Web Exploitation

How to become an online spider

Computer Networks

Modern life would be very different without computer networks. These generally comprise of multiple computers (‘*nodes*’), that are connected together to share data and resources. The most popular Computer Network is *The Internet*, which specifically connects computers that use the Internet Protocol or ‘IP’.

How does the Internet work?

Completely new and need to know the basics? Here is a great article that explains the very basic architecture of the internet and how data is transmitted.

Website Basics

Now information on the Internet is segregated by websites. They are a collection of web pages and are referred to by a domain name (like google.com, facebook.com). Each web page is referred to by its URL or Uniform Resource Locator.

1. What is a web page and website?
   A website is a collection of web pages. So website would be like a house and each webpage would be a room inside the house.

2. Breakdown of a URL:
   https://developer.mozilla.org/en-US/docs/Learn/Common_questions/What_is_a_URL

3. Querying:
   https://en.wikipedia.org/wiki/Query_string

4. Different parts of a website and how to mess with it (HTML, CSS, JS, Backend)
   a. HTML breakdown
      Here is a basic tutorial on HTML:
      https://www.w3schools.com/html/html_basic.asp

b. CSS breakdown
   Here is a basic tutorial on CSS:

   https://www.w3schools.com/css/css_intro.asp

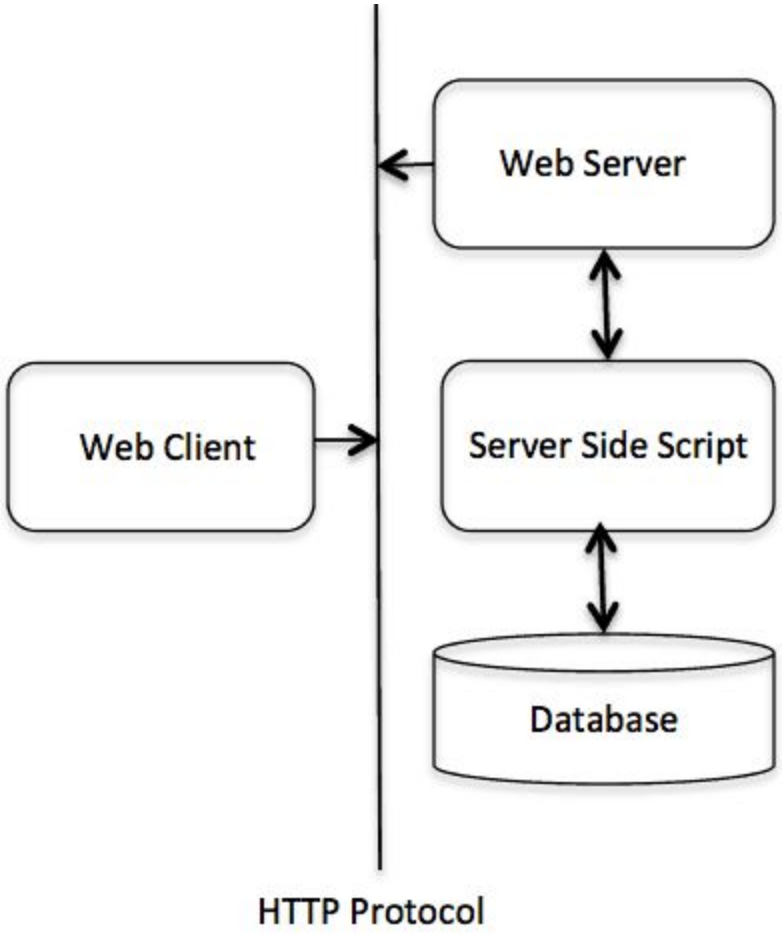
5. Viewing source
   By right clicking on Google Chrome or Firefox you can select the option “View Page Source” to see the code that the website is running on your computer. It allows you to see the HTML and CSS that is running on the website and it will also let you see the Javascript scripts running on your computer. The best part is, that you can edit the HTML directly and see it affect the website, so it lets you modify the website as you desire. You can also select “Inspect Element” to see the code that is running in a specific part of a website.

JS Breakdown

1. Why we need it?
   Javascript is used because it allows us to add interactivity between the user and the website. Javascript allows the user to interact with the website and have the website respond.
2. Basics - Editing elements HTML
   https://www.w3schools.com/js/js_htmldom_html.asp

HTTP breakdown

1. What is HTTP?
   It provides a standardized way for computers to communicate with each other over the internet. HTTP is a communication protocol, that is used to deliver data (HTML files, image files, query results, etc.) over the internet. HTTP dictates how data is sent between clients (you) and servers.
2. GET and POST request
   https://www.w3schools.com/tags/ref_httpmethods.asp
3. Basic Architecture



4. Basic idea of a proxy
   A proxy server is a computer on the web that redirects your web browsing activity. When you try to access any website, your Internet Service Provider (ISP) makes the request for you and gives the website your IP address. So when you use a proxy, your request goes from your ISP to the proxy server to the website you want to go to. This way allows you to mask your IP address as another address so that the websites you access don't know who you are.

Database breakdown

1. What they are and why they are useful
   A database is a collection of information that is organized so that it can be easily accessed, managed and updated. Databases can quickly query data and add/delete data instantly. They are used to hold every kind of data.
2. SQL and others

SQL is Structured Query Language (SQL), a programming language used for managing relational databases. Relational databases are tabular database in which data is defined so that it can be reorganized and accessed in a number of different ways. Relational databases are easy to extend, and a new data category can be added after the original database creation without requiring that you modify all the existing applications. Relational databases are made up of a set of tables with data that fits into a predefined category. Each table has at least one data category in a column, and each row has a certain data instance for the categories which are defined in the columns.

3. How they integrate into sites
   Databases are integrated into websites because they are the most optimal way to display/store data. User information like passwords are stored using databases. Databases also allow for quickly modifying the data displayed on the website. So if someone wants to update information on a website instead of modifying the HTML on the website, they can just change the data on the database that is displayed on the website.

4. Basic SQL syntax
   a. SELECT
      Extracts data
   b. ORDER BY
      Orders the results gotten from SELECT in a specific manner. For example, if one has a table of countries and their populations. One can select the countries starting with the letter R and then order them by their population.
   c. JOIN
      Joins data from two tables depending on a certain characteristic on the table. So if a theres two table one with customer IDs and their addresses and another table with customer IDs and their purchases. You can join both tables so that the customer addresses match their purchases.
   d. DELETE, INSERT
      Allows you to delete data or add new data to a table.
   e. AND,OR
      Allows you to modify queries so that they return information depending on multiple categories.
   f. MIN,MAX
      They return the smallest or largest value of a query.

Injections

1. How to perform a basic SQL injection and how its possible
   SQL Injection (SQLi) refers to an injection attack where an attacker can execute whatever SQL commands they want that control a web application's database server. Websites use the information you give them to query SQL, for example when logging onto a website, the website will query your username and password to see if you are an authorized user. A SQL injection would allow the user to supply their own SQL code and run it on the website.
2. How to safeguard against SQL injection(safely accept user input)
   a. Prepared statements
      Prepared statements are a way to separate code and user supplied input. It's the most common way to avoid attackers from attempting to run SQL code by supplying it as input. Prepared statements will not execute the SQL code and will treat the input as a query and not as code.
3. How to perform a PHP object injection
   https://www.tarlogic.com/en/blog/how-php-object-injection-works-php-object-injection/

Glossary

1. *'IP' or 'Internet Protocol'*: A set of rules that govern how data is transmitted over The Internet.
2. *IP Address*: A unique name given to every computer connected to the internet. It looks like 'a.b.c.d' where each of a,b,c,d is a number between 0 and 255.
3. *Packet*: In terms of the Internet, if the amount of data being transmitted is too large, we break it down into smaller chunks, called packets.
4. *Port Number*: In Networking, a port is an endpoint of the communication and the port number is the specific number associate with a particular port.
5. *Domain Name Service (DNS)* : A database which stores the IP address of each website and its *domain name* (like google.com).
6. *Client:* These are usually computers of users looking to access web pages or search engines. These are usually the ones looking to get a particular service.
7. *Server:* These are computers that store web pages, services or applications. They are usually the ones providing the service.
8. *Webpage:* A single hypertext document that is connected to the World Wide Web.
9. *Website*: A collection of related web pages usually connected to one common domain name.