Introduction To Multimodal Urban Transportation System

Prof. Arkopal Kishore Goswami

Department of RCG School of Infrastructure Design And Management

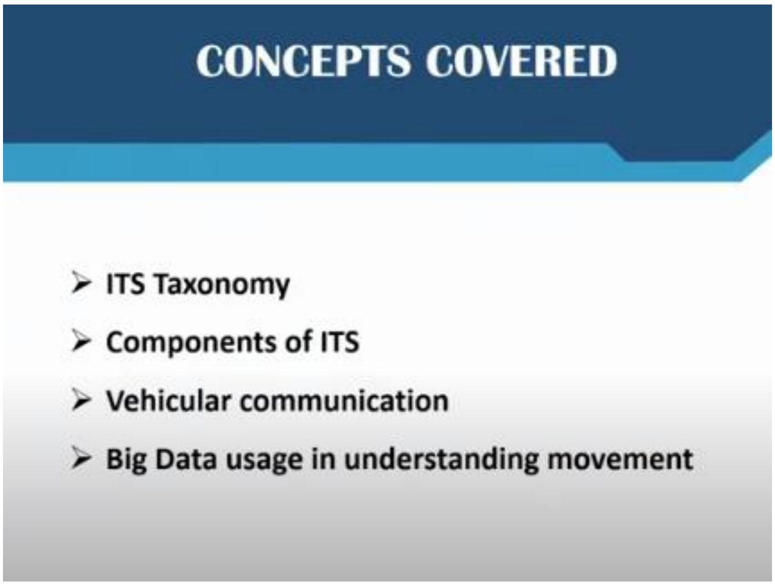
Indian Institute of Technology-Kharagpur

Lecture - 47

ITS Components, Applications and Communication

Welcome back friends. I hope you have enjoyed the first lecture of on ITS.

(Refer Slide Time: 00:29)

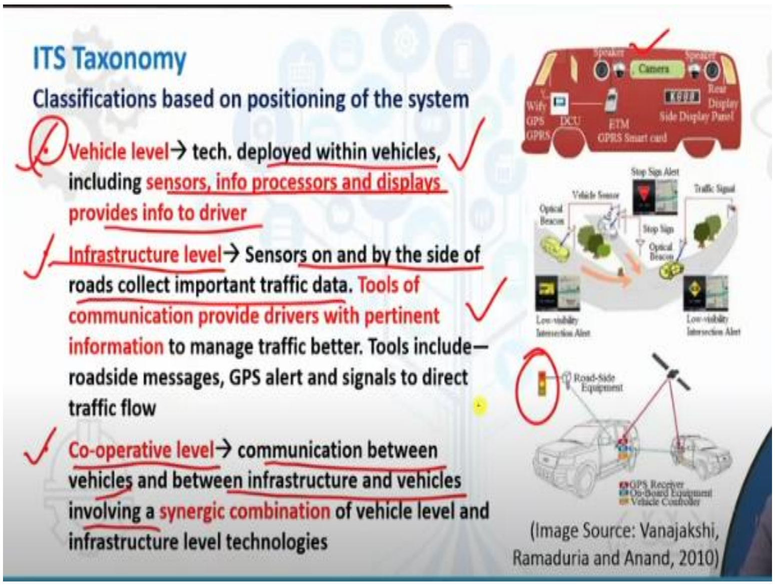


CONCEPTS COVERED

- ITS Taxonomy
- Components of ITS
- Vehicular communication
- Big Data usage in understanding movement

In this lecture, we are going to introduce you to some of the basic taxonomy in ITS; break down the ITS into its different components; tell you about how vehicular communication happens, and also give you an understanding of big data usage in transportation.

(Refer Slide Time: 00:46)



ITS Taxonomy

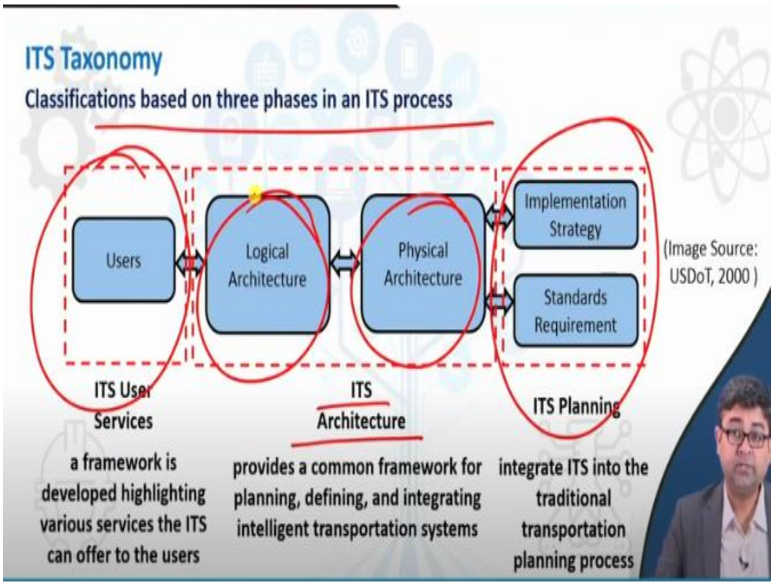
Classifications based on positioning of the system

- Vehicle level→ tech. deployed within vehicles, including sensors, info processors and displays provides info to driver
- Infrastructure level→ Sensors on and by the side of roads collect important traffic data. Tools of communication provide drivers with pertinent information to manage traffic better. Tools include– roadside messages, GPS alert and signals to direct traffic flow
- Co-operative level→ communication between vehicles and between infrastructure and vehicles involving a synergic combination of vehicle level and infrastructure level technologies

(Image Source: Vanajakshi, Ramaduria and Anand, 2010)

When we are talking about ITS, we usually are talking about three different levels. We are either talking at the level of the vehicle or the level of the infrastructure or at a co-operative level. So when we are talking at a vehicular level, we are talking about technology that can be deployed within the vehicles including sensors, processors, as well as displayers which provide information to the driver. There are various cameras that can be fitted within a vehicle that could be sensors that allows the vehicle to not sway too much in the lane left or right. There could be parking sensors. Now when you put a vehicle in reverse gear, it has a camera at the back of the vehicle that allows you to see how much space is there and you can then back into back into a parking space. All of those systems that are in the vehicle are the vehicle level classification of ITS devices. Next is the infrastructure level, which are all the sensors that are on and by the side of the road that collect information about traffic. It can be your vehicle’s signals, it can be at the toll plazas that where your transactions are happening. So all of those kind of devices that are either on the side of the road or on top of the road collecting information about traffic are known as infrastructure level ITS devices. And then there are cooperative level, which is the communication between the infrastructure and vehicles. It is one thing to have just a vehicle level or an infrastructure level ITS capabilities, but unless and until there is communication between the vehicle and the infrastructure, higher efficiency is not going to be achieved. Certainly has to be two way communication between the infrastructure and your vehicle. So for example, in the last lecture, we have told you that how two way communication can be achieved of an ambulance or an emergency vehicle at a signalized intersection. So the signal detects the vehicle that is a communication from the infrastructure to the vehicle. And then the vehicle communicates back saying I am an emergency vehicle, give me right of way. So then that is communication from the vehicle to the infrastructure. And then the infrastructure turns, the signal turns green. That is kind of two way communication and co-operative level ITS devices.

(Refer Slide Time: 03:46)



ITS Taxonomy

Classifications based on three phases in an ITS process

(Image Source: USDoT, 2000)

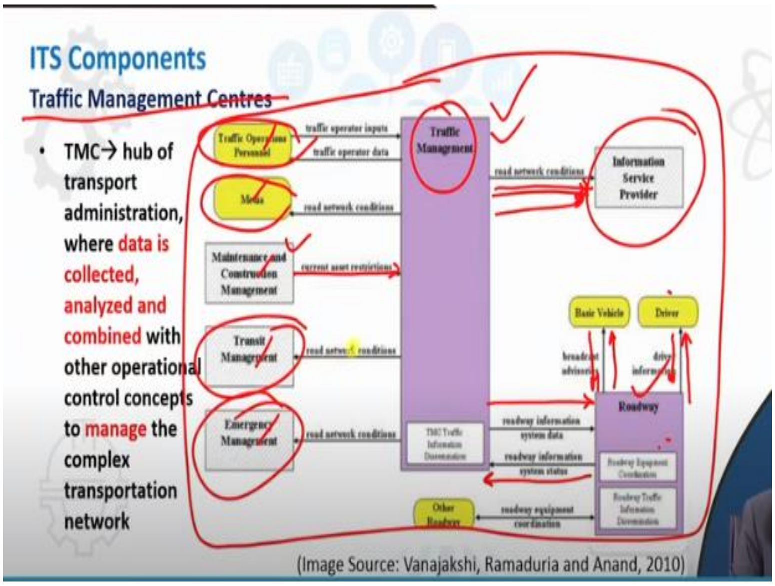
ITS User Services: a framework is developed highlighting various services the ITS can offer to the users

ITS Architecture: provides a common framework for planning, defining, and integrating intelligent transportation systems

ITS Planning: integrate ITS into the traditional transportation planning process

When we look at the classification based on three phases, what we see is there are users on one side and then there are all the strategies and the planning segment on the other side. Users want to have an efficient transportation system. Governments, agencies want to devise their policies, develop their priorities based on the existing system. And in between are all the different ITS architecture that needs to be developed in order to satisfy on both ends. So when you look at an ITS and we will give you an understanding of what is a logical architecture and what is a physical architecture. These are basically how your information flow happens. What are all the hardware and the softwares that are needed in order for a system to be created? Should the information be only one way flow, should be should it be only two way flow, what kind of information should be shared between the different stakeholders, so on and so forth. So all of this is a very interesting new field, a new arena into which transportation professionals are getting into like I said in the previous lecture as well. And this should be something that opens up newer avenues for you as well.

(Refer Slide Time: 05:24)



ITS Components

Traffic Management Centres

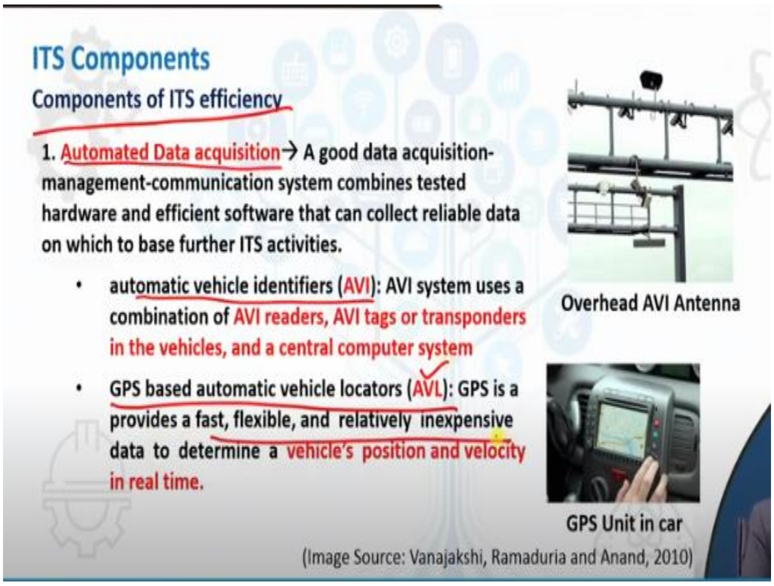
- TMC→ hub of transport administration, where data is collected, analyzed and combined with other operational control concepts to manage the complex transportation network

Traffic Operations Personnel
traffic operator inputs
traffic operator data
Traffic Management
Media
road network conditions
Maintenance and Construction Management
current asset restrictions
Transit Management
road network conditions
Emergency Management
road network conditions
TMC Traffic Information Dissemination
Information Service Provider
road network conditions
Basic Vehicle
Driver
broadcast advisories
driver information
Roadway
roadway information system data
roadway information system status
Roadway Equipment Coordination
Roadway Traffic Information Dissemination
Other Roadway
roadway equipment coordination

(Image Source: Vanajakshi, Ramaduria and Anand, 2010)

We have already told you about what a traffic management center is. The one in the previous lecture that we had showed in Bhopal is the first traffic management center that has opened as a part of the smart city. If you were to understand how information flows between the traffic management center and various other entities, you would have to develop something called a physical architecture which shows the traffic management center and how it is linked to the roadway; what information is the traffic management center sharing with the roadway and what is the roadway sharing back with the traffic management center. Furthermore, how the roadway is connected to the driver as well as to the vehicle. So what information is the roadway getting from or giving to the vehicle and the driver? Is it getting back anything? If it does not seem to be getting any information back then it is a one way type of information that is being given. Similarly, roadway network condition information is going out from the traffic management center to the different ISPs (Information Service Providers). So Google and all are getting all the information about the roadway network condition from these TMCs (Traffic Management Centers) and it is a one way information that is going out. Similarly, there is a lot of one way information that is may be coming into the TMC. It may be the maintenance or construction management. If there is a roadway construction that is going on in one part of the city, that information must be only coming in to the TMC which is then maybe converted into travel time or travel delay, and then forwarding it to the ISPs. A very simplistic diagram allows you to understand how ITS system may work. This is an ITS system meaning this is a traffic management system. How is it connected to the roadway? How is it connected to the basic drivers? How is it connected to the for example, media as well. So if you have to provide information on the TV or on the radio about the traffic conditions, it can they can directly get information from the traffic management center. Traffic operation personnel can get information from this and of course, different bus transit managements and emergency managements can get information on this. So the basic thing to understand is who all the stakeholders are that need to get data or from whom we need to receive data. What kind of data we need to get. And does the data have to be a two way communication or one way communication and so on and so forth. These basic things if one understands then one would understand how to develop a ITS system for their city.

(Refer Slide Time: 08:36)



ITS Components

Components of ITS efficiency

1. Automated Data acquisition→ A good data acquisition-management-communication system combines tested hardware and efficient software that can collect reliable data on which to base further ITS activities.

- automatic vehicle identifiers (AVI): AVI system uses a combination of AVI readers, AVI tags or transponders in the vehicles, and a central computer system
- GPS based automatic vehicle locators (AVL): GPS is a provides a fast, flexible, and relatively inexpensive data to determine a vehicle's position and velocity in real time.

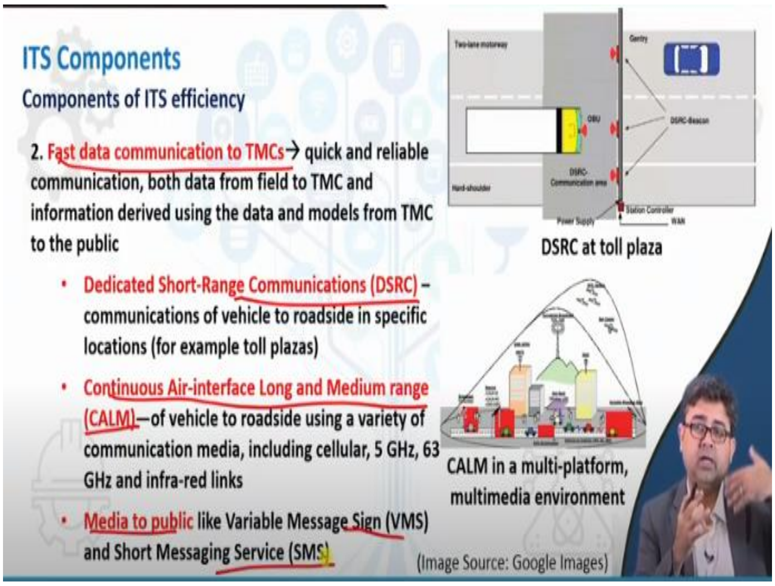
Overhead AVI Antenna

GPS Unit in car

(Image Source: Vanajakshi, Ramaduria and Anand, 2010)

What are the basic components of ITS's efficiency? How does it improve the efficiency in your urban traffic network? First is it allows for automated data acquisitions. We have already given you a brief understanding of that. Automatic Vehicle Identifiers (AVI) are now nowadays used especially in public transportation arena, where can have these AVI sensors on your buses and you can easily locate where the bus is. Based on that you can put up information at various bus stops as to when the next bus is arriving and so on and so forth. You can also use GPS technology, automated vehicle locators AVL, which provides fast flexible and relatively inexpensive data to determine a vehicle's position and velocity in real time.

(Refer Slide Time: 09:35)



ITS Components

Components of ITS efficiency

2. Fast data communication to TMCs→ quick and reliable communication, both data from field to TMC and information derived using the data and models from TMC to the public

- Dedicated Short-Range Communications (DSRC) – communications of vehicle to roadside in specific locations (for example toll plazas)
- Continuous Air-interface Long and Medium range (CALM)—of vehicle to roadside using a variety of communication media, including cellular, 5 GHz, 63 GHz and infra-red links
- Media to public like Variable Message Sign (VMS) and Short Messaging Service (SMS)

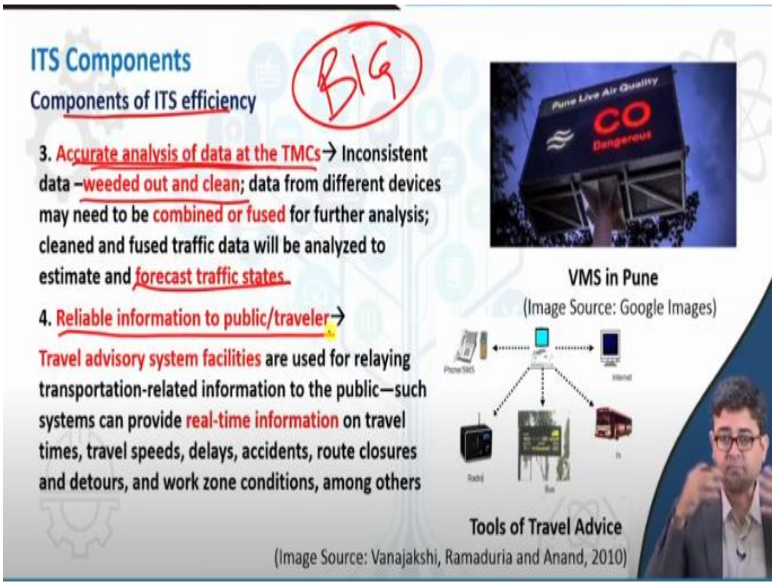
DSRC at toll plaza

CALM in a multi-platform, multimedia environment

(Image Source: Google Images)

Now once you have that data it has to be now communicated to the TMCs as well. So different types of communication protocols can be used. DSRC (dedicated short range communications). CALM (continuous air interface, long and medium) range. And media to public like VMS, short messaging signs. You have VMS is nothing but variable message signs that you might have seen on the side of the road saying that slow down, traffic ahead or accident ahead, that can be easily controlled. Those messages can be controlled from the traffic management center based on data that it is receiving from vehicles or incidents downstream.

(Refer Slide Time: 10:21)



ITS Components

Components of ITS efficiency

BIG

3. Accurate analysis of data at the TMCs→ Inconsistent data –weeded out and clean; data from different devices may need to be combined or fused for further analysis; cleaned and fused traffic data will be analyzed to estimate and forecast traffic states.

4. Reliable information to public/traveler→

Travel advisory system facilities are used for relaying transportation-related information to the public—such systems can provide real-time information on travel times, travel speeds, delays, accidents, route closures and detours, and work zone conditions, among others

VMS in Pune

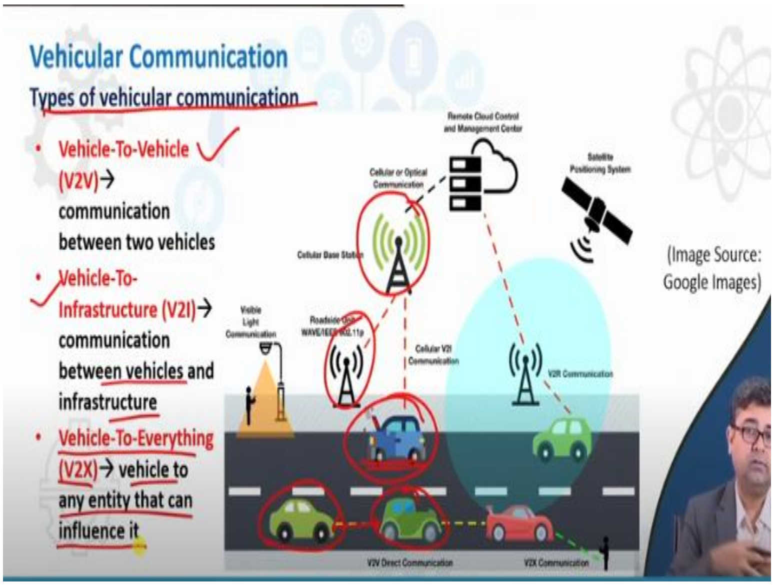
(Image Source: Google Images)

Tools of Travel Advice

(Image Source: Vanajakshi, Ramaduria and Anand, 2010)

The other components of ITS efficiency is accurate analysis of the data. Now when we are talking about all this data, remember this is we are talking about big data. Big data meaning large volumes of data. Now from these large volumes of data, you have to make sense of what the data suggests and how much data you need. Based on that that much volume of data you have to convert now the data into information. It is not very efficient to just have CCTV cameras everywhere. You have to convert the feed of the CCTV cameras into some sort of information. If you have CCTV cameras and you are looking at the intersection, but you are still manually changing the signal timings that does not help. You improve the efficiency of that traffic signal based on the traffic information for example. Maybe the traffic signal feed that you are getting, you have to then try to automate as to when how to change the traffic signal automatically based on data that is coming. So that is an example of what we mean by saying accurate analysis of the data. Poor data has to be weeded out, data has to be cleaned. There will be a lot of noise in the data. Some data are outliers. So we have to check for the outliers as well, whether they are useful information or not. All that has to be done before so as to just establish whether your system is able to capture the current state accurately or not. Once it is able to capture the current state or in other terms once it is calibrated only then you can use it for forecasting traffic for different states, different traffic conditions and so on and so forth. Next, it has to provide reliable information to the public as well as the traveler. Just providing information is not good enough. It has to provide reliable information. For example, the next bus is arriving in arriving at 5:15pm. That bus at 5:15pm has to be reliable. If you do not provide reliable information, what usually tends to happen is the user will no longer have faith in what you are telling. Once that faith is lost, then the user may not use your services. The bigger loss may be that the user may not use public transportation system on a whole now that then he or she may be shifting back to the private vehicle. So reliability of information is very crucial, that allows ITS to improve efficiency of a transportation network.

(Refer Slide Time: 13:29)



Vehicular Communication

Types of vehicular communication

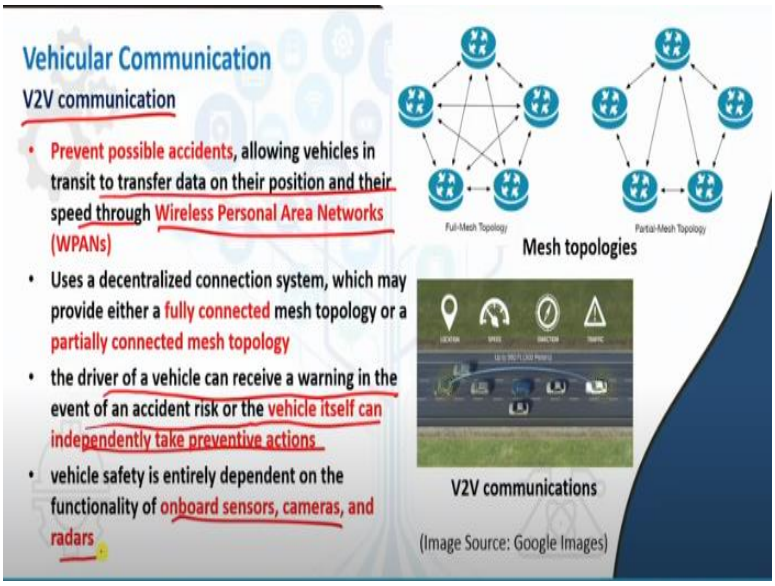
- Vehicle-To-Vehicle (V2V)→ communication between two vehicles
- Vehicle-To-Infrastructure (V2I)→ communication between vehicles and infrastructure
- Vehicle-To-Everything (V2X)→ vehicle to any entity that can influence it

(Image Source: Google Images)

Remote Cloud Control and Management Center
Satellite Positioning System
Cellular or Optical Communication
Cellular Base Station
Visible Light Communication
Roadside Unit WAVE/IEEE 802.11p
Cellular V2I Communication
V2R Communication
V2V Direct Communication
V2X Communication

Now what are the different types of vehicle communications? We had already told you a little bit about it. Vehicle to vehicle communications is communicating to this vehicle saying that I am ‘n’seconds behind you, for example. So maybe it is a time headway. So that is vehicle to vehicle information. Vehicle to infrastructure is communication between vehicles to the infrastructure. So for example, this vehicle may be communicating a base station or a mechanic or to the TMC saying that my vehicle has broken down and this is my location based on cell phone towers and I need some assistance at this location. So this is communication from the vehicle to the infrastructure, in this case to the TMC. Then vehicle to everything, that means vehicle to any entity that it can influence. So it can send vehicle to vehicle, vehicle to infrastructure, vehicle to pedestrian who is walking next to the vehicles or to anybody who can influence the vehicle. The ultimate aim is the vehicle should be able to communicate to anybody who can influence it. We are still at a nascent stage in this regard. However, there is a lot of vehicle to infrastructure communication that is happening, vehicle to vehicle communication.

(Refer Slide Time: 15:29)



Vehicular Communication

V2V communication

- Prevent possible accidents, allowing vehicles in transit to transfer data on their position and their speed through Wireless Personal Area Networks (WPANs)
- Uses a decentralized connection system, which may provide either a fully connected mesh topology or a partially connected mesh topology
- the driver of a vehicle can receive a warning in the event of an accident risk or the vehicle itself can independently take preventive actions
- vehicle safety is entirely dependent on the functionality of onboard sensors, cameras, and radars

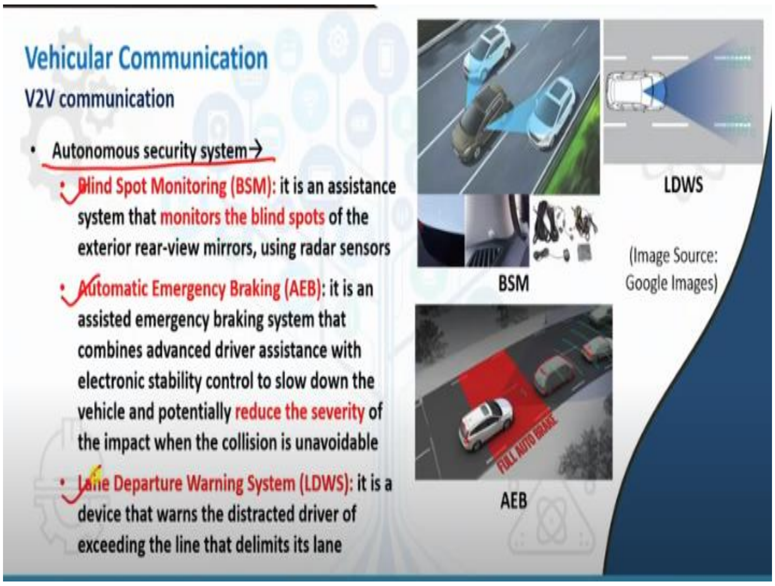
Mesh topologies

V2V communications

(Image Source: Google Images)

Here we will tell you how it is happening. So when we talk about vehicle to vehicle communication it is usually happening through wireless networks, allowing vehicles in transit to transfer data on their position and their speed. Position and speed, these are the two data points that vehicles communicate. What happens is that, for example a driver of a vehicle can receive a warning in the event of an accident risk, or the vehicle itself can independently take preventive actions. You may see that modern cars are coming with automated brake system, ABS. What it says is that if your vehicle gets too close to the vehicle in front of you, suddenly breaks are applied. Your vehicle automatically acts. Even you do not have to put your foot on the brake, but the vehicle automatically brakes. Or there are onboard sensors, cameras and radars that tells you that a vehicle is too close on your right hand side. So do not change your lane right now. Maybe you're a vehicle is in your blind spot. So do not change your lane right now. So these kinds of sensors are already in use in many of the vehicles which allows you to have a safer driving experience. They improve your safety or of the passengers, and when safety is improved in turn the efficiency of the entire network also improves.

(Refer Slide Time: 17:17)



Vehicular Communication

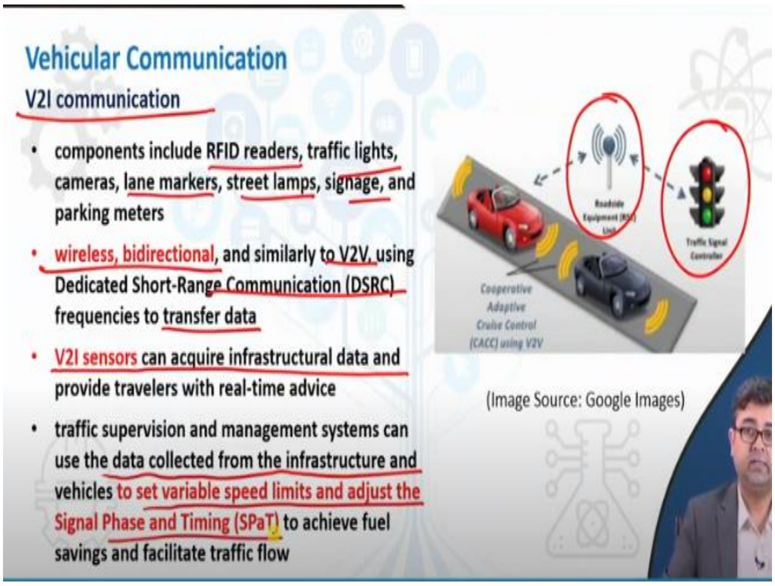
V2V communication

- Autonomous security system→
  - Blind Spot Monitoring (BSM): it is an assistance system that monitors the blind spots of the exterior rear-view mirrors, using radar sensors
  - Automatic Emergency Braking (AEB): it is an assisted emergency braking system that combines advanced driver assistance with electronic stability control to slow down the vehicle and potentially reduce the severity of the impact when the collision is unavoidable
  - Lane Departure Warning System (LDWS): it is a device that warns the distracted driver of exceeding the line that delimits its lane

(Image Source: Google Images)

Vehicle to vehicle communication. The other thing is autonomous security systems. Like we have already said, blind spot monitoring, automatic braking systems, automatic emergency brakes, lane departure warning systems. So we have already discussed about these three things and these are already some things that are available at some of the high end models of the existing vehicles that are available in India as well.

(Refer Slide Time: 17:47)



Vehicular Communication

V2I communication

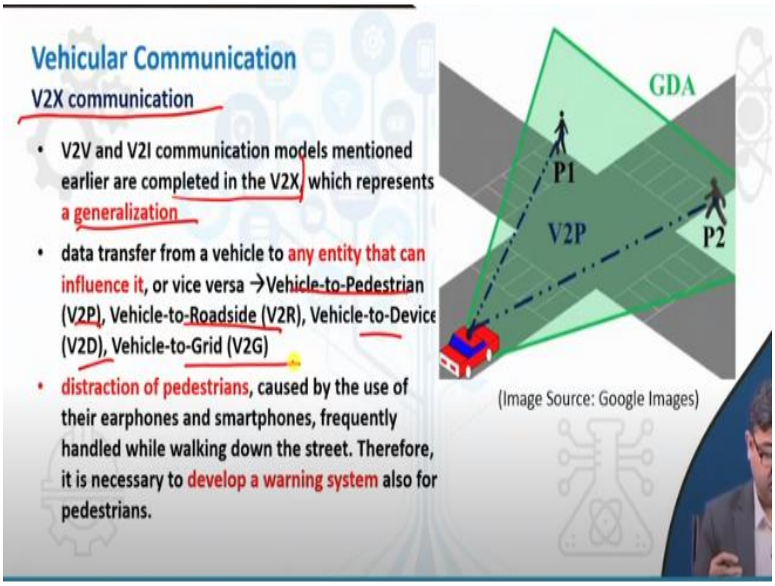
- components include RFID readers, traffic lights, cameras, lane markers, street lamps, signage, and parking meters
- wireless, bidirectional, and similarly to V2V, using Dedicated Short-Range Communication (DSRC) frequencies to transfer data
- V2I sensors can acquire infrastructural data and provide travelers with real-time advice
- traffic supervision and management systems can use the data collected from the infrastructure and vehicles to set variable speed limits and adjust the Signal Phase and Timing (SPaT) to achieve fuel savings and facilitate traffic flow

Roadside Equipment (RSE) Unit
Traffic Signal Controller
Cooperative Adaptive Cruise Control (CACC) using V2V

(Image Source: Google Images)

Now vehicle to infrastructure communication, V2I components include RFID readers, traffic lights, cameras, lane markers, street lamps, signage and parking meters. So this is something interesting that is happening in some of the urban areas as well. You can have on-street parking that can be monitored using vehicle to infrastructure communication. On-street parking ones are created where a vehicle enters into that zone, it is automatically sensed and then there is a mechanism by which it gets your meter. It starts automatically for parking price. And then based on how long you are parked, you can automatically get billed for your parking. So this is something that is being used in the case of some of the cities that are looking at efficient parking management systems for their city. Wireless bidirectional and similarly vehicle to vehicle using Dedicated Short Range Communication (DSRC) frequencies to transfer data. DSRC is something that is necessary in order to transfer data. V2I sensors can acquire infrastructural data and provide travelers with real time advice. Vehicle to infrastructure the most prominent ones are your traffic signal that we have already discussed about. We can significantly reduce the presence of traffic police at a signalized intersection by using such ICT devices. Traffic supervision and management systems can use the data collected from the infrastructure and vehicles to set variable speed limits and adjust signal phase and timing, to achieve fuel saving. Based on the arrival patterns that you see through your CCTV cameras at the TMCs, you can then set your traffic signals more efficiently. That is what we have already mentioned to you about converting the feed that you are getting from your CCTV cameras into information, and then processing that information and using it to forecast your travel time, speed or delay on the corridor. So that is something that is crucial and ICT can help improve the efficiency of a signalized intersection vis-à-vis reduce the delays at such intersections.

(Refer Slide Time: 20:50)



Vehicular Communication

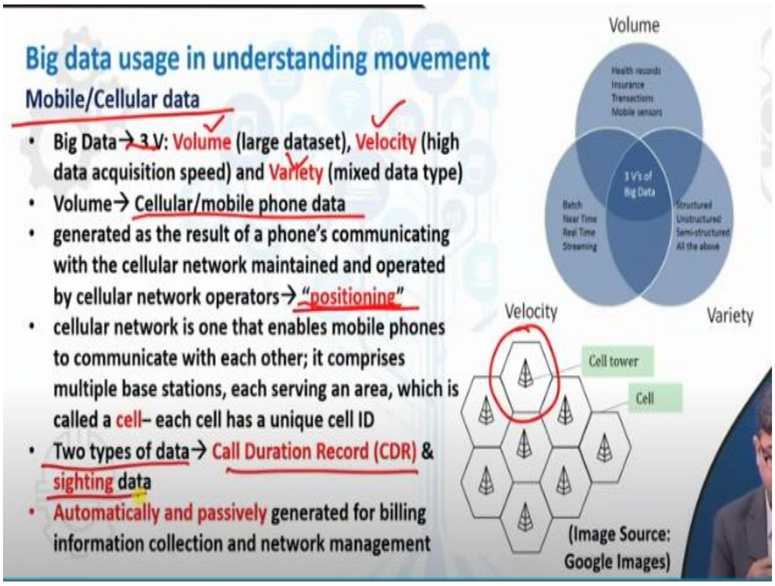
V2X communication

- V2V and V2I communication models mentioned earlier are completed in the V2X which represents a generalization
- data transfer from a vehicle to any entity that can influence it, or vice versa →Vehicle-to-Pedestrian (V2P), Vehicle-to-Roadside (V2R), Vehicle-to-Device (V2D), Vehicle-to-Grid (V2G)
- distraction of pedestrians, caused by the use of their earphones and smartphones, frequently handled while walking down the street. Therefore, it is necessary to develop a warning system also for pedestrians.

(Image Source: Google Images)

Vehicle to X, vehicle to everybody, vehicle to vehicle and vehicle to infrastructure communication models, mentioned earlier, are completed in the V2X, which represents a generalization. Data transfer from a vehicle to any entity that can influence it or vice versa vehicle to pedestrian, vehicle to roadside, vehicle to device, vehicle to grid. So this is something that is being still worked upon. This can alert a pedestrian for example who is trying to cross a street at an unsignalized intersection. And the vehicle can now communicate with the person who is trying to cross the street and maybe send a message to his or her mobile phone saying that the vehicle is approaching, please be alert. This is especially for the current situation because at unsignalized intersection we usually nowadays have our headphones on. We are not looking at the traffic and in such situations these things might be helpful. But these are still in a testing phase and this is something if it becomes operational or viable can help in the safety of the transportation network as a whole. At various locations in our network, we would need to develop such warning systems for pedestrians.

(Refer Slide Time: 22:26)



Big data usage in understanding movement

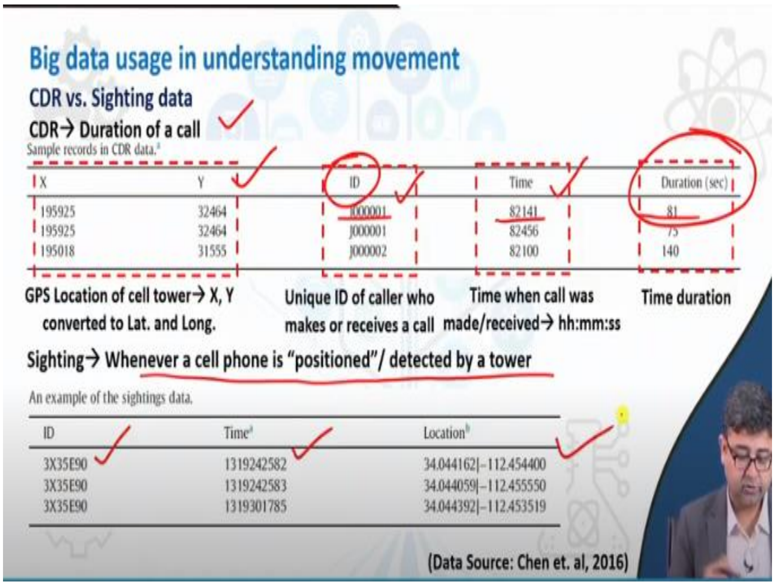
Mobile/Cellular data

- Big Data→3 V: Volume (large dataset), Velocity (high data acquisition speed) and Variety (mixed data type)
- Volume→ Cellular/mobile phone data
- generated as the result of a phone's communicating with the cellular network maintained and operated by cellular network operators→ "positioning"
- cellular network is one that enables mobile phones to communicate with each other; it comprises multiple base stations, each serving an area, which is called a cell– each cell has a unique cell ID
- Two types of data→ Call Duration Record (CDR) & sighting data
- Automatically and passively generated for billing information collection and network management

(Image Source: Google Images)

Now when we talk about big data, let us give you an understanding of what we usually are trying to say. When we talk about big data, we are usually talking about mobile or cellular phone data. There are three V's associated with it. One is volume, large data set. Then velocity, high speed of data acquisition. The speed at which the data is generated and acquired is very high. And variety is a mixed data type. There is a lot of different types of data that is coming in. That is why we say big data first and foremost what that is needed is data cleaning. You may be for example, you may be getting hundred pieces of information, but maybe all you need is 20 pieces of information. So you have to screen the data and weed out those other 80 pieces of information and then only look at the 20 pieces that you need. Maybe you are getting one lakh data points per minute and if you if you talk about only 24 hour data, you may have a huge data set. So even just for working with 24 hours data and to clean up that data it will take you a lot of computing. It will take you computing resources and computing knowledge in your TMC. Cellphone data generated as a result of phone's communication. One thing it gives you is the positioning of the various cell phones. We have already talked about how cell phone towers and can be used in triangulation and trilateration to find out where an object is. From your cell phone data, two types of data can be collected for transportation purposes. Call duration record, CDR or sighting data.

(Refer Slide Time: 24:41)



Big data usage in understanding movement

CDR vs. Sighting data

CDR→ Duration of a call

Sample records in CDR data.[a]

X	Y	ID	Time	Duration (sec)
195925	32464	J000001	82141	81
195925	32464	J000001	82456	75
195018	31555	J000002	82100	140

GPS Location of cell tower→ X, Y converted to Lat. and Long.

Unique ID of caller who makes or receives a call

Time when call was made/received→ hh:mm:ss

Time duration

Sighting→ Whenever a cell phone is "positioned"/ detected by a tower

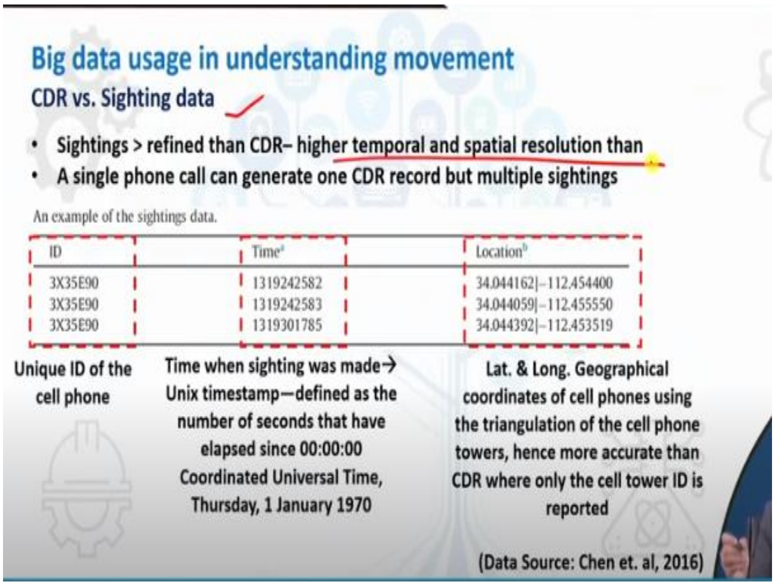
An example of the sightings data.

ID	Time[a]	Location[b]
3X35E90	1319242582	34.044162|-112.454400
3X35E90	1319242583	34.044059|-112.455550
3X35E90	1319301785	34.044392|-112.453519

(Data Source: Chen et. al, 2016)

CDR or call duration records usually gives you a latitude longitude of your particular Mac ID. At what time that Mac ID was located at that location, and for how long your call went on for. So through this information, one could understand is that nobody can trace it back to you, because all the personal identification features have been anonymized. We only know there is a certain phone with a Mac ID that started a phone call at this time, and that phone call lasted for 81 seconds. So we can come to know about his or her call duration record. Sightings is whenever a cell phone is positioned and detected by a tower. This is just sightings. You do not have to make a call in order to find out where you are. The cell phone towers can just detect if you have your GPS signal on your phone. The cell phone towers can just detect where you are based on triangulation.

(Refer Slide Time: 25:59)



Big data usage in understanding movement

CDR vs. Sighting data

- Sightings > refined than CDR– higher temporal and spatial resolution than
- A single phone call can generate one CDR record but multiple sightings

An example of the sightings data.

ID		Time[a]		Location[b]
3X35E90		1319242582		34.044162|−112.454400
3X35E90		1319242583		34.044059|−112.455550
3X35E90		1319301785		34.044392|−112.453519

Unique ID of the cell phone

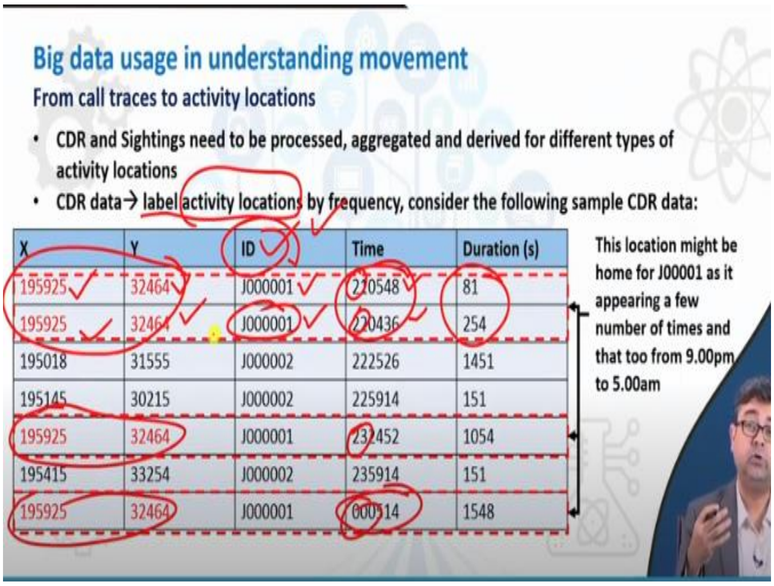
Time when sighting was made→ Unix timestamp—defined as the number of seconds that have elapsed since 00:00:00 Coordinated Universal Time, Thursday, 1 January 1970

Lat. & Long. Geographical coordinates of cell phones using the triangulation of the cell phone towers, hence more accurate than CDR where only the cell tower ID is reported

(Data Source: Chen et. al, 2016)

Sightings are a little bit more refined than CDR and can do a lot of temporal and spatial resolution. CDR usually has to you have to make a phone call in order for you to locate the person. In this case, you do not have make a phone call, you just have to have your GPS signal on in order to track.

(Refer Slide Time: 26:23)



Big data usage in understanding movement

From call traces to activity locations

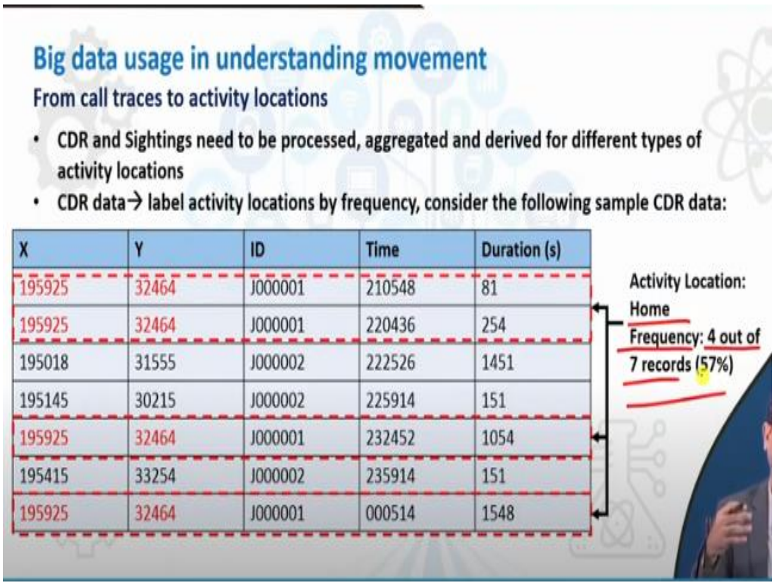
- CDR and Sightings need to be processed, aggregated and derived for different types of activity locations
- CDR data→ label activity locations by frequency, consider the following sample CDR data:

X	Y	ID	Time	Duration (s)
195925	32464	J000001	210548	81
195925	32464	J000001	220436	254
195018	31555	J000002	222526	1451
195145	30215	J000002	225914	151
195925	32464	J000001	232452	1054
195415	33254	J000002	235914	151
195925	32464	J000001	000514	1548

This location might be home for J00001 as it appearing a few number of times and that too from 9.00pm to 5.00am

CDR and sightings data can be then converted into further information, which the traffic or transportation engineers are usually interested in the origin destination data. We are always interested in where a person is located and where he or she is traveling to. Our primary concern is whether he or she is located in residential zone, a commercial zone, or a retail zone and how long does it take him or her to travel between different zones. That is what essentially we are trying to find out. So from the CDR data, you can then find out different activity location in your city as well. For example, ID J0001 is making different calls during these times of the day, and from these same x, y coordinate locations right. So usually, what you can infer is that these are night times, evening 9pm, 10pm, 11pm, night, midnight, later midnight. So usually what that would should tell you is this location should be his or her home location. Usually a person would not be in his or her office at such later time in the day. Usually he or she should be at home. Now you have this IDs data for about a year or so. Once you have it for a year and then you can generalize and see yes, this may be his or her home location or then you can have multiple of such IDs, and then figure out where their home locations could be. Similarly, if you have data during the day times and then you can figure out during the day times J0001 is making different calls from a different x, y location. Maybe that location during the daytime could be his work location. As a transportation engineer or a traffic planner, now you have information about your origin and destination of your different users based on just cell phone data or calling data records. The volume of data that we are getting is improving vastly. Otherwise in order to get origin destination information, we used to conduct questionnaire surveys and the sample size would inevitably be always low. And will not give very good information. But now through cell phones, we are able to get larger volumes of data that can be validated from our questionnaire survey.

(Refer Slide Time: 29:52)



Big data usage in understanding movement

From call traces to activity locations

- CDR and Sightings need to be processed, aggregated and derived for different types of activity locations
- CDR data→ label activity locations by frequency, consider the following sample CDR data:

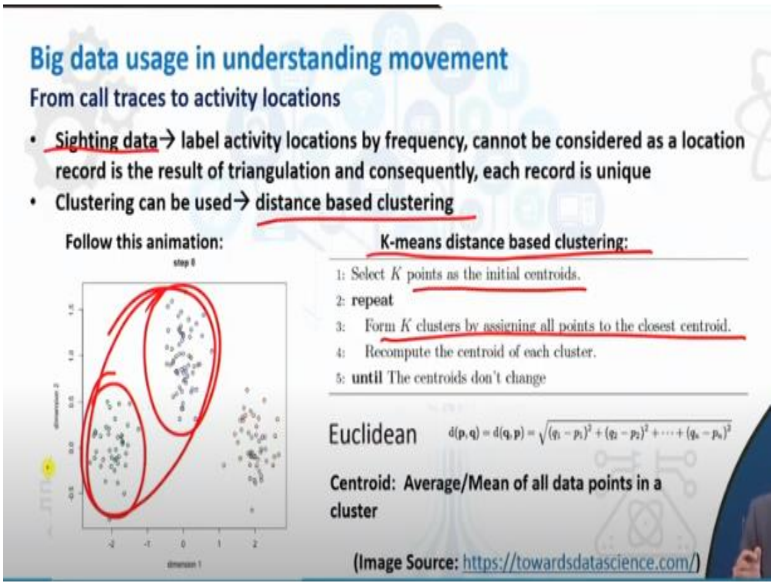
X	Y	ID	Time	Duration (s)
195925	32464	J000001	210548	81
195925	32464	J000001	220436	254
195018	31555	J000002	222526	1451
195145	30215	J000002	225914	151
195925	32464	J000001	232452	1054
195415	33254	J000002	235914	151
195925	32464	J000001	000514	1548

Activity Location: Home

Frequency: 4 out of 7 records (57%)

So like I said the activity location home can be inferred because four out of seven record show that he or she is making these calls at these different times. So with certain amount of confidence we can say it is his or her home location.

(Refer Slide Time: 30:10)



Big data usage in understanding movement

From call traces to activity locations

- Sighting data→ label activity locations by frequency, cannot be considered as a location record is the result of triangulation and consequently, each record is unique
- Clustering can be used→ distance based clustering

Follow this animation:

K-means distance based clustering:

```
1: Select $K$ points as the initial centroids.
2: repeat
3:    Form $K$ clusters by assigning all points to the closest centroid.
4:    Recompute the centroid of each cluster.
5: until The centroids don't change
```

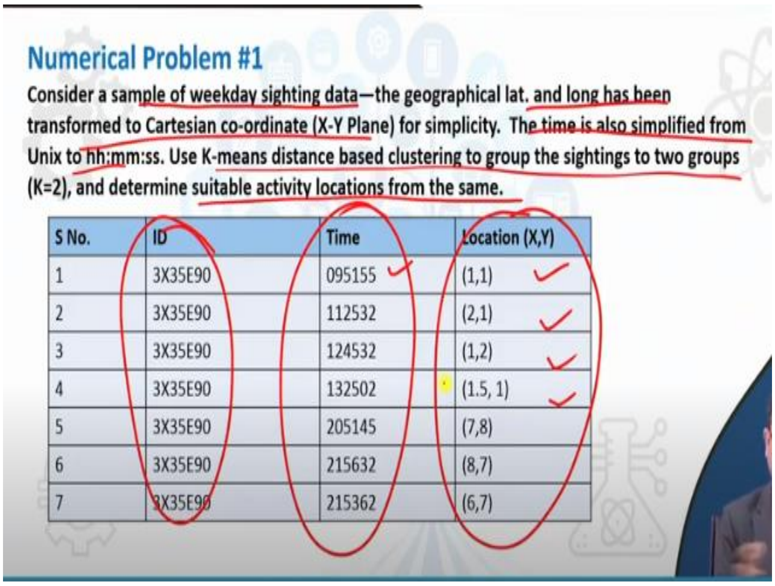
Euclidean $d(\mathbf{p},\mathbf{q}) = d(\mathbf{q},\mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}$

Centroid: Average/Mean of all data points in a cluster

(Image Source: https://towardsdatascience.com/)

Now how can sighting data be used? Sighting data can be used to determine the activity locations by something called distance based clustering. You are encouraged to take more and more statistical classes, especially if you want to work in ITS because this involves a lot of big data analytics. Statistics becomes more and more crucial in your ability to be able to work with such large volumes of data. One of the method is K-means based distance clustering. What it essentially tells you is that if there are K points with initial centroids and if these centroids do not change after forming newer clusters, then you can combine those points into one cluster location. There may be different points, but can you say that this is one cluster and this is another cluster. Or do you want to say that this entire thing is a cluster. How do you determine what is a cluster. So this K-means distance based clustering is a very simple method by which you can determine different clusters. And once you know different clusters, then you can say that this is a home cluster versus this is an office cluster versus this is a shopping cluster and that gives you an understanding of origin and destination in your city.

(Refer Slide Time: 31:48)



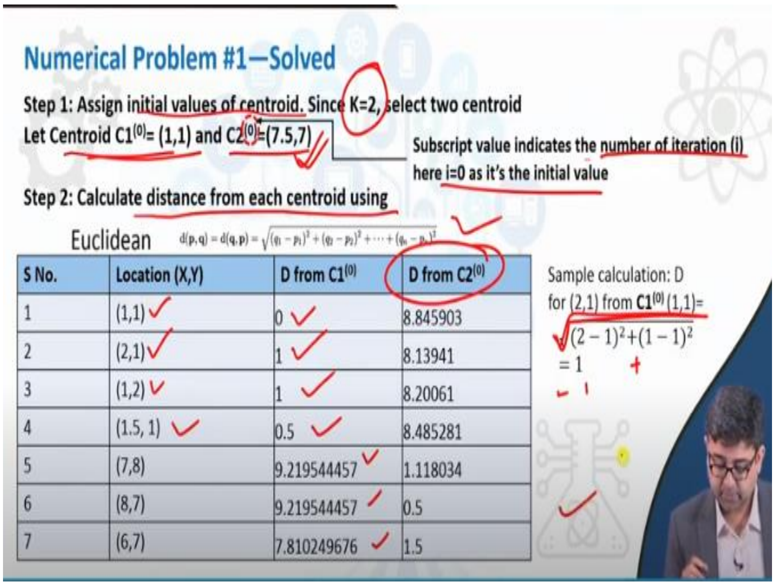
Numerical Problem #1

Consider a sample of weekday sighting data—the geographical lat. and long has been transformed to Cartesian co-ordinate (X-Y Plane) for simplicity. The time is also simplified from Unix to hh:mm:ss. Use K-means distance based clustering to group the sightings to two groups (K=2), and determine suitable activity locations from the same.

S No.	ID	Time	Location (X,Y)
1	3X35E90	095155	(1,1)
2	3X35E90	112532	(2,1)
3	3X35E90	124532	(1,2)
4	3X35E90	132502	(1.5, 1)
5	3X35E90	205145	(7,8)
6	3X35E90	215632	(8,7)
7	3X35E90	215362	(6,7)

So let us see an example that can allow you to simplistically understand how K-means clustering works. Consider a sample of a weekday sighting data. The geographical latitude and longitude has been transformed into Cartesian X, Y coordinates. So these are mostly X, Y coordinates for simplicity. The time is also simplified from Unix to hh:mm:ss. Use K-means based clustering to group the sightings into two groups and determine suitable activity locations for the same. You have one ID and have been tracking that ID and seeing that it has been at these different locations during these different times. So now, you want to, based on this location data and time data, see if there is any activity patterns that you can determine. If you can determine that they are at a home location or an office location and where spatially they could be. Spatially you already know that they are in different latitude and longitude. So can you determine conclusively based on this information, that which is the home location and which is the work location?

(Refer Slide Time: 33:11)



Numerical Problem #1—Solved

Step 1: Assign initial values of centroid. Since K=2, select two centroid

Let Centroid $C1^{(0)}$= (1,1) and $C2^{(0)}$=(7.5,7)

Subscript value indicates the number of iteration (i) here i=0 as it's the initial value

Step 2: Calculate distance from each centroid using

Euclidean $d(\mathbf{p},\mathbf{q}) = d(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2+(q_2-p_2)^2+\cdots+(q_n-p_n)^2}$

S No.	Location (X,Y)	D from $C1^{(0)}$	D from $C2^{(0)}$
1	(1,1)	0	8.845903
2	(2,1)	1	8.13941
3	(1,2)	1	8.20061
4	(1.5, 1)	0.5	8.485281
5	(7,8)	9.219544457	1.118034
6	(8,7)	9.219544457	0.5
7	(6,7)	7.810249676	1.5

Sample calculation: D for (2,1) from $C1^{(0)}$ (1,1)=

$\sqrt{(2-1)^2+(1-1)^2}$

= 1

So what essentially what first you do is you assign initial values of centroid. Just pick based on what is available here from these different Cartesian coordinates X, Y plane or in other words you just have initial values. Since K is only 2, you have said only two groups. So let us assign two centroids. Let the first centroid be (1, 1). And the second centroid be (7.5, 7). And the superscript here tells the number of iterations in this case i = 0 because we are making the first iteration. So this is the 0th iteration meaning the base value. Now calculate the distance from each centroid using the Euclidean equation, as below.

Euclidean $d(\mathbf{p},\mathbf{q}) = d(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2+(q_2-p_2)^2+\cdots+(q_n-p_n)^2}$

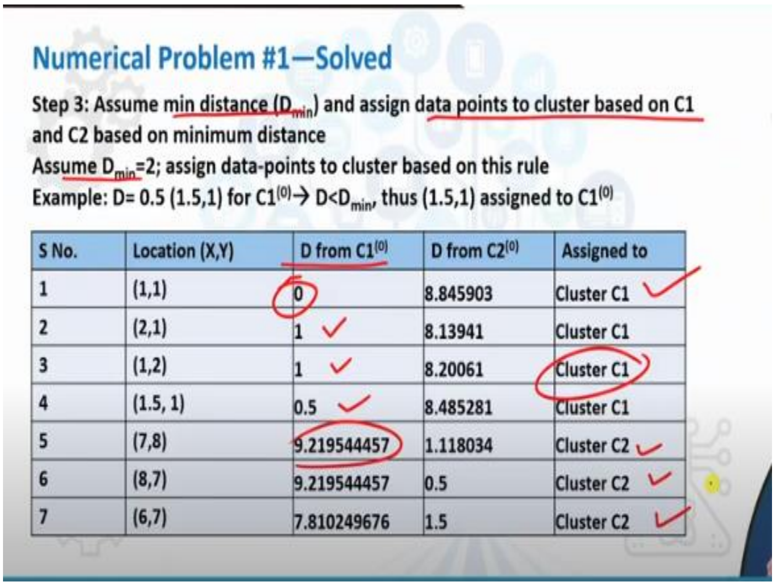
So the first centroid was chosen as (1, 1) thus the distance of this point from that centroid itself, would be obviously zero. For the point (2,1) all you have to do is you have to subtract 2 from 1, square it and plus 1 minus 1 square it and then do a square root of the entire thing which will give you the Euclidean distance of 1.

Sample calculation: D for (2,1) from **$C1^{(0)}$** (1,1)=

$$\sqrt{(2-1)^2+(1-1)^2} = 1$$

Similarly, you do it for all of the other values. And calculate the distance from your second centroid as well. So your distance the second centroid, remember is (7.5, 7). Because you are already told that to group the sightings into two clusters. You have initially assumed two centroids based on the x, y plane and from both of the centroids you are now calculating the distance in the second step.

(Refer Slide Time: 35:04)



Numerical Problem #1—Solved

Step 3: Assume min distance (D_{min}) and assign data points to cluster based on C1 and C2 based on minimum distance

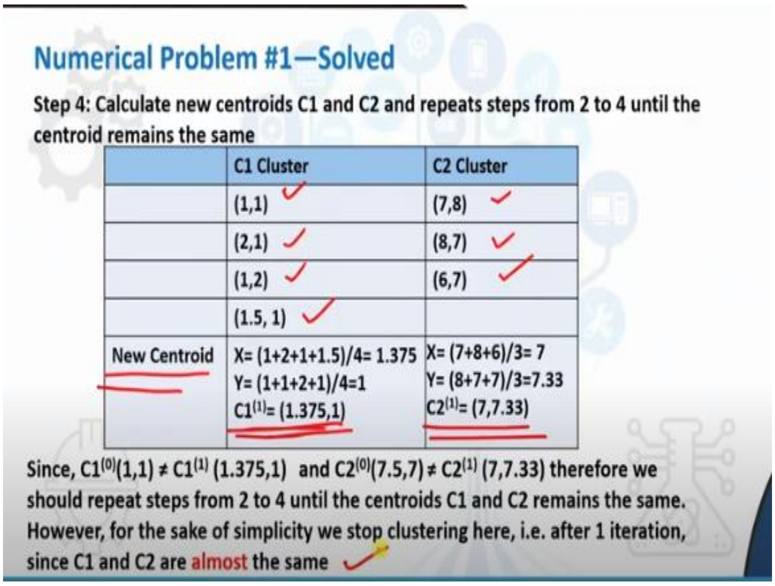
Assume D_{min}=2; assign data-points to cluster based on this rule

Example: D= 0.5 (1.5,1) for $C1^{(0)}$ → $D<D_{min}$, thus (1.5,1) assigned to $C1^{(0)}$

S No.	Location (X,Y)	D from $C1^{(0)}$	D from $C2^{(0)}$	Assigned to
1	(1,1)	0	8.845903	Cluster C1
2	(2,1)	1	8.13941	Cluster C1
3	(1,2)	1	8.20061	Cluster C1
4	(1.5, 1)	0.5	8.485281	Cluster C1
5	(7,8)	9.219544457	1.118034	Cluster C2
6	(8,7)	9.219544457	0.5	Cluster C2
7	(6,7)	7.810249676	1.5	Cluster C2

Once you have calculated the distance, you assume a minimum distance and assign data points to cluster C1 or C2 based on minimum distance. Say you assume a minimum distance of 2. So this is some kind of criteria or threshold that you put. Now if anything is less than 2, then you say that it belongs to cluster C1 and if the distance is greater than 2, then you say that it belongs to cluster C2. This is an iterative process. The iterations will stop once the centroid positions do not change. Once the centroid positions do not change between one iteration and other iteration, then you can say that yes these two are the centroids and all the points around it belong to this cluster and all the points along around it belong to the other cluster. So this is just the first iteration or the first step that we are telling you. If you assume, in this case the threshold of D min to be 2, then you would say that, the distance from C1 is 0. So since 0 is less than 2, this should belong to cluster C1. Similarly, all of this are less than 2. So all of these are belonging to cluster C1. Whereas, all of this is greater than 2. So this should be belonging to C2. So that is that is how you initially assign the points to these different clusters.

(Refer Slide Time: 36:35)



Numerical Problem #1—Solved

Step 4: Calculate new centroids C1 and C2 and repeats steps from 2 to 4 until the centroid remains the same

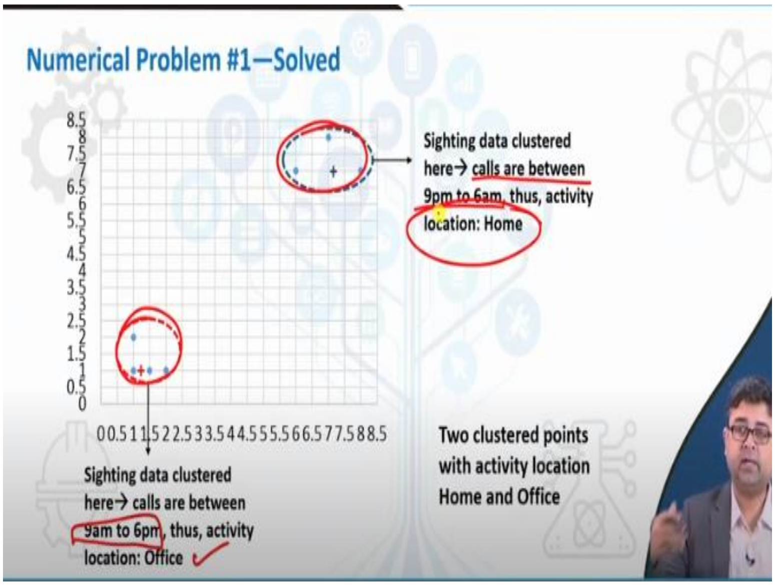
	C1 Cluster	C2 Cluster
	(1,1) ✓	(7,8) ✓
	(2,1) ✓	(8,7) ✓
	(1,2) ✓	(6,7) ✓
	(1.5, 1) ✓	

New Centroid	X= (1+2+1+1.5)/4= 1.375
	Y= (1+1+2+1)/4=1
	$C1^{(1)}$= (1.375,1)
		X= (7+8+6)/3= 7
		Y= (8+7+7)/3=7.33
		$C2^{(1)}$= (7,7.33)

Since, $C1^{(0)}(1,1) \neq C1^{(1)}$ (1.375,1) and $C2^{(0)}(7.5,7) \neq C2^{(1)}$ (7,7.33) therefore we should repeat steps from 2 to 4 until the centroids C1 and C2 remains the same. However, for the sake of simplicity we stop clustering here, i.e. after 1 iteration, since C1 and C2 are almost the same ✓

Now that you have assigned these points to different clusters, C1 cluster has these four points, C2 cluster has these three points, next calculate the new centroid for each of those two clusters. You have you had initially assumed two centroids remember. Now you have, two new centroids and if these two new centroids had to be equal to the initial centroid that you had assumed, then your iteration would stop and you would say that C1 and C2 are two distinct clusters. In this case, although they are not for simplicity for this exercise, we are saying they are almost the same although they are not really same. So you have to repeat this for a next iteration. But for simplicity, let us say that they are both same, because 1.375 and 1 are very close to each other. So let us for this exercise, say that C1 and C2 are almost the same. And we say that the iterations stop here.

(Refer Slide Time: 37:40)



Numerical Problem #1—Solved

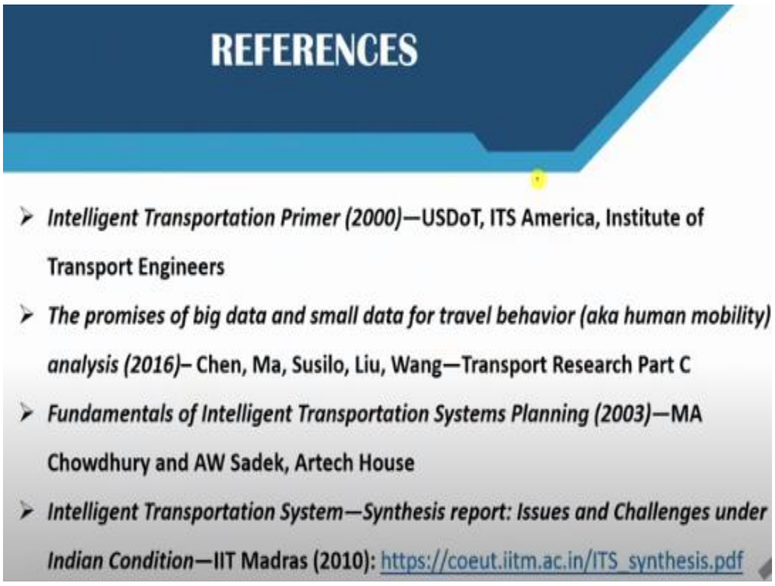
Sighting data clustered here→ calls are between 9pm to 6am, thus, activity location: Home

Sighting data clustered here→ calls are between 9am to 6pm, thus, activity location: Office

Two clustered points with activity location Home and Office

And hence, since we stopped the iterations here, we find out that these two are distinct clusters. And since these two are distinct clusters, and all the calls were between 9am and 6pm in this cluster, so we can say that maybe this is an office location 9am and 6pm, because these calls in this cluster came from 9am to 6pm. And in this cluster, all the calls came between 9pm and 6am. And hence, this could be the home cluster. So you see, now you have based on just cell phone data records of that one ID you have been able to decide that ID's, possible office location and possible home location. And then you can further use mobile phone sighting data to find out the speed in which maybe these this ID is traveling between these two. So you get an idea or you get a feel for how cell phone data can be used in the transportation arena to find out not only activity. In this case activity locations, but in future also we will tell you how this can be used in calculating travel times and travel speeds.

(Refer Slide Time: 39:00)

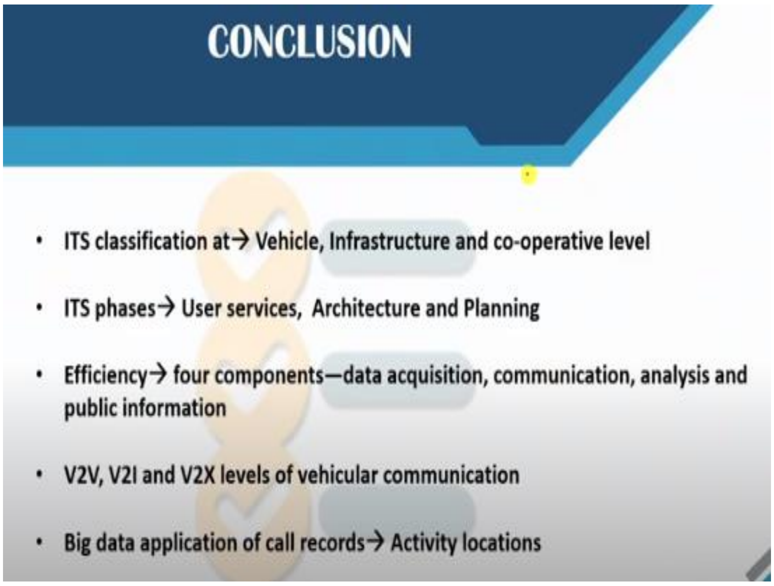


REFERENCES

- ➢ *Intelligent Transportation Primer (2000)*—USDoT, ITS America, Institute of Transport Engineers
- ➢ *The promises of big data and small data for travel behavior (aka human mobility) analysis (2016)*– Chen, Ma, Susilo, Liu, Wang—Transport Research Part C
- ➢ *Fundamentals of Intelligent Transportation Systems Planning (2003)*—MA Chowdhury and AW Sadek, Artech House
- ➢ *Intelligent Transportation System—Synthesis report: Issues and Challenges under Indian Condition*—IIT Madras (2010): https://coeut.iitm.ac.in/ITS_synthesis.pdf

Hopefully, you are getting a feel for all of this. All the references are given here. We understand that many of you may not have an IT or electrical or electronics background. But like we said, this is a new arena in transportation engineering and traffic planning. So you should at least get an understanding of how all of this works. So please, do go through the references.

(Refer Slide Time: 39:24)



CONCLUSION

- ITS classification at→ Vehicle, Infrastructure and co-operative level
- ITS phases→ User services, Architecture and Planning
- Efficiency→ four components—data acquisition, communication, analysis and public information
- V2V, V2I and V2X levels of vehicular communication
- Big data application of call records→ Activity locations

And just in conclusion in this slide, in this lecture, what we have looked at the different classifications at vehicle infrastructure and cooperative levels. The different ITS phases. What are the user services and what are architecture and how do we plan. And then we have looked at the four different components of efficiencies, and how big data can be applied in finding out activity locations. Thank you.