4월 11일 목요일 회의록

**- url 받아오기**

1. wireshark로 http패킷 확인해봤으나 host, 즉 도메인 주소가 나타나는 부분에 ahnlab.com~~같은 것만 나오거나 패킷 내용이 암호화 되어 있거나 해서 알아낼 수 없었음

2. javascript코드로 현재 url알아내는 코드 돌렸으나 이거는 내 생각으로는 코드 상에 돌아가는 페이지들의 url을 알아내는 것 같음 -> 현재 윈도우에서 사용자가 보고있는 브라우저의 url을 가져오는 것이 아닌 것 같음

3. url로 피싱사이트 탐지하는 연구(논문)들 찾아보면서 사용자 브라우저 url 어떻게 받아왔는지 알아보고 있었음

-> 그러다가 url hooking하는 방법을 썼다길래 찾아봄 -> <https://ezbeat.tistory.com/470> 읽고 해봐야함

**- 이미지 크롤링**

1. 현재 사이트 이미지 저장은 접속한 첫 페이지 화면에서 들어갈 수 있는 모든 트랜잭션에 대한 웹사이트 이미지를 저장함

2. 저장된 이미지의 웹사이트에서 들어갈 수 있는 트랜잭션에 대해서도 웹사이트 이미지를 저장해야함 -> 그래서 저장된 이미지의 url에 대해서도 크롤링을 할 수 있도록 코드를 고쳐야함.

-> 그래서 기존 크롤링 코드를 수정하여 저장된 이미지의 url에서 들어갈 수 있는 웹사이트도 이미지를 저장하도록해야함