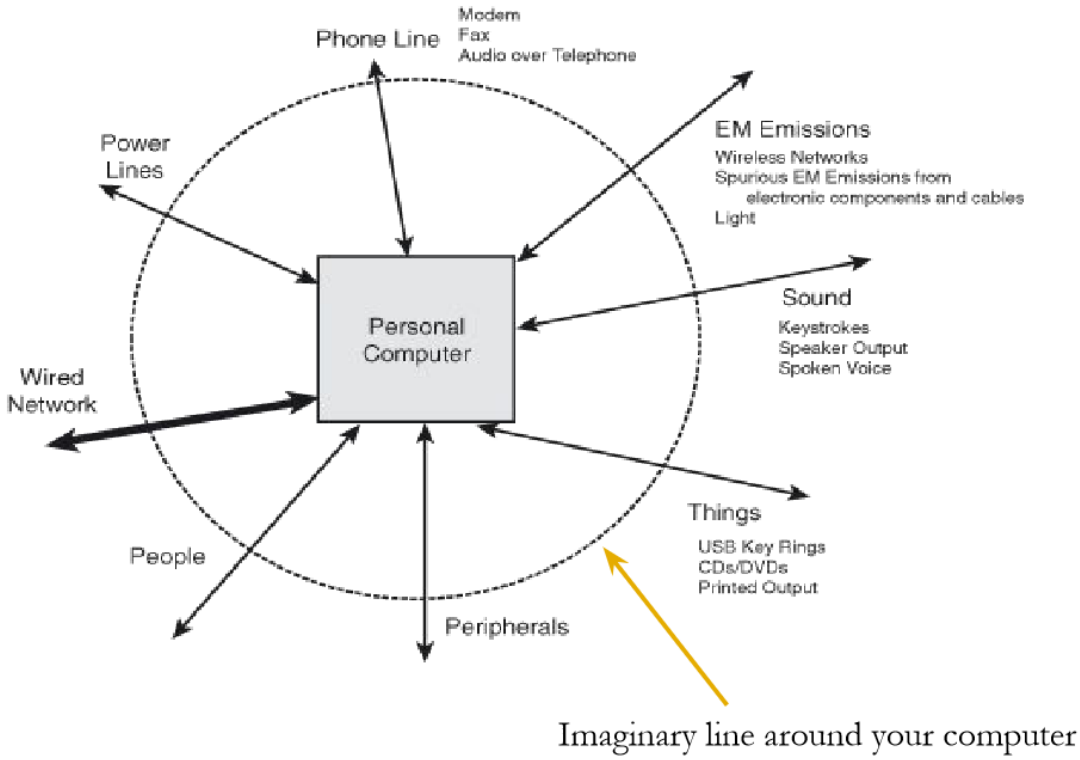
==========Start lecture 9 notes==========

Q. List five channels through which information can leak from a personal computer.



Information Flow & Leakage

Networks

▪ Wired, Wireless and Telephone networks

▪ Biggest threat vector for web-based information leakage

▪ Turning on your machine initiates a number of networking protocols and applications which reveal sensitive information

▪ ARP, DNS, Email, AV software

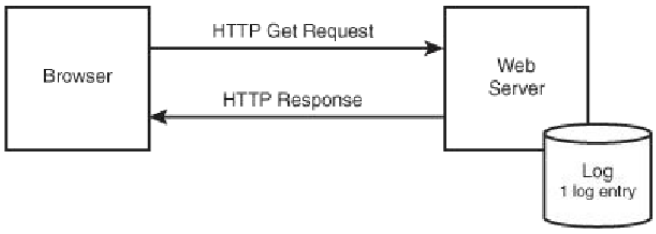
▪ Wireless machines broadcast their traffic so it is easier sometimes for an attacker to gain access. They do not need physical access.

Footprints/log file

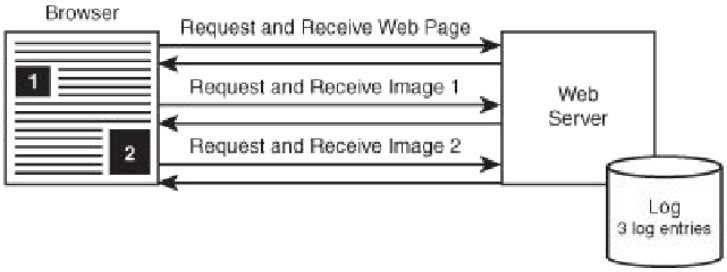
When accessing information online, we all leave a distinct trail in the logs of each server we interact with These logs can be aggregated, mined and used to create user profiles to help organisations with targeted advertising.

Web Interaction & Logs

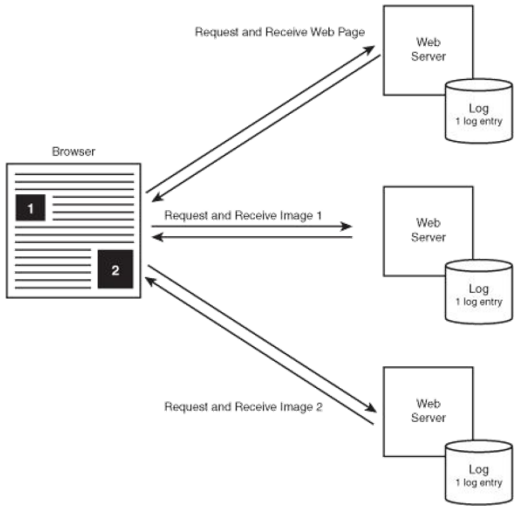
Basic Interaction for a web page:



Normal Interaction including embedded media:



Interaction with media from other 3rd Party sites



Q. List six types of information stored on a webserver logfile in response to connecting a browser navigating to a page on that server.

The Trail You Leave Behind

Web Server Logs

▪ Each time a browser accesses a web page, image or object, an entry is placed in a web server log

▪ Example of a Log Entry: ▪ 86.X.X.X - - [27/Feb/2013:04:49:28 -0700] "GET / HTTP/1.1" 200 15384 "http://www.google.co.uk/search?hl=en&q=MALWARE &meta=" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 2.0.50727)“

Breakdown:

▪ 86.X.X.X –> IP Address of device making the request

▪ 27/Feb/2013:04:49:28 -0700 -> Date and time of request

▪ GET / HTTP/1.1" 200 15384 -> The HTTP request

▪ http://www.google.co.uk/search?hl=en&q=MALWARE &meta= -> The referring URL, the search term used and that the user speaks english.

▪ Version of the browser used, the client OS

The Trail You Leave Behind

- IP Address
- Browser Header fields
- Cookies
- HTTP Referer data

Q.What are cookies used for?

Q.What privacy risks do cookies present?

Cookies

▪ Web servers use cookies to mark web browsers with identifying information due to the fact that HTTP is stateless and cannot keep track of previous interactions.

▪ Two main types of cookies exist:

▪ Persistent cookies which can exist for many years in a user's browser cache, repeatedly identifying the user to the issuing website across many visits.

▪ Session cookies which exist only for the duration of a single online visit.

▪ Cookies are issued either by the web server of the site a browser is visiting (first-party cookies) or by a third party when a given page includes embedded content, such as an advertisement or video, provided by a third-party server.

▪ Cookies are small pieces of data that are passed to and stored by the browser by a given website. When needed, such as on a repeat visit, the cookie is passed back to the web server to uniquely identify the user

▪ Although it is possible to set a web browser to block cookies, doing so breaks many online applications

▪ Each browser on a given computer, such as Firefox, IE, or Chrome, maintains a separate set of cookies

▪ In addition, most OS maintain a separate profile for each user of the browser, each with its own distinct set of cookies

▪ Cookies uniquely identify a browser/user account combination for a given individual

▪ The same cookie used from different IP addresses would allow an online company to map the networks of different domains that the user visits, perhaps including home, workplace, school, and travel destinations

Q. Give account of inter website tracking.

Inter Website Navigation

Click-through Tracking

▪ JavaScript is used on web pages to provide interaction

▪ JavaScript can also be used on a webpage to allow the current web server to be able to detect and log which link a user clicked on a page

▪ This technique is used by Google which ties together search queries with the links that you click

▪ If JavaScript is turned on in your browser, the actual links Google returns point back to Google

Q. Give account of intra website tracking.

Intra Website Navigation

Intra website navigation tracks a user's browsing pattern within a single domain It has been used for website analytics for a number of years Navigation within a particular website can reveal user's desired goals and could also be used to identify them Most users exhibit unique or at least uncommon navigation behaviours that uniquely identify them

Uniqueness and Behavioural Targeting

So far, we have looked at the different ways in which a user discloses information to online companies either knowingly or not

▪ IP Addresses

▪ Cookies

▪ Browser Header Fields

▪ User Accounts

▪ Semantic Information

Large organisations then use this information to uniquely identify a user, profile their activities and possibly link all of this to discover their real-world identity

Q .Describe behavioural targeting. What are its benefits to companies?

Behavioural Targeting

Also known as **User Profiling**, this uses web interaction data to categorise a user's interests and allow targeted advertising.

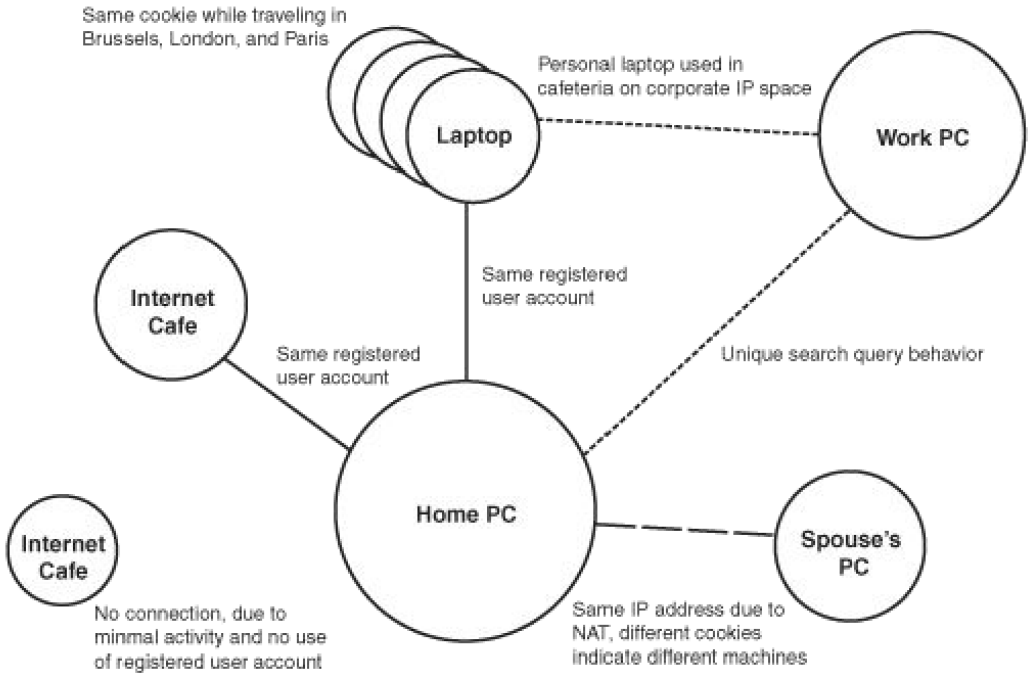
It allows companies to increase the effectiveness of their advertising campaigns by capturing data generated by website and landing page visitors.

When it is done without the knowledge of users, it may be considered a breach of browser security and illegal by many countries' privacy, data protection and consumer protection laws

The typical approach to this starts by using web analytics to break-down the range of all visitors into a number of discrete channels

Each channel is then analysed and a virtual profile is created to deal with each channel.

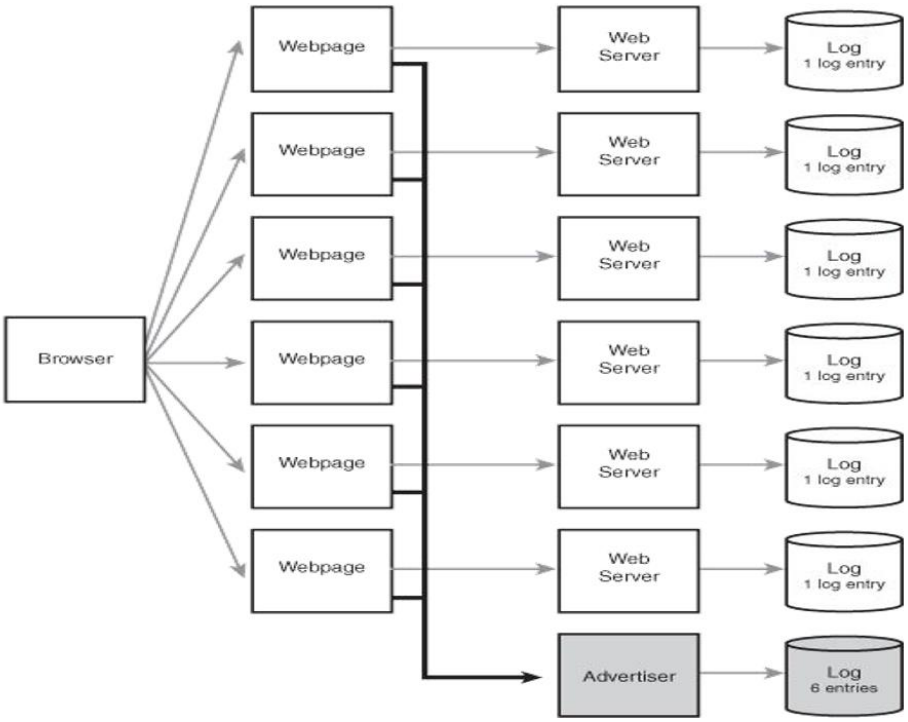
Q.Give account of how a user's home, work PC and laptop can be linked together.



Information which we reveal in our searches

▪ Locations ▪ Religious Affiliations ▪ Medical Conditions ▪ Business Plans ▪ Hobbies and Interests ▪ Stage of Life ▪ Hopes and Dreams

<u>Cross-Site-Tracking</u>



Cross Site Tracking

In the previous slide we see an example of Cross Site Tracking

▪ A user visits 6 distinct websites each hosting content from a single advertiser

▪ In turn, the user's visits create one set of log entries on each of the six legitimate servers

▪ However, because each visit contained an advertisement from a single advertiser, the advertiser is able to log all six visits

Advertising

Advertising is the fuel behind virtually all free tools online It is also the means for tracking your web activity across the internet Web Advertisements are big business with Google earning almost $50 billion in 2013 Some of the largest services are offered by Google. ▪ AdSense ▪ AdWords ▪ DoubleClick

AdSense

▪ Service which allows webmasters to earn advertising revenue by hosting AdSense ads ▪ The revenue from hosting these ads can range from a few hundred dollars a month to $50,000 a year ▪ It is an extremely popular service ▪ The ads are context-sensitive textual ads served by Google based on the hosting site's content ▪ Unfortunately, just visiting these websites allows Google to collect a user's IP address and log their visit

Q. Describe counter measures to protect your online identity

1.Counter Measures

No clear solution exists to protecting your identity online It is necessary to employ a number of different techniques but you must decide which ones suit you as locking down your information completely can hinder your web browsing experience Techniques ▪ Patching Users ▪ Technical Protection

2.Raised Awareness

▪ Realise that a problem exists

Know what you are divulging

▪ Some things you need to consider ▪ Consider each new tool carefully and understand the implications of the personal information that is disclosed through its use ▪ Consider how much of what you are sharing you would ever share with your parents, partner, or co-workers ▪ Read what privacy policies say directly and what they say between the lines ▪ Disclose the minimum information to accomplish a task ▪ Think in terms of years' or decades' worth of disclosures, not single instances or even days

3.Usable Security

▪ Protection against web-based information disclosure demands usable security ▪ Usable security seeks to help people remain secure by creating systems that are designed to be effective, efficient, understandable, and easy to use ▪ Such systems don't waste users' time and attention; instead, they seek to disturb users only when they need to make a decision ▪ A trade-off exists between usability and privacy using today's technology—each current approach comes at a cost

4.Controlling Cookies

▪ Reduce or eliminate cookies that accumulate on your computer ▪ Modern browsers provide intelligent cookie management ▪ Check out your browsers cookie management capability and what it sees as best practise

5.Protect your network address

▪ Along with cookies, your network IP address identifies you to online companies every time you visit ▪ Your address might also be included in the header information of each email you send. Blocks of IP addresses are allocated to ISPs, companies, organizations, and educational institutions ▪ Perhaps the best hope for anonymous web browsing is the anonymising proxy (TOR). Anonymising proxies act as intermediaries between you and the destination website. They make requests on your behalf, filter some identifying information (such as cookies, browser header fields, etc.), replace your IP address with theirs, and pass responses back to you

==========End lecture 9 notes==========