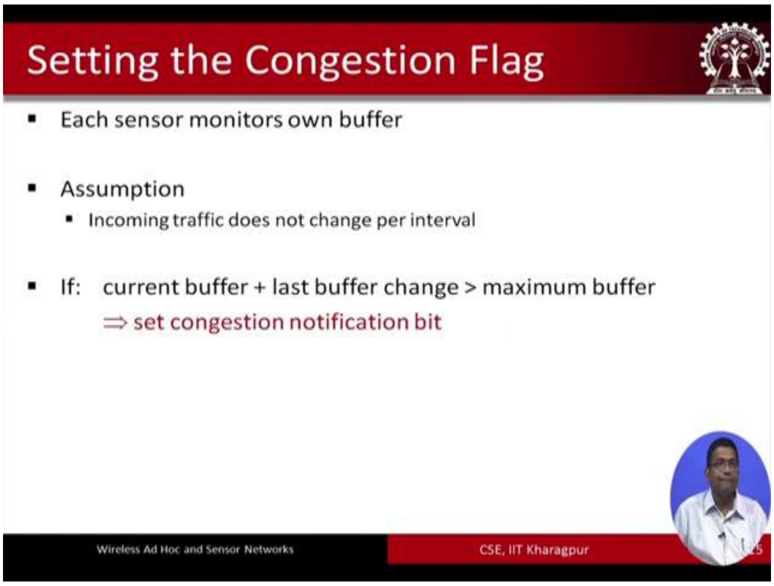
And then, follow the action in NC, HR. And likewise, you know for the other two states as well. So, how these are obtained? These prescriptions, the analysis of it are there in the ESRT protocol paper, which whose reference is given at the end. So, these are the different formulas which capture the action of the ESRT protocol. Corresponding to the different states, it is shown here. And, f is the reporting frequency. So f, the first one means that f i plus 1; that means, at the next iteration, next time interval f i 1 i plus 1 equal to f i; that means, the current one reporting frequency divided by this eta i, which is basically the reliability factor. So, eta i, if you recall is the reliability factor, reliability indicator, which is the ratio between the observed event reliability and the desired event reliability.

(Refer Slide Time: 21:46)

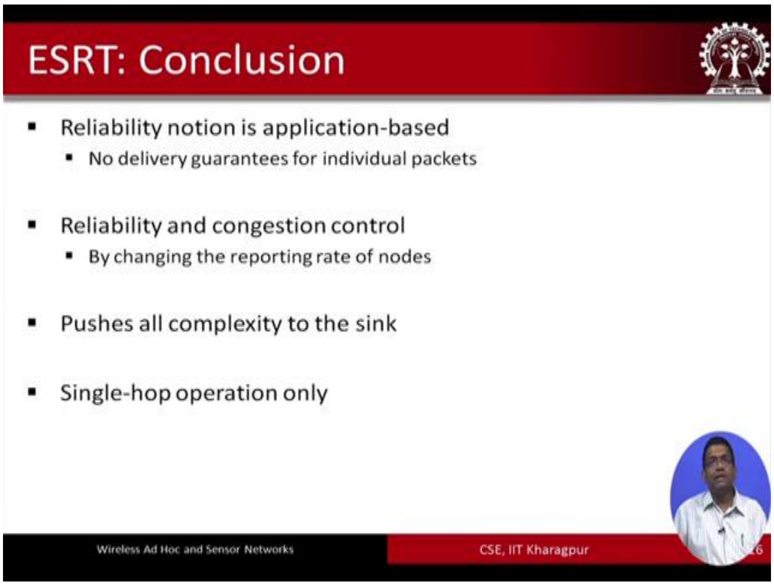


So, both congestion detection and reliability level taking care of it are done at the sink. In this particular protocol, ESRT congestion is taken care in this way that the nodes, monitor, their buffer sizes. And, they inform the sink by setting the flag bit of the header of the packets, if overflow occurs. For the reliability level, it is calculated by the sink at the end of each interval. And, if the flag bit is set to 1, then it is considered to be a case of high congestion, high reliability. And, the recommendation in this particular case is to reduce the rate of reporting.

(Refer Slide Time: 22:29)

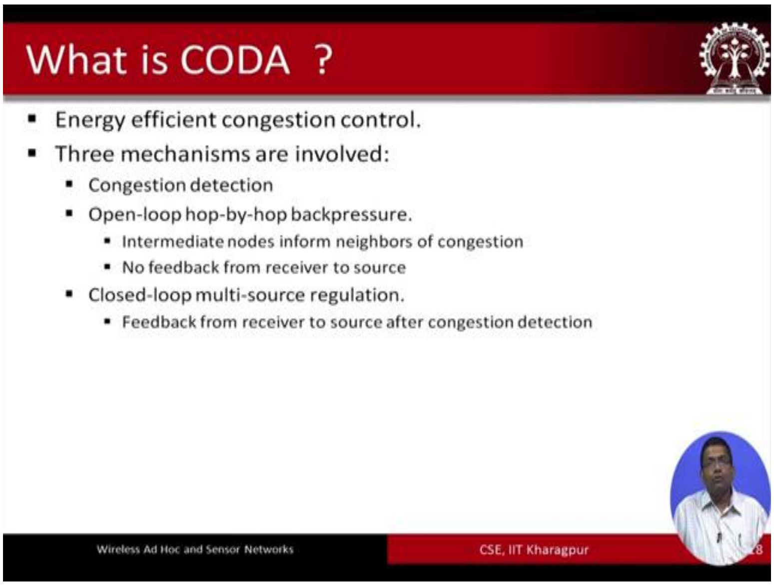


(Refer Slide Time: 22:35)



So, we will skip this. And, these are some of the different concluding points, concluding remarks about ESRT. The reliability notion that is prescribed in ESRT is application specific. And, there is no delivery guarantee for individual packets. Reliability and congestion control are changed by changing the reporting rate of the nodes in ESRT and all the complexity decision making, etcetera, etcetera are controlled from the sink. It is what we have seen. So, this is what we have seen that the sink node basically controls of these operations. So, this is what ESRT does. The next protocol after ESRT is the CODA protocol. The full form of which is congestion detection and avoidance.

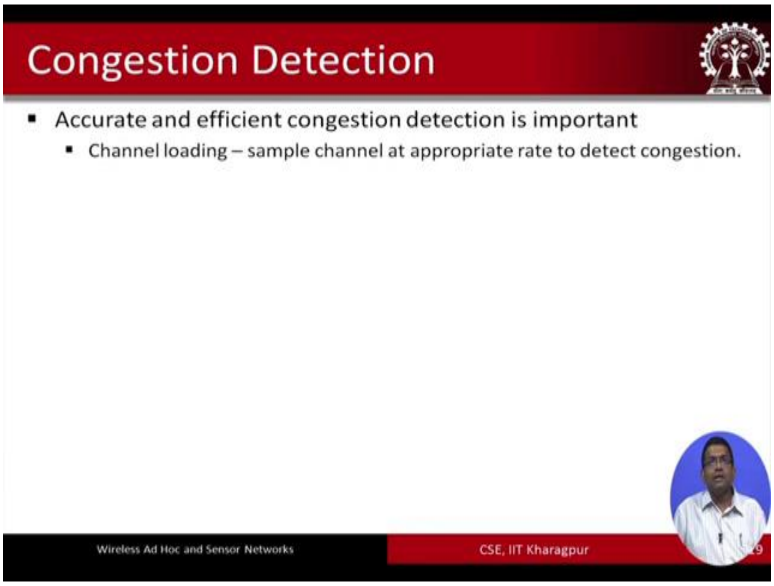
(Refer Slide Time: 23:28)



So, only CODA basically takes the energy issue also into account. And, there are three mechanisms that are involved in implementing CODA. First thing is that CODA detects congestion. Second is after congestion is detected, it prescribes a open loop hop by hop backpressure mechanism, where the intermediate nodes inform the neighbors of congestion. And, there is no feedback from the receiver towards the source.

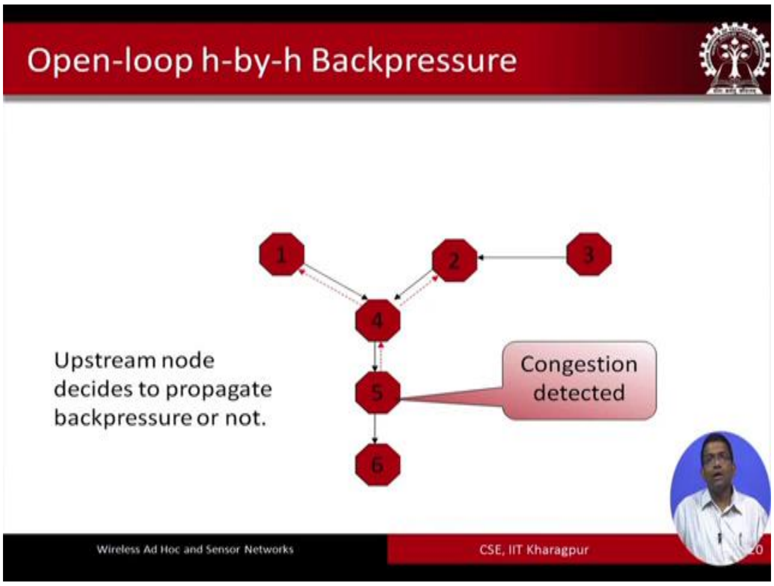
The second is the closed loop multi source regulation, where the feedback from the receiver is sent to the source after congestion is detected. So, the difference between open loop and close loop mechanisms are in open loop the intermediate nodes are informed of congestion. And, in closed loop the receiver basically informs the source about the congestion, occurrence of congestion. So, explicitly it will notify the source node.

(Refer Slide Time: 24:26)



So, congestion detection is all about accurately and efficiently detecting congestion. And, so this is very important issue.

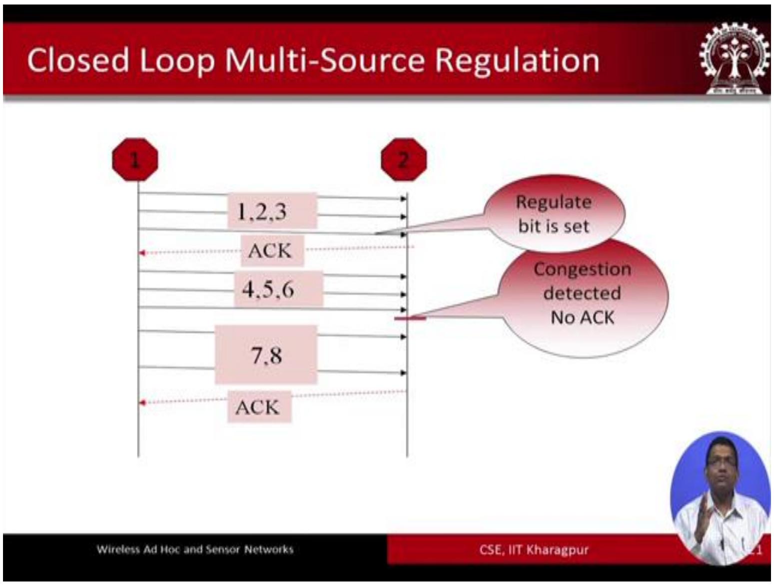
(Refer Slide Time: 24:40)



So, this is what a CODA basically prescribes that after congestion is detected, perform open loop hop by hop backpressure. So, which as we have seen in open loop, what is done is to inform the neighbors. So, in this particular example let us consider that a node s, a node 5 detects that there is some congestion. So, then this particular thing is notified to the neighbor node 4, this red colored arrow, dotted arrow. Basically, shows it that it is notified to the red color to the neighbor, which is towards the source node.

So, basically this neighbor is the one from which 5 had received the packet. So, it has detected congestion. So, it will notify this node 4. And from here, then what we have is the node 4 also does the same. It will, in fact notify its neighbors. So, nodes 1 and 2, among which one of them is the source node is also going to get notified about, it occur, occurrence of congestion. So, as you can see that it is a backpressure kind of mechanism, hop by hop backpressure, from 5 to 4, 4 to 1 or 4 to 2 and so on. So, hop by hop, this backpressure mechanism is sent towards, sent to the neighbors and eventually from, towards from the neighbors towards the source node.

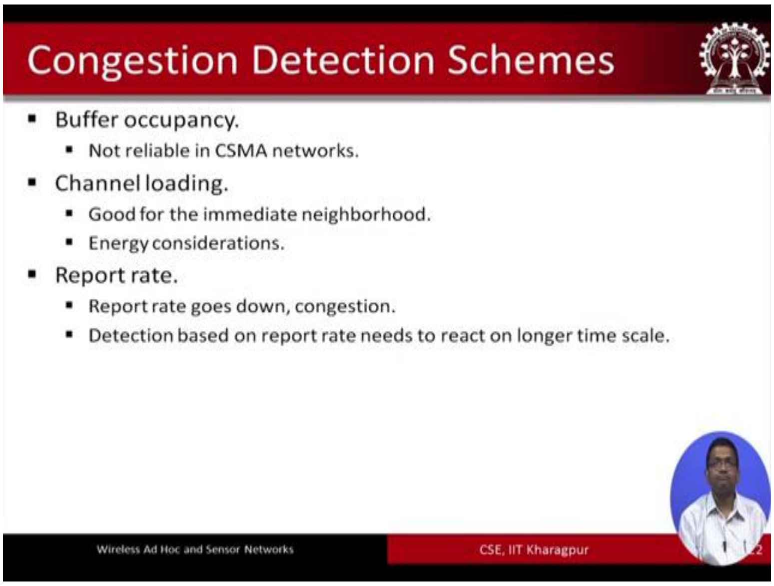
(Refer Slide Time: 26:13)



So, this is one thing. And, the second thing is that after hop, you know this open loop approach, the closed loop approach is followed, where explicitly the source node is notified. So, in this particular case that let us consider we have a scenario like this. We have the node 1 and node 2. Let us say that three packets 1, 2 and 3 are sent. And then, the corresponding acknowledgement is sent back then 4, 5, 6. And after 4, 5, 6, as you can see over here that there is no acknowledgement sent because congestion has been detected, you know, sending of the acknowledgement would imply acknowledgement with the regulate bit set would imply that congestion is taking place.

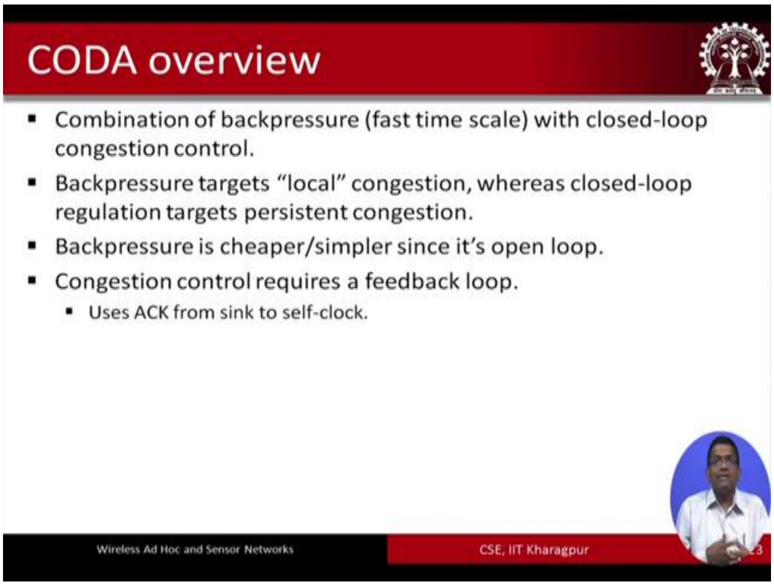
So, congestion is taking place. And, so in this particular case the acknowledgement is not this. It was not send by 2 towards 1. And then, let us say that packet 7 and 8 are received. And thereafter, acknowledgement is sent. So, this means that things have been become and there, you know congestion has been taken here. And, you know so the acknowledgement is going to be sent back to signify that.

(Refer Slide Time: 27:34)



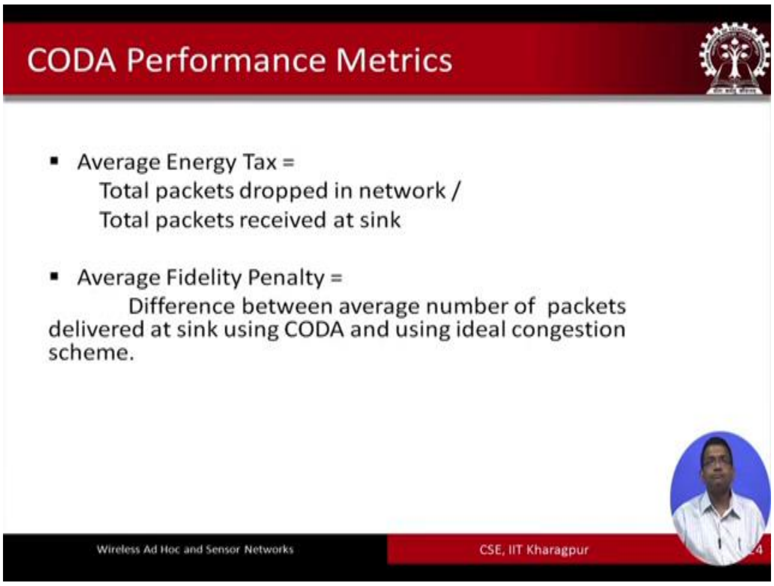
So, there are different issues of congestion detection. Buffer occupancy is very important. And, this is not a very reliable thing in CSMA class of networks channel. Loading is another important issue and the third is the reporting rate. So, these are different issues that have to be taken care of for detecting congestion.

(Refer Slide Time: 27:56)



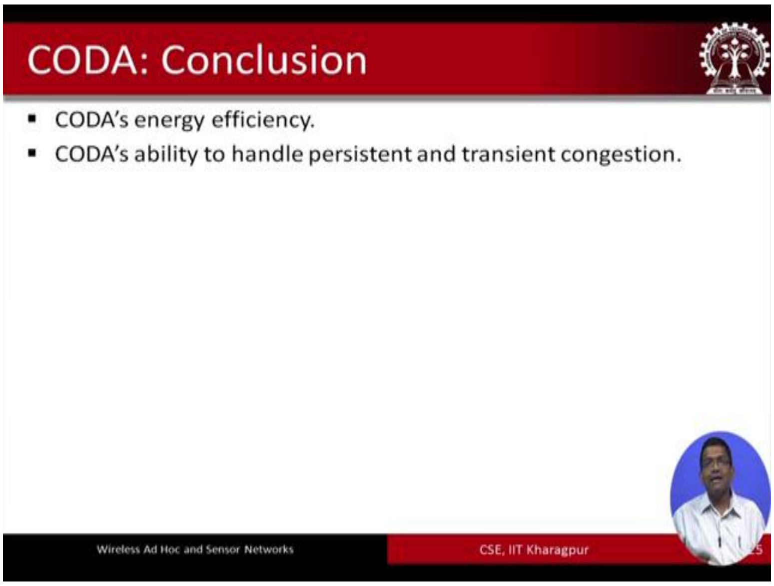
So, CODA basically what it does is after congestion is detected, it combines the back pressure; that means, open loop approach with the closed loop approach for congestion control. So, the backpressure mechanism, the open loop one basically targets local congestions; whereas, the closed loop one basically regulates the target. Closed loop regulation targets the persistent congestion. So, this is what the closed loop one. Persistent one is not a temporal one. It, you know, so it stays for longer duration of time. And that is how the source node has to be notified explicitly. And, the source from the source node towards the node which has experienced congestion, the congestion control mechanism is going to be invoked.

(Refer Slide Time: 28:41)



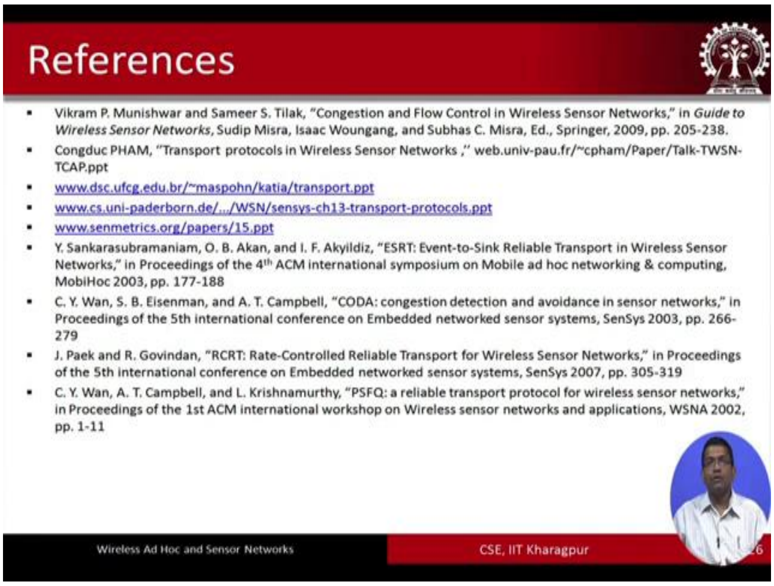
Here, some performance metrics that are considered by the code of CODA protocol for evaluating its performance and comparing it with the other congestion control mechanisms. Average energy tax is one which is basically the ratio between the total packets that are dropped in the network to the total number of packets that are received at the sink. And, the second one is the average fidelity penalty, which is basically the difference between the average number of packets delivered at the sink using CODA and using ideal congestion scheme. So, these are the different performance metrics that have been proposed.

(Refer Slide Time: 29:17)



So, CODA basically has explicitly compared to ESRT, it has explicitly taken energy issues into consideration for, you know, detection and control of congestion. And, CODA is able to handle persistent as well as transient congestions successfully. This is what the CODA protocol and the corresponding paper basically reports.

(Refer Slide Time: 29:43)



So, here are list of differences for not only these protocols that we have discussed, but also the other protocols that we will discuss in the next set of slides. And, all these differences are there. And, some of the, some of the slides that we have taken are basically adopted with minor modifications from different sources. And, these sources are given in these references.

So with this, we come to an end of the first part of congestions and flow control in sensor networks.

Thank you.