

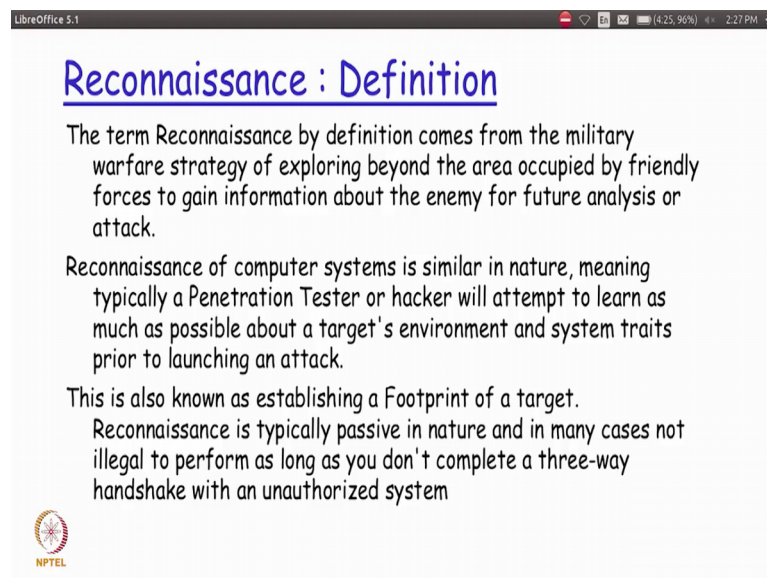
Information security-IV
Prof. Vasan
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture - 15
Reconnaissance

Over the last few sessions, we are actually been talking about different steps an attacker would typically make use of for doing penetration testing. So, if you recall the first step that we were actually talking about what is called as reconnaissance.

Starting from this particular session onwards, we will actually be looking at what each of the steps are supposed to be doing in detail and also look at, what are the different kinds of tools that Kali Linux is actually providing us for each of those steps which attacker or ethical hacker could actually make use of for accomplishing the objectives of that particular step. So, the first step that we will be actually looking at in detail is reconnaissance.

(Refer Slide Time: 01:00)




Reconnaissance : Definition

The term Reconnaissance by definition comes from the military warfare strategy of exploring beyond the area occupied by friendly forces to gain information about the enemy for future analysis or attack.

Reconnaissance of computer systems is similar in nature, meaning typically a Penetration Tester or hacker will attempt to learn as much as possible about a target's environment and system traits prior to launching an attack.

This is also known as establishing a Footprint of a target.
Reconnaissance is typically passive in nature and in many cases not illegal to perform as long as you don't complete a three-way handshake with an unauthorized system

 NPTEL

So, what exactly we mean by the term reconnaissance. So, reconnaissance by definition basically came right from the time of history where military battles actually started happening, wherein this was basically strategy of exploring, what the area occupied by the enemy forces was actually trying to do. So, it was more information gathering phase which one side of the out of the two sides encountering the battle will actually be

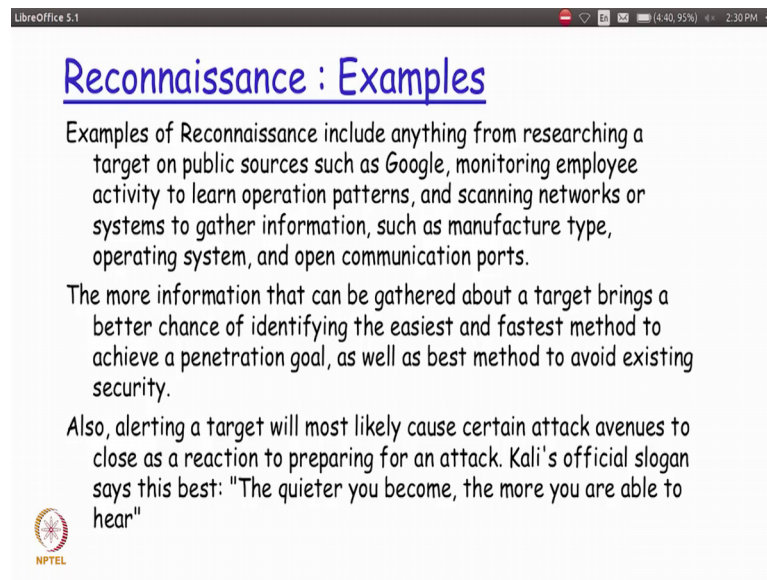
deploying for to get access about details on what exactly is the enemy area or enemy aside is actually trying to really do right now.

So, it was basically used as the first step before really launching the attack in the similar way the reconnaissance of computer systems is basically very similar in nature where a penetration tester before we actually starts going ahead with the job of doing the testing will first try to find out; how much ever information is possible about the target environment. So, he would really try to find out on what kind of business really the target is actually into what kind of systems are deployed and so on and so forth.

So, that the at the end of the face at the end of this particular step of reconnaissance, the person that attacker the presentation tester will have all the necessary details in his hand to really start doing the attack after getting the details. We also tell you call this really has establishing a footprint of a target where we try to basically make sure that the information pertaining to the attack having taken place not present as much as possible and the same time have some sort of a backdoor mechanisms for the penetration tester to make use of to enter into the same target system in future, right.

So, broadly speaking the reconnaissance is something very similar to what the military teams, basically try to do in terms of getting more details about the enemy before really launching the attack in a similar way, the penetration tester would actually tried to get the details about what actually is happening or what is the strategy that the other party is deploying. So, that the relevant tools and the relevant vulnerability could be attacked a straight away to find out how much vulnerable this the target environmental.

(Refer Slide Time: 03:50)




Reconnaissance : Examples

Examples of Reconnaissance include anything from researching a target on public sources such as Google, monitoring employee activity to learn operation patterns, and scanning networks or systems to gather information, such as manufacture type, operating system, and open communication ports.

The more information that can be gathered about a target brings a better chance of identifying the easiest and fastest method to achieve a penetration goal, as well as best method to avoid existing security.

Also, alerting a target will most likely cause certain attack avenues to close as a reaction to preparing for an attack. Kali's official slogan says this best: "The quieter you become, the more you are able to hear"

 NPTEL

So, what could be some typical examples of reconnaissance that we all would have heard of would you possibly facing in our day to day life. So, anything that we actually try to use a public source like a Google right is something that could potentially be made use of by the penetration tester because anything that we actually try to put in Google or put on the internet.

So, to say is something that as we would be seeing a little later is just recorded for posterity. So, there is no way that even if the original owner of the data is actually deleting the data from the original source, the data is going to be continuing to be remaining somewhere in some corner of the internet which separation tester could actually make use of at the right time.

So, other things could really monitoring the employing activity to learn what kind of operation patterns scanning the network. So, systems gather the information. So, some things that could be gathered really here would potentially me the type of the manufacturer, what kind of operating system is actually running on that particular system if it is really a server and then what kind of communication ports are open.

So, all these kind of things, we are going to be actually seeing subsequently in our sessions down below where we will be really demonstrating the different tools which will give us this detail. So, more information that I could gather about a target the

penetration tester will be in a much better position of identifying what is the most quickest mechanism to achieve the penetration goal right.

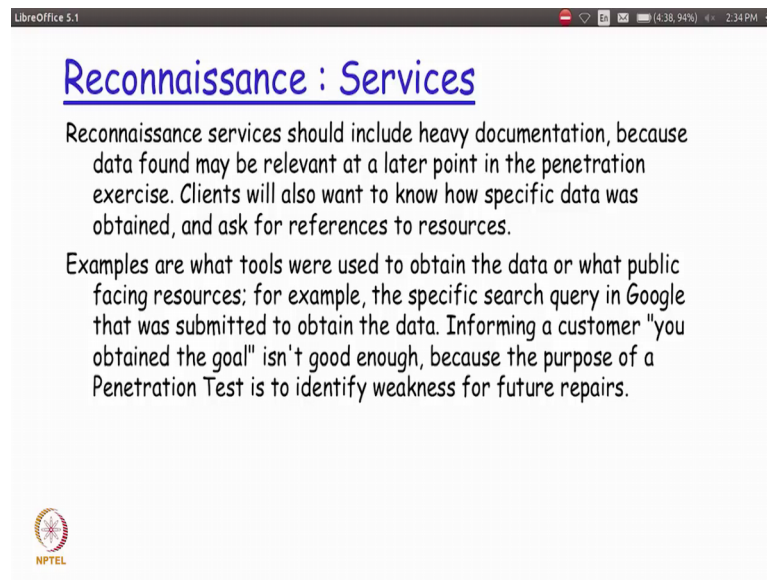
So, the penetration goal is to detect vulnerabilities make uses vulnerabilities and then attack that particular target the information gathering phase that is that is going to be actually getting done as part of reconnaissance will basically held the trade tester to achieve the goal as quickly as possible right and at the same time is also very important. That the tester make sure that the existing security mechanisms do not find out about his effort, that he has actually putting to gather the details because if that has actually been found out by the end administrator before the tester goes from recognisance phase to the actual attacking phase that vulnerability would possibly be getting addressed by the administrator right.

So, it also has to be ensured that this is actually done in a stealth manner as much as possible. So, that the objective of the penetration testing which was actually a set forth initially is successfully made. So, the objective you should always are the target should not be able to find out that the attack has actually going to commence at any point in the near future from the reconnaissance stage that has actually been done, right.

So, this is also very much in line with what is the official slogan for Kali Linux which basically says that the quieter you become the more you are able to hear right. So, we could actually feel that happening with in our human conversation itself and this is all the more.

So, in case of penetration testing where right from the reconnaissance to the actual attack and maybe the breaking down of that particular target; it should all be done in as quiet as possible in quite as quiet a manner as possible without making too much noise for it to be completely successful and the objective met.

(Refer Slide Time: 07:46)



Reconnaissance : Services

Reconnaissance services should include heavy documentation, because data found may be relevant at a later point in the penetration exercise. Clients will also want to know how specific data was obtained, and ask for references to resources.

Examples are what tools were used to obtain the data or what public facing resources; for example, the specific search query in Google that was submitted to obtain the data. Informing a customer "you obtained the goal" isn't good enough, because the purpose of a Penetration Test is to identify weakness for future repairs.

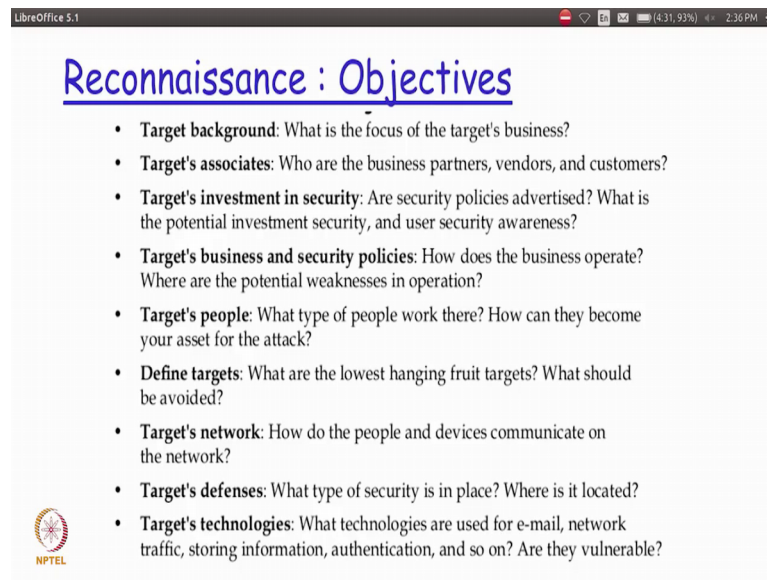
NPTEL

So, what kind of services typically go in reconnaissance, it should basically have extensive amount of documentation because the data that we actually find out as part of doing the reconnaissance is something that is going to be used as part of the penetration testing subsequently in our next set of steps.

So, i should have the necessary detail documentation of whatever has actually been found out as part of this reconnaissance. So, that when I am actually as a penetration tester starting to do the attack I do have all the details in front of me in black and white and not really depending on my memory capacity as a penetration tester. So, what kind of detail should go into the extensive documentation? So, the details could be like what kind of tools were actually used to obtain the data what kind of a publicly available resources were actually made use of.

So, let us say that we actually try to find out from Google search; what exactly was the operating system that could be I mean what could be the potential vulnerabilities in this particular operating system. For example, that specific search query that we are actually done in Google should be also recorded. So, that we actually can do the same search subsequently in order to get a background details if at all during a, our actual testing face we want to know more details about that particular target.

(Refer Slide Time: 09:21)



LibreOffice 5.1 (4.21, 93%) 2:36 PM

Reconnaissance : Objectives

- **Target background:** What is the focus of the target's business?
- **Target's associates:** Who are the business partners, vendors, and customers?
- **Target's investment in security:** Are security policies advertised? What is the potential investment security, and user security awareness?
- **Target's business and security policies:** How does the business operate? Where are the potential weaknesses in operation?
- **Target's people:** What type of people work there? How can they become your asset for the attack?
- **Define targets:** What are the lowest hanging fruit targets? What should be avoided?
- **Target's network:** How do the people and devices communicate on the network?
- **Target's defenses:** What type of security is in place? Where is it located?
- **Target's technologies:** What technologies are used for e-mail, network traffic, storing information, authentication, and so on? Are they vulnerable?

NPTEL

So, in terms of the objectives what exactly could be the objective of the reconnaissance mission is that first is to try to understand the background of the target what exactly is a business to the target is actually in right now because knowing the business will give us more details on the terms of what and how we could potentially attack the identified targets. Who are the targets associates in terms of vendors in terms of customers and so on because that could be a very valuable amount of information that the tester could basically get from this detail right.

So, how much of investment has been done by the target in terms of the security. So, in terms of the extensive investment for the network for the individual servers individual client desktop machines if as a tester I could actually have details about what kind of investment has been made then that will help me as part of doing the reconnaissance to appropriately handle it.

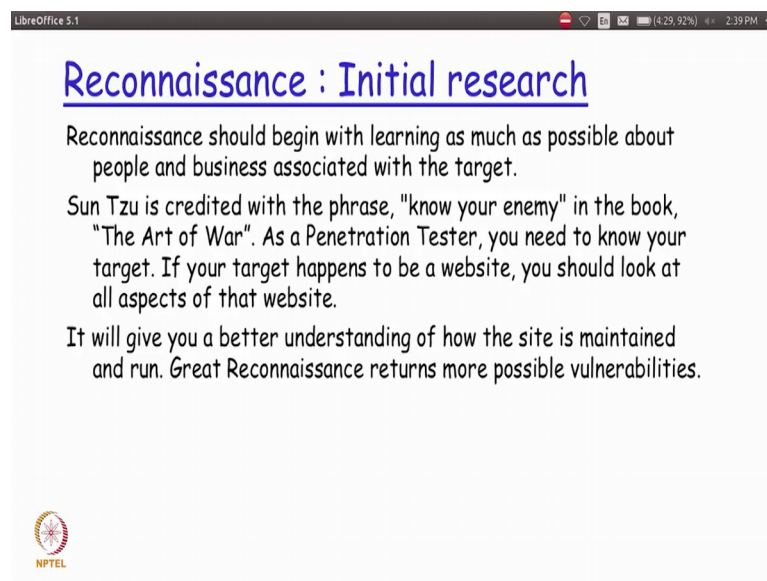
So, what is the targets a business in security policy is with which I will be able to find out what potential weakness that I could actually leverage on what is the background of the people working for the target what are the targets exact location and a definition. So, I would actually be trying to attack a network in which let us say there are twenty different servers right.

So, out of which typically speaking ideally I would not really try to concentrate on all the 20 servers at the same time, but I would really try to find out among the 20 servers for

example, which one is potentially very critical or very important that I should first concentrate on for attacking based on which possibly the remaining servers will become inaccessible or will automatically come down.

So, the definition the target is very important then identifying the targets network. So, what kind of differences are there for that particular network and then what kind of technologies are actually being used. So, basically things like what kind of network topology? What kind of network protocols? What are the databases? What kind of web technologies? and so on and so forth will be a very critical information for the attacker because the with each of those technologies the attacker will be basically have access to the set of possible vulnerabilities that is there in those technologies and once the attacker comes to know what kind of technology is deployed in the targets environment it basically makes the job of the attackers that much more easier.

(Refer Slide Time: 12:00)




LibreOffice 5.1

Reconnaissance : Initial research

Reconnaissance should begin with learning as much as possible about people and business associated with the target.

Sun Tzu is credited with the phrase, "know your enemy" in the book, "The Art of War". As a Penetration Tester, you need to know your target. If your target happens to be a website, you should look at all aspects of that website.

It will give you a better understanding of how the site is maintained and run. Great Reconnaissance returns more possible vulnerabilities.

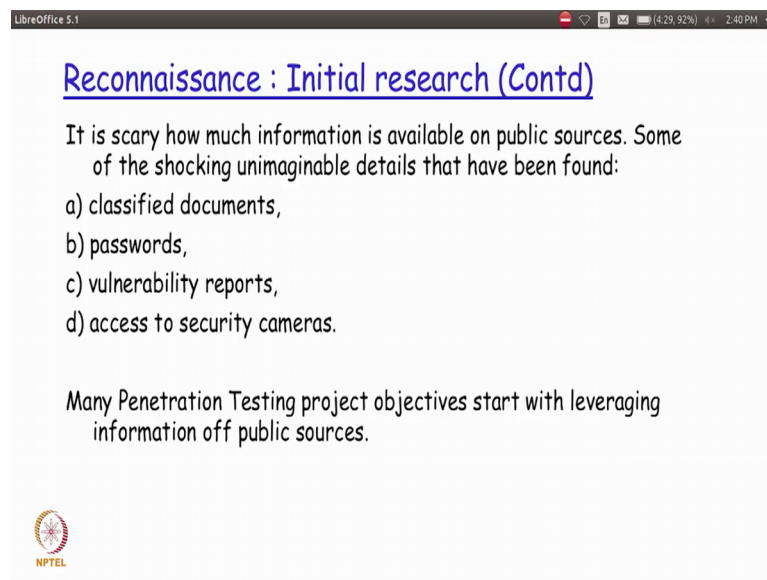
 NPTEL

Initial research is basically what reconnaissance is all about. So, it should basically try to learn as much as possible about the target. So, all the objectives that we actually saw in the previous slide we will actually try to do initial research on that and then try to get more details about it and extensively documented right. So, as the famous authors San Tzu says know your enemy in the book the art of war. So, penetration testing is all nothing, but all war strategy that is actually going to be deployed and so, knowing the

enemy that is actually mentioned in the book is as equally applicable here for penetration testing as it is applicable for fighting a big battle.

So, if for example, if the target is going to be a website it is very important that we actually try to have all the details are all the aspects of the website readymade with us; before we really launch the attack because that makes our actual attack job that much more easier.

(Refer Slide Time: 13:06)



LibreOffice 5.1

Reconnaissance : Initial research (Contd)

It is scary how much information is available on public sources. Some of the shocking unimaginable details that have been found:

- a) classified documents,
- b) passwords,
- c) vulnerability reports,
- d) access to security cameras.

Many Penetration Testing project objectives start with leveraging information off public sources.

NPTEL

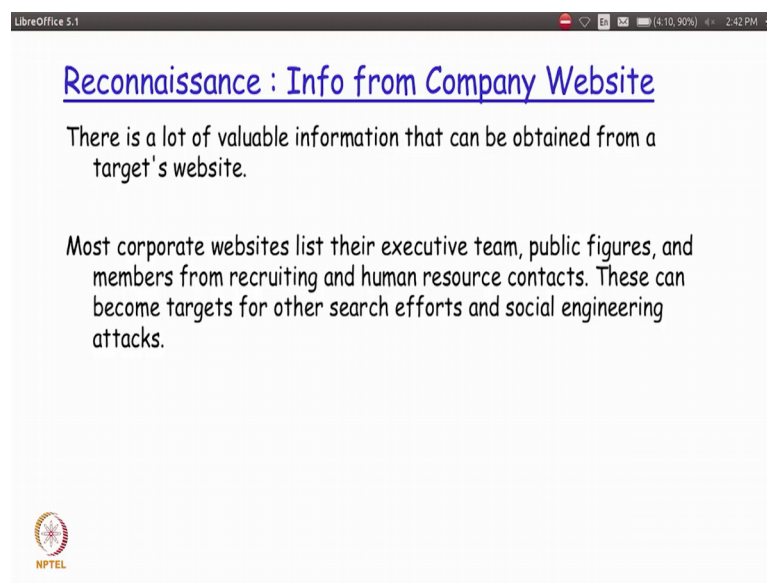
So, in terms of a initial research, what I could potentially do is I could actually try to find out various kinds of documentation a details and various kinds of data some of it which will give me as an attacker lot of details are some set of classified documents that could potentially be available as a leak on the internet passwords information, vulnerability reports and access security cameras right. So, each of these kind of details we could very clearly understand would help the attacker in doing his or her job from one aspect of it.

So, for example, if I basically tried to get any kind of details of the passwords from public sources then we could very comfortable say that half the job of the attacker is actually done because if the passwords details for the patterns, even if not the exact passwords even in the patterns are actually found out by the attacker. It basically helps him to find out what different possible combinations of the patterns.

The person could be using as the password for the particular system or network server or whatever it is and thereby get access into that particular system very easily right. Likewise getting access to classified documents or any kind of vulnerability reports or security camera recordings will basically help the attacker in quickly trying to come to conclusion on how easily that I could actually be done as part of the attack phase, right.

So, all these kind of details you will find in the internet in some form or other and we will actually be seeing a couple of examples in this particular session also.

(Refer Slide Time: 14:59)



The image shows a screenshot of a presentation slide within a LibreOffice 5.1 window. The window title bar includes the text 'LibreOffice 5.1' and system icons for network, volume, and battery, along with the time '2:42 PM'. The slide content is as follows:

Reconnaissance : Info from Company Website

There is a lot of valuable information that can be obtained from a target's website.

Most corporate websites list their executive team, public figures, and members from recruiting and human resource contacts. These can become targets for other search efforts and social engineering attacks.

In the bottom left corner of the slide, there is a small circular logo with the text 'NPTEL' underneath it.

So, there is a lot of valuable information that can be obtained from target websites. So, mostly corporate website if you find give lot of details about the email addresses about the background the company and so on and so forth. So, what I could possibly get from an email address right. So, from an email address, I could potentially very easily get the domain name with which at particular company or the target is registered with, right.

Now when I get the name of the domain very easily the attacker would basically have lot of data points in the public sources from which he could find out what are the possible IP address ranges that are actually assigned to him assigned to that particular target and so on right. So, likewise, if as an attacker, if you really view any company website as an attacker, you will very easily find out the kind of depth information depth that is actually available in it which then attacker could potentially make use of on a platter, right.

(Refer Slide Time: 16:08)

Reconnaissance : Regional Internet Registries (RIRs)

Look at: www.arin.net

Network	
Net Range	50.76.50.112 - 50.76.50.127
CIDR	50.76.50.112/28
Name	FACEBOOK
Handle	NET-50-76-50-112-1
Parent	CBC-SFBA-17 (NET-50-76-32-0-1)
Net Type	Reassigned
Origin AS	
Customer	FACEBOOK (C03029402)
Registration Date	2012-06-10
Last Updated	2013-12-09
Comments	
RESTful Link	https://whois.arin.net/rest/net/NET-50-76-50-112-1
See Also	Upstream network's resource POC records.
See Also	Upstream organization's POC records.
See Also	Related delegations.

NPTEL

So, if you for example, look at a site called arin; so, if you actually look at this particular site arin dot net this is something which an attacker penetration tester would actually be making use of as a first cut measure possibly. So, RIR basically stand for regional internet registry. So, we will actually see a very small clipping on how one could actually make use of this site to get more details about the IP addresses that the target is possible making use of.

(Refer Slide Time: 16:40)

Videos

S15-ARIN.mp4

Google Chrome

American Registry for Internet Numbers

Secure <https://www.arin.net>

Your IPv4 address is 14.139.160.229

SEARCH Whois

NUMBER RESOURCES | PARTICIPATE | POLICIES | FEES & INVOICES | KNOWLEDGE | ABOUT US | FEEDBACK

ARIN ONLINE

Log in and password are case sensitive

username

password

log in

Announcements

Fri, 22 Dec 2017
ARIN Celebrates 20th Anniversary

Wed, 20 Dec 2017
Happy Holidays from ARIN!

Wed, 13 Dec 2017

Holiday Closing Information

Highlights

Request Resources

Waiting List for Unmet Requests

Draft Policies & Proposals

Internet Governance

Resource Revocation, Returns, and Reinstatement

IPv6 Info Center

New to ARIN?

Submit Payment

Whole Inaccuracy Reporting

Billing Info Management

ARIN Mailing Lists

Jobs @ ARIN

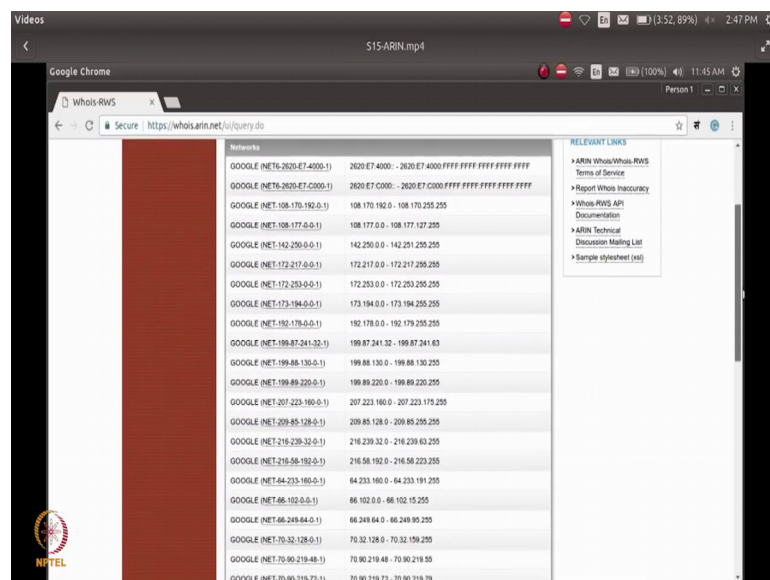
TRANSFER INFORMATION

NPTEL

So, if you actually use the URL www dot arin dot net, you will actually come across this website where first the home page you will find there is an IP address that is displayed and this IP address is the IP address from which you have actually logged in which of access this particular server, right. So, you have access the arin dot net sites from this particular IP address and this is basically what it is right now mentioning.

So, now if I basically go and type any domain name as part of this search who is database right I you will find that it basically list down all the IP addresses for that particular domain name that was actually searched, right.

(Refer Slide Time: 17:20)



So, this is basically what I was telling just a short while back to the moment, I know the domain name of my target environment maybe by looking at the IP address of some of the employees of that particular company in their website. So, if the email address will always have the domain name associated with it like for example, it could be the company name is a b c d right the name of the employee at a b c d dot com is what will be the email id typically given to that particular employee.

Now, if I have this email ID is available as part of some of the top level management, employees in a company in my website. The penetration tester could first make use of that detail from that email ID, find out what is the domain name that this particular target is actually making use of and then use the details that is actually available from sites like

arin and then get the corresponding the IP addresses that is potentially made use of by that particular target right.

So, if because the example that we actually gave year was Google dot com you find the there are so many IP addresses that is actually displayed here where has for any other kind of an organisation which is actually a small organisation you might not have this many IP address ranges because they would have actually registered let us say an example only one or two or maybe a very handful number of IP address alone if I basically go into any of those listed IP addresses listed IP address range that I have, I would basically get more details about that particular registration, right.

So, for example, so, if you see here I actually have the autonomous system that is being used for that particular range right. So, autonomous system as we were actually seeing in our previous is 3 course it is basically something which is used by the b g p routing protocol whenever it has to route from one autonomous system to another autonomous system. So, this particular will number the origin a s number will basically give me details about what is autonomous system ID right and so on and so forth, right.

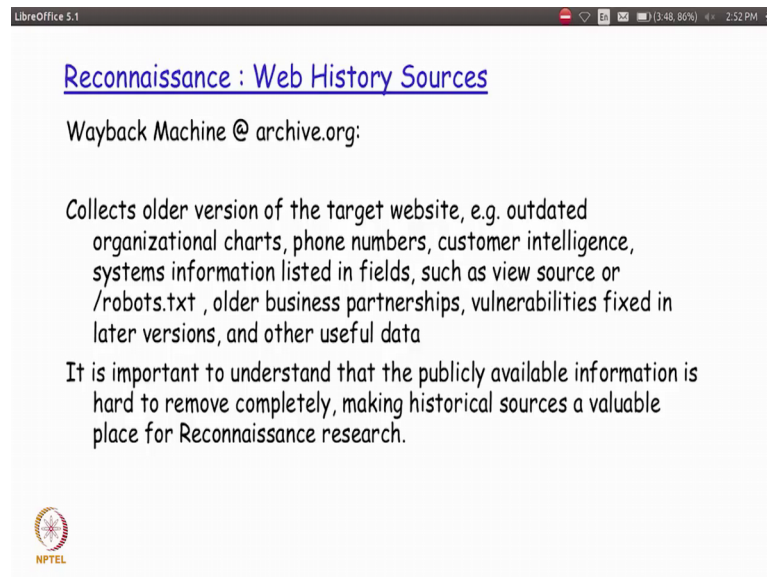
So, like for example, I have details on the registration date when was it last updated and I could potentially go and try to get further details by clicking on these links also right. So, if you are find here that the there is a whole lot of information that is actually available as part of just a simple a DNS entry registration that has actually been done by the target environment.

Similarly, if you look fat YouTube for example, is another domain that I am searching, I could actually get the similar kind of details for that particular domain name also here right. So, for this YouTube I find that the net range that has actually been allocated one of the net range that is actually allocated is this particular number. So, I get what is the corresponding c i d r that is actually used and so on and so forth, right.

So, from here we just come to an understanding of what exactly is used what kind of detail is actually available by somebody just by looking at the DNS registration that has been actually done by the target. So, first how do I get the domain name is very easy I just go and access the targets homepage a site that they actually have and you will find that most of the corporate website will typically have at least a few email ID's there. So, from the email ID, one can easily get the mapping of the hostname that is actually the

domain name that is actually used the domain name in RIR 's by doing something like it as in a r n dot net web site and then get the corresponding IP addresses that is possibly being used by that particular target right now.

(Refer Slide Time: 21:50)




Reconnaissance : Web History Sources

Wayback Machine @ archive.org:

Collects older version of the target website, e.g. outdated organizational charts, phone numbers, customer intelligence, systems information listed in fields, such as view source or /robots.txt , older business partnerships, vulnerabilities fixed in later versions, and other useful data

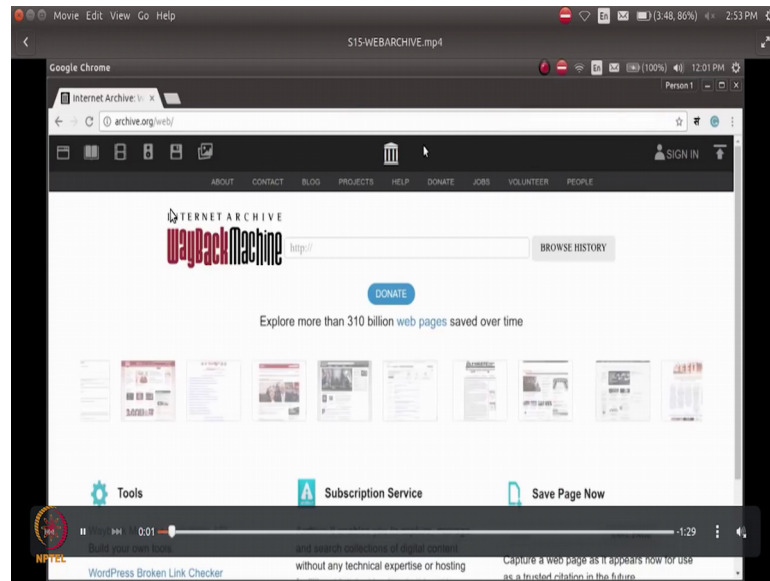
It is important to understand that the publicly available information is hard to remove completely, making historical sources a valuable place for Reconnaissance research.

 NPTEL

So, this is one level of information or rather the first level of information that penetration tester will try to get as part of the reconnaissance for now the next thing detail that is actually available which the penetration test, I will actually try to make use of as part of reconnaissance a huge amount of information available as archive.

Now, how do I actually access is archive is there is a site called archive dot org which will help you to get details about a particular website as it was containing the data as of particular data right.

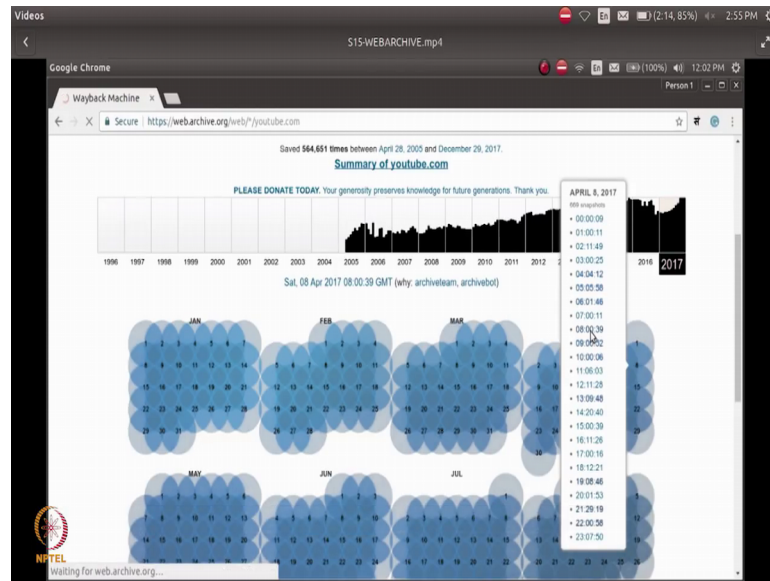
(Refer Slide Time: 22:18)



So, we will actually have a small video demonstration for that also. So, if you look at this particular site it is actually referred to as archive dot org slash web and you just need to type this particular URL as part of the browser URL field and then here you could potentially type what is the domain name or the web page that you want to get the archive listing for right.

So, if I basically want to find out what was the archive listing for you tube dot com for all the past year let me say for the past eight years, ten years or whatever it is I just give that particular domain name and then present and then you find that right from 2005 onwards up to 2017, right. I actually have got the complete arcade of all these past years whatever this particular webpage was actually storing.

(Refer Slide Time: 23:03)



So, this is basically what we were referring to as the point that once you actually put it on to the internet there is no way that information will actually get removed even if you remove it by yourself from the original place of where the data was actually kept right. So, the archive mechanism will actually take place automatically by what is called as generally robos that basically go ahead and archives the contents and something like YouTube dot com, you find that the archive information is available all the way from 2005 onwards up to the current date right.

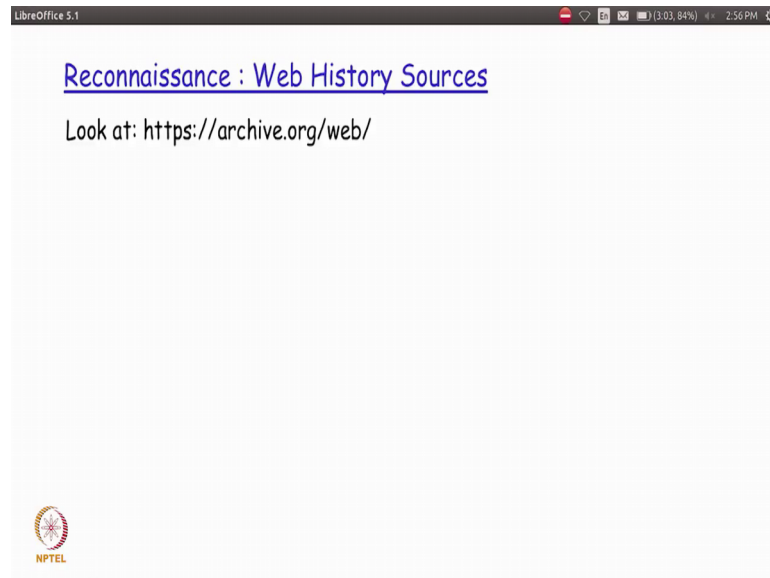
So, the entire archival for 2017 is actually displayed here and I could potentially go and choose any particular date or time within that particular date. So, for example, if I basically choose April 8 2017, I have all these times on this archive has been taken and then I go and click on that say one particular time that I am interested in, right.

So, I just go and click at 800, 39 time slot in the morning at which point in time the archive cable is actually been taking taken. So, this particular u r l field basically tells also; what is the time of archival. So, it basically 2017 0408 and then the rest of it is a time 0800, 39. So, 8 hours in the morning 39 seconds in 8 hours, 39 seconds in the morning is when this is actually been archived.

Now I basically have the details from this particular site on how what kind of data is actually stored and displayed at this particular date and time right of this particular domain that I had actually searched for right.

So, now with this you could very easily find out what kind of information that the attacker could potentially get from accessing these kind of data sets and thereby find out what whatever basics level of information; that he requires to continue is remaining reconnaissance effort and concluded after which he will all be ready for the for doing the actual attack, right.

(Refer Slide Time: 25:10)



So, any kind of data that is there in any of the very popular social media portal sites whether it is a Facebook or YouTube or LinkedIn or whatever it is all these are actually available as part of this web archive.

Thank you.