

**Information security - IV**  
**Prof. M J Shankar Raman**  
**Department of Computer Science and Engineering**  
**Indian Institute of Technology, Madras**

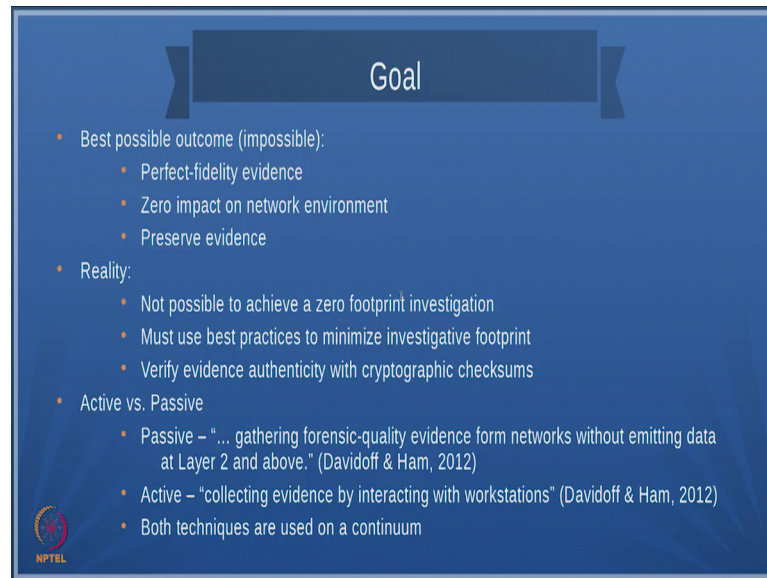
**Lecture – 35**  
**Technical Fundamentals for Evidence Acquisition – 1**

Hi there, a welcome to this third module on network forensics. Until now you would have seen two modules main modules one on the introduction and what we are going to look in this course and the second on operating system security. With respect to operating system security you would have learnt how to do penetration testing, how to do scanning, and essentially you will have looked at the logs that are generated by the operating system such as this logs which can be used for securing the systems or finding out who has entered the system, at what, time etcetera.

In this module we are going to look at the networking aspects of security and network forensics where we collect data from various places in the network. Once we collect data from the various parts in the network we go ahead and do an analysis to find out what has really happened then an attack has taken place or whether someone is trying to probe into your network and then see what are the resources that you are having and how we can attack those resources.

So, before we go into network forensics we need to have some technical fundamentals and we also have to establish a goal for the forthcoming modules. So, one of the goal best possible outcome that can come out of network forensics is that you have a perfect fidelity evidence which essentially means you are able to pinpoint what has happened.

(Refer Slide Time: 02:08)



The slide is titled "Goal" and is set against a dark blue background. It contains a list of bullet points. The first main bullet point is "Best possible outcome (impossible):" followed by three sub-bullets: "Perfect-fidelity evidence", "Zero impact on network environment", and "Preserve evidence". The second main bullet point is "Reality:" followed by three sub-bullets: "Not possible to achieve a zero footprint investigation", "Must use best practices to minimize investigative footprint", and "Verify evidence authenticity with cryptographic checksums". The third main bullet point is "Active vs. Passive" followed by three sub-bullets: "Passive – "... gathering forensic-quality evidence form networks without emitting data at Layer 2 and above." (Davidoff & Ham, 2012)", "Active – "collecting evidence by interacting with workstations" (Davidoff & Ham, 2012)", and "Both techniques are used on a continuum". In the bottom left corner of the slide, there is a small circular logo with the text "NPTEL" below it.

Now, this is really hard. I mean it is its literally impossible to find out a perfect fidelity evidence. So, what we have to do is you can use correlations and try to establish that something wrong has happened ok. The other goal that you want to do is whenever you are doing network forensics or when you are observing the network. You might either observe a live network, but if you observe a live network then you should have 0 impact on the network environment it is like I am trying to collect some evidence there by distorting the evidence or destroying the evidence that should not be the case when we try to do network forensics and the third aspect of what we are supposed to do is to preserve the evidence.

Now, you go do all the hard work of trying to find out what are all the evidences then you collect them, you do all the hard work of correlating everything and then trying to establish something and finally, if you do not have a method of preserving the evidence then everything is lost. So, these are the 3 possible ways by which you can you, I mean there are 3 possible ways in which you are to do the excise carefully.

In reality what is going to happen? One it is really very difficult for us to achieve a 0 footprint investigation. That means, when you do investigation say let us take one police officer trying to find out a murder case and he goes to the murder scene even though he is going to search for evidence of the murderer or the supposed murderer he is also going to leave his marks in the place of crime, So, similar to that when we are trying to get

network evidence we are also going to leave our footprint on the network investigation, but what makes you a really good network forensic specialist is that the amount of evidence that you will leave should be as small as possible.

Now, this is also necessary because if someone is trying to look at your network or trying to monitor your network, you should not be visible to the other person. So, even that is very important. Second one that you should do is that you should come up with the best possible procedures whereby you actually limit the investigative footprint. And the third one is that checksums are very useful when you want to authenticate an evidence, I mean hashing checksum digital signatures, all these things even after you collect any evidence you have to digitally sign it and then store it correctly etcetera etcetera.

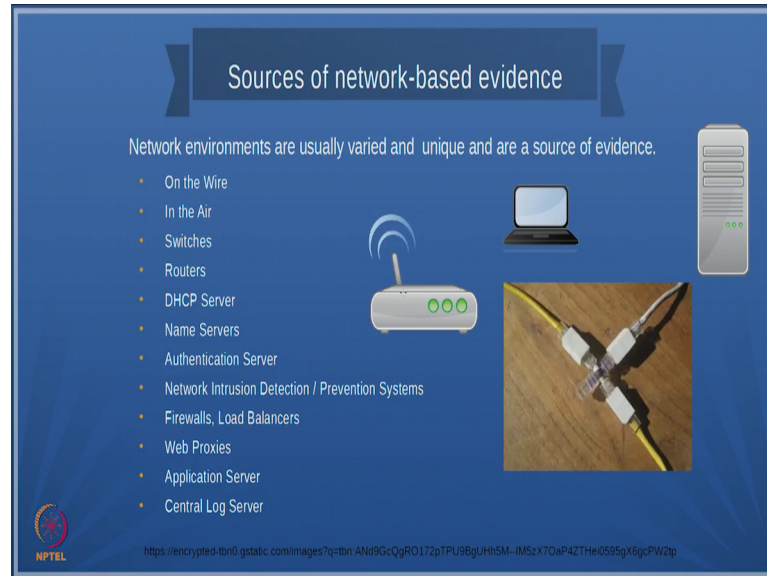
There are two types of evidence collection techniques one is known as the passive technique as David often Hamm given in their book its one is the passive technique and other is the active technique. The passive techniques gathers forensic quality evidence from networks without emitting data layer two or above, and the active method collects evidences by interacting with the workstations. And when you are doing forensic analysis you really have to take these two aspects together. I mean I mean you will be collecting live data you will be collecting passive data if you know that someone is trying to attack your network then you will be collecting the live data. If you, if something incident has already happened then you will be collecting a passive data from the loss this apart.

Let us now, now the whole idea behind this 2 or 3 sessions is going to be a to understand what are the technologies involved in network forensics, and b the techniques that are involved, and c the tools that are involved, and d what are all the logs from which you have to we have to look at. And in during the course of this modules will also try to show some live demo, live demo in the sense how to use certain tools and then how to study those tools and then how to do some correlation and all that.

Of course, if the doing the exact correlation and all this will involve lot of techniques like machine learning and then because the amount of data that you are going to get is really huge and we will see that what is the kind of data that you are going to get. Of course, it

is all text data but then you have to do lot of scanning, you have to do filtering and all those activities it is actually computation intensive.

(Refer Slide Time: 06:48)



Let us look at the sources of network based evidence. So, if I take a network you see that it is usually varied and unique and definitely network logs as well as the what is traveling the packets here are all sources of evidences ok. So, the network environments from where we will have to collect data will include on the wire environments on the wire includes fiber optics, cables as well as Ethernet cables and in the wire environment which essentially is wireless IEEE 802 dot 11 ok, so abg etcetera.

And as far as this is the medium in which the data gets transmitted then there are those devices which actually transmit this data. One is the switches which would be very familiar with, the routers switches actually operate at layer 2, routers operate at layer 3 then these there are these DHCP servers which try to assign IP addresses. Now, remember IP at collecting the IP address is very important evidence of course, IP address can be spoofed, but then you have to overcome all these problems to establish that a particular victim or a particular attacker actually processed that IP address at that particular instant of time and if you can prove that that would be a very good evidence, if tomorrow anyone is doing some kind of a crime investigation.

Then there are these name servers which assign IP addresses to names. The most important is the authentication server because usually authentication server is the one

that provides security from enterprise point of view. So, and sometimes if an internal employee for example, he gets into your network and steals important data then these authentication servers can actually establish when this imply had logged in and all those details ok. So, authentication servers actually provide one level of security by not allowing and unauthorized the third party to enter into the network. But then the evidences from evidence from the authentication server can also be used to pinpoint if an internal attack happens within an organization.

Then there are this regular intrusion detection and prevention systems ok, they also they also provide some kind of filtering. And then they provide lot of logs, then you might have firewalls and load balancers for big organizations and web proxies I mean then you can have application server and then the central log server. Now, a few words about these web proxies, application server and the central log server.

I hope you are all very familiar with firewalls because these are the first level of defense that any organization has. These web proxies are usually what they do is they actually cache the data that sometimes gets exchanged between the server on the client, the client could be the internal network and the server could be an external network, and these web proxies actually ensure that then the data does not go directly into the internet all the http data does not go directly into the network, but go through this web proxies. And web proxies acts like a it is a kind of what should I say a door where it cannot let and you are to you make a kind of security check when you leave or enter into the store.

Then these application servers actually provide lot of logs and of course, I mean all these logs should be enabled I mean worst case is that the network administrator or the person man managing these devices. If he does not enable any of the logs and then you got it I mean there is no way you are going to identify what has happened. Sometimes what happens is some organizations use center log server where either the data that we collect from these devices and this medium is sent to a central log server for analysis one, the advantage of doing a centralized analysis is that you can provide lot of correlation ok.

For example, at time instant  $x$  something has happened, at time instant  $x + 1$  this has happened and so on. And the point is that the central log server has to be very capable and it will be implementing lot of machine learning algorithms or big data based algorithms etcetera because the data is essentially going to be huge and you need very

powerful servers to do some sort of analysis after getting all the data. So, let us take a look at each one of these.

(Refer Slide Time: 11:43)

The slide, titled "On the wire", discusses network tapping. It lists physical cabling types (copper twisted pair/coaxial and fiber-optic) and their forensic value. It also details three tap types: Vampire, surreptitious fiber, and infrastructure. A diagram illustrates a network tap implementation on a switch, showing a tap device connected to the switch's ports to capture traffic in both directions.

**On the wire**

- Physical cabling carries data over the network
- Typical network cabling:
  - Copper : twisted pair or coaxial cable
  - Fiber-optic lines
- Forensic Value:**
  - Wire tapping can provide real-time network data
  - Tap types
    - "Vampire" tap – punctures insulation and touches cables
    - Surreptitious fiber tap – bends cable and cuts sheath, exposes light signal
    - Infrastructure tap – plugs into connectors and replicates signal – (usually used in coordination with ISPs for security monitoring)

**Network Tap Implementation**

The diagram shows a central switch with two ports. A tap device is connected to both ports. Red arrows indicate data flow from the left port to the switch and then to the right port. The tap device has two output ports, one for each direction of traffic, capturing data from both.

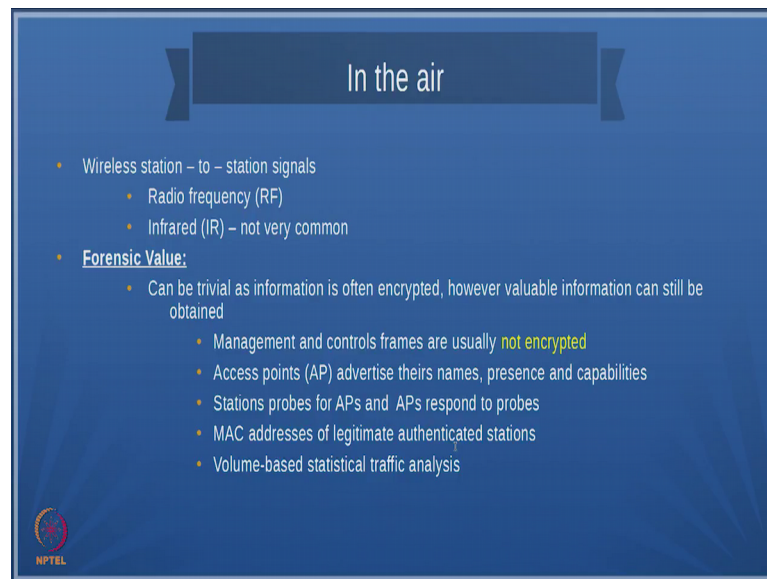
[http://www.nextgigsystems.com/net\\_optics/10\\_GigaBit\\_Fiber\\_Tap\\_files/AppL\\_C\\_SlimTap.png](http://www.nextgigsystems.com/net_optics/10_GigaBit_Fiber_Tap_files/AppL_C_SlimTap.png)

This devices and the methodology by which we are going to acquire data, ok; The first one is on the wire and it actually requires a kind of physical cabling ok. So, for example, if you are going to use Ethernet and if you see this diagram here we see we are able to tap a we have the Ethernet wire that goes from one source to the other source and we are trying to tap the data and take the data out and obviously, as you know this is for some sort of live analysis. And this is the way we actually tap the data out. And if it is optical network we use this reflector tap what it essentially does is you take the data from here and then sends it via this port and that data is captured by one of these ports then the other data that is traversing the other direction could be captured by these ports.

Essentially what happens is that any data that enters the organization can actually if it is a optical port it can be mirrored and you can store the data obviously, I mean the data will run to gigabits. But what usually we do is we store the metadata I mean that is metadata is stet about the data. So, typical network cabling is copper and then fiber optic lines and copper cable is usually used in the Ethernet and that is we saw how we can tap these physical devices and the advantage the forensic value of this on the wire tap is that you can have real time network data ok. There are various types of taps I mean you can look into the references are look into the internet on different types of taps.

One of this sometimes what happens is that in the optical taps the signal strength reduces ok. So, you need to be very careful because the signal travels to a particular distance and then you insert a tap in between and there could be signal degradation and that could affect the data transmission etcetera, and in wired networks it could introduce some sort of a noise. Usually these kind of infrastructure tap which actually replicates the signals is used by the ISPs for monitoring data traffic.

(Refer Slide Time: 14:00)



The slide is titled "In the air" and contains the following content:

- Wireless station – to – station signals
  - Radio frequency (RF)
  - Infrared (IR) – not very common
- Forensic Value:
  - Can be trivial as information is often encrypted, however valuable information can still be obtained
    - Management and controls frames are usually **not encrypted**
    - Access points (AP) advertise their names, presence and capabilities
    - Stations probes for APs and APs respond to probes
    - MAC addresses of legitimate authenticated stations
    - Volume-based statistical traffic analysis

NPTEL logo is visible in the bottom left corner.

So, coming to in the air this essentially is wireless ok. Essentially there are two types one is the radio frequency where if the data gets broadcasted so that you can put some sort of a listener, and then try to get all the data the other is the infrared type which is a point to point kind of communication. And sometimes what happens is if you put a point to point communication you actually intercept the data going on the other side. So, it could be used for interception of data if you are going to go for infrared kind of a stuff.

The advantage the forensic value of this is that sometimes you get interrupted I mean even in wired networks you might get encrypted data. So, decrypting the data yes it is a its a its a area by itself So, our for the time being we are assuming that the that the data you is either encrypted or not encrypted we just have to get the data that is that is what we are we are interested in right now. Later if the data is encrypted then you have to apply a decryption algorithms you have to find out the keys, you have to break the encryption and all those things that is a separate part in itself ok. But one of the things

that we know is that the management and control frames usually are not encrypted ok. For example, I mean if you encrypt an IP address then the router will not know how to route the data.

So, usually what happens is and the other point with encryption is that it actually slows down your network. So, people usually do not encrypt the IP addresses and all that, unless they need excellent security and the router is also configured etcetera, but essentially we can be sure that management and control frames are usually not encrypted and therefore, the collecting the metadata that is the data about the data from these management and control frames is given.

The second one is the devices that are connect themselves they advertise some information. Say for example, if you have a wireless access network and you try to scan the wireless access networks, see it tells you what are all the wireless snacks networks that are there in your environment, you would observed it in your office or in your schools. These access points some of course, you can actually make the access point not transmit their names ok. But we usually do not do it ok. So, access points can advertise their names their presence and their capabilities. So, this is another place where you can the data can be taken ok.

And you can actually probe for a piece there are small devices that are available which can actually scan the wireless networks and then probe what type of device it is, even if the device does not advertise you can actually find out what type of device you are using etcetera. The other important aspect is that you know Mac addresses, see Mac addresses you can establish since Mac addresses are usually unique in a network in a network a local network, this can also serve as an evidence to pinpoint what which machine has got affected etcetera ok. And most important thing is that volume based at statistical traffic analysis can also be used. I mean these are all the these the forensic value that in their networks in their technology gives.

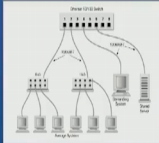
Then we come to two important aspects one is the switch and the routers this is essentially at the layer two ok. And we know that switch just actually put the lands together ok.



(Refer Slide Time: 17:21)

## Switches

- "Switches are the glue that hold LANs together" (Davidoff & Ham, 2012)
- Multiport bridges that physically connect network segments together
- Most networks connect switches to other switches to form complex network environments
- **Forensic Value:**
  - **Content addressable memory (CAM) table**
    - Stores mapping between physical ports and MAC addresses
  - Platform to **capture and preserve** network traffic
  - Configure **one port to mirror traffic from other ports** for capture with a packet sniffer



NPTEL [www.nptel.ac.in/mauritus-mu.blogspot.in/2011/10/connect-two-computers-to-my-internet.html](http://www.nptel.ac.in/mauritus-mu.blogspot.in/2011/10/connect-two-computers-to-my-internet.html)

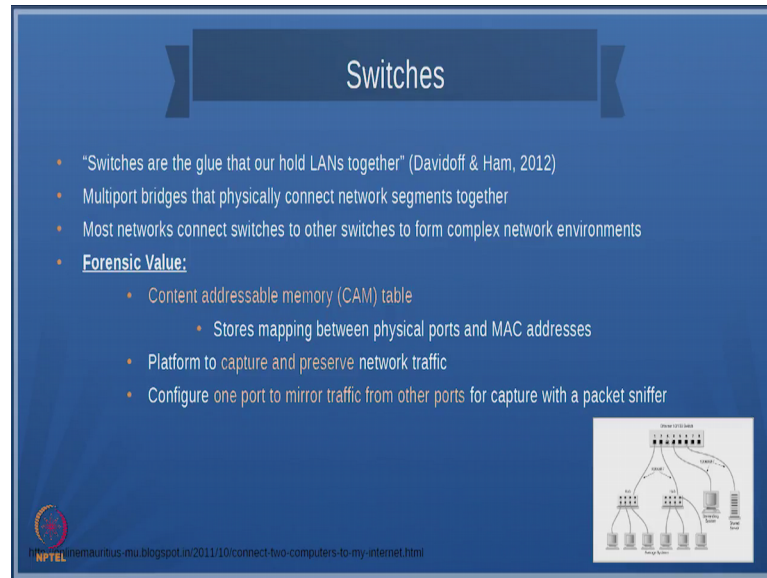
What switches, sometimes switches are also intelligent; that means, they learn the networks and some switches also throughout log information ok. Usually what happens is that if you see this diagram here switches combined two other switches and then that others which becomes very intelligent and essentially tries to learn where all the Mac addresses and all that, and routes the data based on the Mac addresses ok.

So, switches can be grouped in a hierarchical fashion and then data can be routed from one place to the other place. The acquaint the forensic value that switches provide is one is the content addressable memory, where it actually tells you in network in which segment of the network the device is available. For example, the switch will really know that this is on I mean this switch will know that this machines Ethernet address actually is used is in this can be reached via this link. Similarly this network can be reached via this link. So, if a data comes from here this switch will exactly know where to route the data based on the network the Ethernet address of these devices.

So, in that way it stores a mapping between physical ports. So, this is the port. So, it says port 2 ok, has this Ethernet address port 2 also has this Ethernet address and port to also has this Ethernet address. So, in that way the switch actually keeps track of how something can be reached. And the advantage of switch is that you can actually capture and preserve the network traffic whatever is exchanged between these devices and there is something called a port mirroring. So, what it does is that you can actually replicate

the data that comes out of one port ok, to a particular port so that means, for example, let us say I have port 3. So, I can configure the switch in such a way that all data that comes way a port 1 is also sent on port 3 that is known as port mirroring.

(Refer Slide Time: 19:38)



The slide is titled "Switches" and contains the following text:

- "Switches are the glue that hold LANs together" (Davidoff & Ham, 2012)
- Multiport bridges that physically connect network segments together
- Most networks connect switches to other switches to form complex network environments
- **Forensic Value:**
  - Content addressable memory (CAM) table
    - Stores mapping between physical ports and MAC addresses
  - Platform to capture and preserve network traffic
  - Configure one port to mirror traffic from other ports for capture with a packet sniffer

In the bottom right corner, there is a diagram of a network switch with multiple ports and connections to other devices. In the bottom left corner, there is a logo for NPTEL and a URL: <http://nptel@nptel.ac.in>

So, what we can do is it is a kind of tapping, I mean you are not physically tapping it, but you are programming the switch to tap the data. So, if you look, if you if you remember previously with respect to Ethernet or the optical device we were actually putting a physical a segment to tap it ok. Here what we do is the device is available and you program the device to tap the data. So, this is another way of capturing the data ok. So, one port to mirror the traffic from other ports for capture with a packet sniffer; So, you have a packet sniffer is could be a deviser or software which actually looks into this medium and then takes all the data that traverses in the medium.

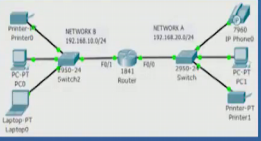
Now, now in fact, it is suspected see internationally when some of these if you if you rear the if you have the most recent cases like wiki leaks etcetera people suspect ok. If course, it may not be a fact, but people suspect that when some of these devices are say for example, devices from some companies are bought by some ISPs and then they are placed on the internet or now, these devices actually do some sort of a port replication and then send data to some other locations. I mean this is suspected. You would have heard about these cases recently also where some of the applications were banned because they were transmitting data to unknown destinations.

So, something like this could happen and this exactly one of the ways by which you can do it is port mirroring there are other ways to do it, but port mirroring is one other ways.

(Refer Slide Time: 21:30)

## Routers

- Connect traffic on different subnets or networks
- Allows different addressing schemes to communicate
- MANs, WANs and LANs are all possible because of routers
- Forensic Value:
  - Routing tables
    - Map ports on the router to networks they connect
    - Allows path tracing
  - Can function as packet filters
  - Logging functions and flow records
  - Most widely deployed intrusion detection but also most rudimentary



<https://broadcaststormlog.wordpress.com/2016/03/01/1-purpose-and-functions-of-network-devices-routers-switches-bridges-and-hubs/>

**NPTEL**

Coming to routers the routers actually connect different subnets and networks. So, and this is the one which actually allows the IP addresses and these routers use the IP addresses to communicate between different devices the MANs, metropolitan area networks, wide area networks and LANs are all possible because of routers and the forensic value that the routers provide is one they can provide routing tables. Routing tables are very important it is like telling at what time what data has been routed by a what port or at what time what data was routed by a what port ok, so what part port to what network.

So, this is a very important information ok. And this routing tables allow path tracing back tracing means how do how did the data get from the source to the destination ok. So, how does one find out that see if you read some reports they say that an attack happened in country xx and the hackers were in country yy. So, how are they able to identify this country yy its basically they do some kind of a path tracing and that path tracing tells you how the packet has traversed from the source to the destination or from the attack to the victim. The routers can also do other things they can block unwanted packets ok, usually this blocking of unwanted packets is done for prevention methods.

For example, if there is a ddos attack routers can be configured to drop packets so that it does not clog the internal network the routers actually can have flow records and then they can also do locking functions. So, the routers are much more capable of generating huge amount huge amount of logs ok. So, when I talk about filtering you see that the routers can actually block the flows and there are two things, one is the flow and another is a packet. I mean a flow is actually a series of related packets is a flow ok. For example, I can have a html session http session and the packets completely corresponding to that hp (Refer Time: 23:41) could be called as a flow anyway. We will come to these definitions when we look at the forensics.

So, routers provides some sort of a rudimentary method by which you can filter packets and get the logs.

(Refer Slide Time: 23:57)

The slide is titled "DHCP Servers" and contains the following content:

- Dynamic Host Configuration Protocol
- Automatic assignment of IP addresses to LAN stations
- **Forensic Value:**
  - Investigation often begins with IP addresses
  - DHCP leases IP addresses
  - **Create log of events**
    - IP address
    - MAC address of requesting device
    - Time lease was provided or renewed
    - **Requesting systems host name**

On the right side of the slide, there is a diagram illustrating the DHCP process between a "DHCP Client" (represented by a laptop) and "DHCP Servers" (represented by a server rack). The process is shown as a sequence of four messages:

- IP lease request (from client to server)
- IP lease offers (from server to client)
- IP lease selection (from client to server)
- IP lease acknowledgment (from server to client)

At the bottom left of the slide, there is a logo for NPTEL and a URL: <http://I4wisdom.com/linux-with-networking/dhcp-server.php>

Next one is the dynamic host configuration protocol servers. Now, DHCP protocol is almost used by I mean many these provides you very quick way of assigning IP addresses or automatic way of assigning IP addresses ok. The forensic value is that DHCP tells you at what time an IP address was least, at what time the least was revoked etcetera. So, in that way DHCP leases you can identify when DHCP actually leased out IP addresses, it can also create log of events that what is the IP address that was given, to which Mac address it was given ok, and then when the time lease was provided, and when that time lease was removed ok, and also what is the hostname of the system.

What we actually do is that you should collect all the logs if for example, is the DHCP protocol it says it is say its starting at 4 o'clock in the morning the IP address was assigned to this particular machine, which had this Ethernet address and then it it was released at 8 o'clock. So, that tells you that that machine was on the network for about 4 hours, ok. So, these type of logs can be got from DHCP servers. So, what we will do is in the next module we will look at other types of data, the other type of devices and the forensic value that the other types of devices provide to us.

Thank you.