웹 브라우저란 HTML 문서와 이미지, 음성, 동영상 등을 www 즉 월드 와이드 웹을 기반으로 한 인터넷의 컨텐츠에 접근하기 위한 응용 프로그램을 뜻합니다. 여러분이 자주 보시는 웹 브라우저로는 크롬, 엣지, 모바일용 삼성 인터넷 등등이 있습니다. 이 브라우저는 페이지를 다운로드하기 위해 응용 계층의 HTTP 프로토콜을 이용해 데이터를 송신/수신 합니다.

이번 시간에는 우리가 웹 브라우저에서 어떻게 컨텐츠를 화면에서 볼 수 있는지 그 과정에 대해 알아보고, 추가적으로 이 과정 속에서 일어날 수 있는 일들을 방지하는 방법에 대해 알아보도록 하겠습니다.

설명에 들어가기 앞서 간단한 용어 정리부터 하겠습니다.
isp는 개인이나 기업체에게 인터넷 접속 서비스, 웹사이트 구축 및 웹호스팅 서비스 등을 제공하는 회사 즉 통신사입니다. 여러분이 라우터 공유기라도 부르는 곳에 와이파이 혹은 유선으로 연결하는 것을 lan이라고 하죠? 이 공유기들에 연결된 AS라고 하는 자율 지스템에 연결되어 광범위의 인터넷 연결을 가능하게 합니다. isp가 이 as의 일종입니다.

ip는 네트워크 상에서 정보를 수신하고 송신하는 통신 규약입니다. 이후에 더 자세하게 다룰 예정입니다.
ip 주소는 이 ip통신을 하기 위해 필요한 고유한 주소입니다. 일반적으로 ip는 숫자와 .으로 나누어진 체계를 따르고 있습니다.
도메인은 인터넷에 연결된 컴퓨터를 사람이 쉽게 기억,입력할 수 있도록 문자로 만든 IP로 url이라고 생각하시면 됩니다. url 은 기본적으로 www.~~.~~.~~로 이루어져 있죠 이것이 도메인이라고 생각하시면 됩니다.
프로토콜은 컴퓨터 내 외부에서 데이터의 교환 방식을 정의하는 규칙 체계로 위에서 보신 ip도 프로토콜의 일종으로 인터넷 프로토콜의 약자입니다.
TTL은 time to live 데이터의 생존 시간으로 데이터가 발송되고 네트워크상에서 소멸하기까지의 시간을 의미합니다.

이제부터 웹의 접속 과정에 대해서 들어가보도록 하겠습니다.
여러분이 만약 웹에서 원하는 정보를 얻으려고 할 때 일단 브라우저부터 실행하게 됩니다. 이렇게 브라우저를 실행하게 되면 자동으로 설정된 웹 페이지가 실행이 됩니다. 만약 특정한 웹 페이지를 요청한다면 다음과 같이 url을 주소창에 입력하게 됩니다.
구글을 예로 들어 설명하겠습니다. 여러분이 www.google.com을 입력한다면 브라우저는 해당 url에 해당하는 ip주소를 찾아야 합니다. 아까 말씀드렸듯이 url은 사람이 보기 편리하게 만들었기 때문에 ip주소로의 변경이 필요합니다. 이것이 DNS입니다.

dns에 대해 설명하기 전 도메인 이름에 대해 한번 자세하게 살펴보겠습니다.
도메인 네임은 일반적으로 다음과 같이 구분되어 있습니다. 앞에 응용 프로토콜이 위치하고 다음으로 도메인 이름 그리고 상세 페이지로 이루어집니다 도메인 체계는 오른쪽 그림과 같이 국가도메인 혹은 일반 도메인으로 이루어져 .을 통해 나누어져 있습니다.
아래 예시를 보시면 구글.co.kr와 google.com 이 있는데 각각 최상위 도메인이 맨 마지막에 위치하는 것을 확인해 볼 수 있습니다.

이러한 도메인은 종종 nameserver라고 불리기도 합니다. DNS는 거대한 분산 시스템으로 Domain Name Space는 DNS가 저장,관리하는 계층 구조를 가지게 됩니다. 여기서 각각의 . 으로 구분된 도메인들을 nameserver라고 부를 수 있습니다. 이러한 도메인을 제공하는 회사로는 대표적으로 cloudfare가 있습니다. 아래 예를 통해 보시면 이해가 더 쉬울 것입니다.

기초 지식을 숙지하셨으니 이제 DNS가 어떻게 url을 통해 ip주소를 찾는지에 대해 에측해 볼 수 있을 것입니다. 앞서 말했듯이 dns란 도메인을 ip주소로 변환해주는 행동을 취합니다.
이 과정에서 dns 캐싱이라는 기법을 사용하게 됩니다.

dns 캐싱이란 운영체제에서의 캐싱과 동일합니다. 나중에 재사용할 수 있는 도메인 이름에 대한 ip주소 유지 기법으로 isp에서부터 캐시를 확인하여 도메인 이름에 대한 ip를 추적하는 방식입니다. 로컬에서 찾는 동시에 루트, 하위 nameserver에서 탐색하며 ip주소를 특정하게 됩니다. 방법으로는 재귀, 반복이 있는데 대부분 재귀적 탐색을 사용하고 루트에서부터 그 아래로 내려가며 재귀적으로 ip주소를 탐색하게 됩니다. 재귀적 방법은 효율적이지만 다양한 위험성이 있습니다.

첫째로는 dns 증폭 공격이 있습니다. 모든 증폭 공격은 공격자와 대상 웹 자원 간의 대역폭 비용 차이를 이용합니다. 아까 말했듯이 재귀적 탐색은 많은 메모리를 필요로 하게 됩니다. 여기서 dns 증폭 공격은 출발지 ip 주소를 조작하여 dns 요청에 대한 응답이 조작된 ip 주소로 전송 되도록 하는 공격 방법입니다.
예를 들자면 다음과 같습니다. DNS 증폭에서의 봇은 식당에 전화해 "거기 있는 메뉴를 모두 1인분씩 주문할 테니 내게 전화해서 내가 주문한 걸 다 말해주세요"라고 말하는 악동을 생각하면 됩니다. 식당에서 어디로 전화하냐고 물을 때 목표 피해자의 전화번호를 알려주는 것입니다. 그러면 목표 피해자는 식당으로부터 본인이 요청하지도 않은 엄청난 양의 정보를 받게 됩니다.

다음으로는 캐시 포이즈닝이 있습니다. 캐시 포이즈닝은 위와 비슷하게 쿼리를 조작하여 잘못된 웹 사이트로 연결하게 만듭니다.

dns 캐싱에서의 작동방식을 이해하셨다면 dns 캐싱에 문제점이 뭔지 알 수 있을 것입니다. 바로 사용자의 요청이 기록된다는 것입니다.

자 지금까지 ip 주소를 알아가기 위한 여정이었습니다 이제 dns를 통해 ip 주소를 특정했습니다. 이렇게 ip 주소를 알게 되면 브라우저가 해당 ip인 서버와 tcp 연결을 시작합니다.

tcp 연결을 알기 위해 먼저 프로토콜에 대해 간단한 설명 하고 넘어가겠습니다. 프로토콜은 컴퓨터 사이 통신을 위한 통신 규약으로 하나의 프로토콜로는 데이터 송수신의 문제를 해결할 수 없다.

프로토콜의 계층은 4계층으로 이루어져 있습니다. 어플리케이션 계층, 전송 계층, 인터넷 계층 네트워크 인터넷 계층으로 이루어져 어플리케이션에서 전송한다면 tcp ip lan을 통해 인터넷

에 전송되게 됩니다.

응용단계는 이미 많이 사용해보셨으니 생략하고 ip로 바로 넘어가겠습니다. ip는 인터넷 프로토콜로 네트워크 상에서 정보를 전달하는 통신 규약입니다. ip는 패킷 단위로 이루어져 다양한 정보를 포함하고 있습니다. 이 패킷 단위로 데이터가 이동하게 됩니다.

a 클라이언트에서 서버로 ip패킷을 보내면 인터넷 상에서 노드들을 통과하며 서버에 전달되게 됩니다.

이렇게 전달되는 ip 패킷에는 한계가 있습니다. 첫째로 클라이언트는 서버의 상태를 파악할 수 없어 ip 패킷을 받을 대상이 없거나, 서비스 불능 상태여도 패킷이 전송되게 됩니다. 둘째로 중간 노드(서버)의 장애 발생 시 클라이언트가 파악할 수 없어 패킷이 중간에 소실될 수 있습니다. 셋째로는 패킷의 순서를 보장할 수 없음. 마지막으로는 동일 ip를 사용하는 서버와의 통신이 둘 이상일 경우 어떤 정보를 누가 요청했는지 모호함이 생기게 됩니다

이 한계를 보완하기 위해 tcp 개념이 탄생했습니다. tcp 패킷을 만들어 그 위에 ip 패킷으로 덮어서 부족한 정보를 보완했습니다.

이렇게 만들어진 tcp 연결은 3 ways handshake 기법을 통해 이루어집니다. 클라이언트에서 서버에 접속을 요청하는 SYN 패킷을 전송하고 서버에서 그에 해당하는 요청을 수락한다는 패킷인 ACK와 SYN을 전송하고 클라이언트가 ACK 패킷을 서버에게 전송해 연결이 성립되게 됩니다.

이 3ways handshake를 사용하면 서버의 상태를 클라이언트가 추측할 수 있고, ip의 비신뢰성을 보완할 수 있고 순서를 보장해 줄 수 있고 추가정보를 통한 모호함을 해결할 수 있습니다.

자 이제 tcp연결이 이루어졌습니다. 이제 요청을 보내고 받기만 하면 됩니다. 브라우저에서 웹 서버에 http 요청을 전송하게 됩니다.
url 과 ip 정보를 포함한 http 요청을 보내고 tcp 연결을 통해 해당 ip 주소에 해당하는 서버에 요청이 가게 됩니다.

해당 서버에서 요청을 받고 해당하는 응답을 전송해줍니다. 결과적으로 브라우저에서 요청한 주소에 해당하는 html문서를 확인해 볼 수 있습니다.

지금까지 웹에서 주소를 요청하고 해당 주소를 확인하는 과정에 대해 간단하게 살펴보았습니다. 이 과정에 다수의 취약점이 있다는 점을 캐치해 볼 수 있습니다. 다양한 ip주소의 유출, 중간 패킷의 탈취 등등의 사건이 발생할 수 있습니다. 이번에는 그 중에서 여러분의 ip주소가 탈취당하지 않는 방법에 대해 몇 가지 선택사항들을 제시해 보겠습니다.

첫째로는 vpn입니다. vpn이란 Virtual Private Network로 사용자가 사설망에 연결된 것처럼

인터넷에 접속하게 해주는 서비스입니다. 그림을 보시면 아시겠지만 사용자가 가상 네트워크에 연결되고 해당 네트워크에서 요청 서버를 호출해 사용자에게 해당 요청사항을 보내줍니다. 이 과정을 거치게 된다면 isp로부터 여러분이 어떤 요청을 했는지 숨겨지게 되고 요청한 페이지에서도 어떤 ip 주소가 요청했는지 알 수 없게 됩니다.

하지만 vpn에도 완전한 익명성을 보장하지 않습니다. 앞에서 isp로부터 여러분이 어떤 요청을 했는지 숨겨지게 되고 요청한 페이지에서도 어떤 ip 주소가 요청했는지 알 수 없게 됩니다. 라고 했는데 vpn 서버와 request서버 vpn 서버와 user와의 연결은 드러나 있습니다. 이 외에도 다양한 위험들이 존재합니다.

그를 보완하는 시스템으로는 tor 시스템이 있습니다. tor 시스템은 어찌 보면 가상 네트워크와 비슷하다고 볼 수 있습니다. 가상 터널 네트워크로 3개의 임의 서버를 통해 트래픽을 전송하고 각 계층별 암호화를 수행합니다.

이런 tor로 완벽하지는 않습니다. 물론 거의 완벽하지만 어떤 사용자가 요청했는지 찾을 수 있는 방법이 존재합니다. 네트워크상에서 완벽한 익명성은 기대하지 않는 것이 좋습니다.