시스템프로그래밍 과제 보고서 Proxy 2-4

수업 명: 시스템프로그래밍 월5수6

담당 교수: 김태석 교수님

학과: 컴퓨터정보공학부

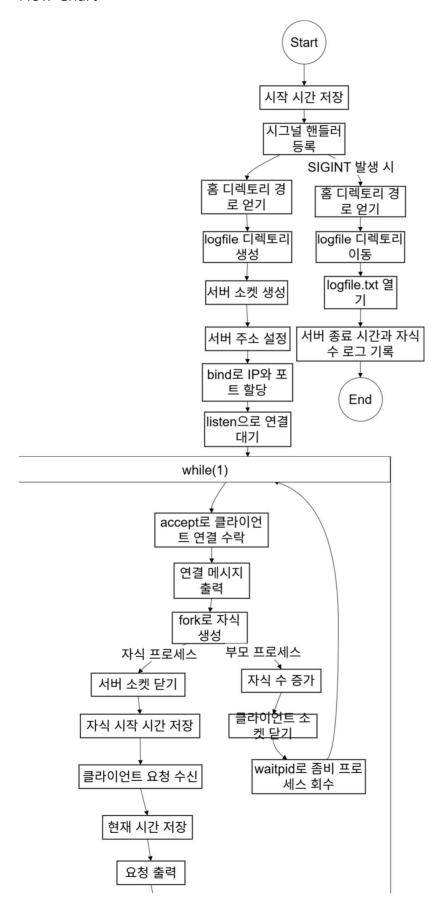
학번: 2023202070

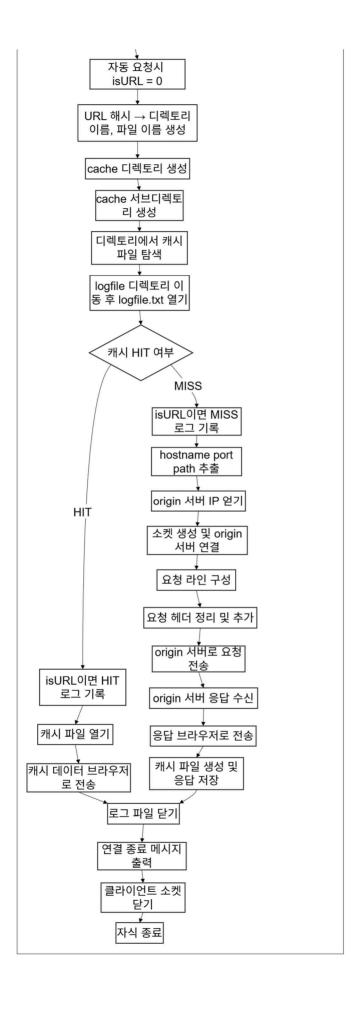
이름: 최현진

제출일: 2025.05.22

Introduction

이번 Proxy 2-4 과제는 이전 Proxy 2-3 과제에서 구현한 기능을 기반으로, 프록시 서버의 캐시 및로그 처리 방식에 대한 세부 동작을 확장하여 구현하는 실습 과제이다. 프록시 서버는 웹 브라우저로부터 수신한 HTTP 요청에 대해 SHA-1 해시 기반의 캐시 구조를 활용하여 HIT 또는 MISS를 판별한다. 사용자가 직접 입력한 요청의 경우에만 로그 파일에 기록하며, 자동으로 발생하는 요청 (favicon, css 등)은 로그에는 남기지 않고 캐시 파일에만 저장한다. 캐시 저장은 요청마다 고유의 해시값을 생성하여 파일 단위로 수행되며, 기존 캐시 여부에 따라 해당 응답을 웹 브라우저에 전달한다. 서버 종료 시에는 SIGINT 시그널을 통해 프로세스를 정리하고, 실행 시간 및 자식 프로세스 개수를 로그 파일에 기록한다. 본 과제를 통해 멀티프로세스 구조에서의 시그널 제어, 파일 기반 캐싱 처리, 선택적 로깅 기법 등을 실습하며, 시스템 프로그래밍의 세부 동작 흐름에 대한 이해를 심화할 수 있다.





Pseudo code

```
main:
프로그램 시작
  시작 시간 저장
  SIGINT 핸들러 등록
  홈 디렉토리 경로 얻기
  ~/logfile 디렉토리 생성
  서버 소켓 생성
  서버 주소 설정
  bind()로 IP/PORT 할당
  listen()으로 클라이언트 연결 대기
무한 반복:
  accept()로 클라이언트 연결 수락
  연결 메시지 출력
  fork()로 자식 생성
  자식이면:
     서버 소켓 닫기
     자식 시작 시간 저장
     클라이언트 요청 수신
     현재 시간 저장
     요청 출력
     method가 GET이 아니거나 URL이 없거나 .ico/.css/.txt 포함되면 isURL = 0
     SHA1 해시 수행
        앞 3자리 → 캐시 디렉토리 이름
        뒤 37자리 → 캐시 파일 이름
```

~/cache 디렉토리 생성

~/cache/xxx 디렉토리 생성

디렉토리에서 캐시 파일 탐색

~/logfile 디렉토리 이동 logfile.txt 열기

캐시 HIT이면:

isURL이면 [HIT] 로그 기록 캐시 파일 열기 캐시 데이터 브라우저로 전송

캐시 MISS이면:

isURL이면 [MISS] 로그 기록

요청에서 hostname, port, path 추출 origin 서버 IP 얻기 소켓 생성 및 origin 서버 연결

요청 라인 구성 (HTTP/1.0) 요청 헤더 정리 및 추가 origin 서버로 요청 전송

origin 서버로부터 응답 수신 응답을 브라우저로 전송

캐시 파일 생성 및 응답 저장

로그 파일 닫기 연결 종료 메시지 출력 클라이언트 소켓 닫기 자식 종료

부모면:

자식 수 증가 클라이언트 소켓 닫기 waitpid로 좀비 프로세스 회수

SIGINT 발생 시:

홈 디렉토리 경로 얻기
~/logfile 디렉토리 이동
logfile.txt 열기
서버 종료 시간과 자식 수를 로그로 기록
프로그램 종료

결과 화면

1. proxy 2-4

```
1. make
 cw2023202070@ubuntu: $ ls
Desktop Documents Downloads Makefile Music Pictures proxy_cache.c Public Templates Videos kw2023202070@ubuntu:-$ make
gcc proxy_cache.c -o proxy_cache -lcrypto
kw2023202070@ubuntu:-$ ls
Desktop Documents Downloads Makefile Music Pictures proxy_cache proxy_cache.c Public Templates Video
$ ls: Makefile과 소스 코드를 작성하였다.
$ make: make를 실행하여 컴파일을 수행하였고, gcc 컴파일 명령어가 정상적으로 실행되었다.
$ ls: 컴파일 결과, 실행 파일 proxy_cache가 생성된 것을 확인했다.
```

2. Operation

http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html 을 브라우저 상에 입력했다.

```
kw2023202070@ubuntu:~$ ./proxy_cache
[192.168.245.128 : 56676] client was connected
Request from [192.168.245.128 : 56676]
GET http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html HTTP/1.1
Host: gaia.cs.umass.edu
User-Agent: Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:136.0) Gecko/20100101 Firefox/136.0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
Connection: keep-alive
Upgrade-Insecure-Requests: 1
Priority: u=0, i
======= MISS =======
======= [192.168.245.128 : 56676] connect with origin server ========
======= send request =======
GET /wireshark-labs/HTTP-wireshark-file3.html HTTP/1.0
Host: gaia.cs.umass.edu
User-Agent: Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:136.0) Gecko/20100101 Firefox/136.0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
Connection: keep-alive
Upgrade-Insecure-Requests: 1
Priority: u=0. i
======= response buffer ========
HTTP/1.1 200 OK
Date: Wed, 21 May 2025 20:38:54 GMT
Server: Apache/2.4.6 (CentOS) OpenSSL/1.0.2k-fips PHP/7.4.33 mod_perl/2.0.11 Perl/v5.16.3
Last-Modified: Wed, 21 May 2025 05:59:02 GMT
ETag: "1194-6359f0fd92c74"
Accept-Ranges: bytes
Content-Length: 4500
Keep-Alive: timeout=5, max=100
Connection: Keep-Alive
Content-Type: text/html; charset=UTF-8
<html><head>
```

```
<a name="10"><strong><h3>Amendment X</h3></strong></a>

The powers not delegated to the United States by the Constitution, nor prohibited by it to the states, are reserved to the states respectively, or to the people.
</body></html>
[192.168.245.128 : 56676] client was disconnected
```

56676 client가 요청을 보내오고 child process는 MISS임을 판단했다. 이후 origin web server로 연결하고 write() 통해 보낸 요청과 read()를 통해 받은 응답 버퍼를 cmd에 출력해봤다. (너무 길어서 중간에 잘랐다.) 해당 응답은 캐시 파일에 저장된다. 이후 web server와의 연결이 끊기고 client 연결과 child process가 종료된다.

```
[192.168.245.128 : 56678] client was connected
Request from [192.168.245.128 : 56678]
GET http://gaia.cs.umass.edu/favicon.ico HTTP/1.1
Host: gaia.cs.umass.edu
User-Agent: Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:136.0) Gecko/20100101 Firefox/136.0
Accept: image/avif,image/webp,image/png,image/svg+xml,image/*;q=0.8,*/*;q=0.5
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
Connection: keep-alive
Referer: http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html
Priority: u=6
______
======== MISS =======
======= [192.168.245.128 : 56678] connect with origin server ========
======= send request =======
GET /favicon.ico HTTP/1.0
Host: gaia.cs.umass.edu
User-Agent: Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:136.0) Gecko/20100101 Firefox/136.0
Accept: image/avif,image/webp,image/png,image/svg+xml,image/*;q=0.8,*/*;q=0.5
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
Connection: keep-alive
Referer: http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html
Priority: u=6
======= response buffer ========
HTTP/1.1 404 Not Found
Date: Wed, 21 May 2025 20:39:00 GMT
Server: Apache/2.4.6 (CentOS) OpenSSL/1.0.2k-fips PHP/7.4.33 mod perl/2.0.11 Perl/v5.16.3
Content-Length: 209
Keep-Alive: timeout=5, max=100
Connection: Keep-Alive
Content-Type: text/html; charset=iso-8859-1
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">
<html><head>
<title>404 Not Found</title>
</head><body>
<h1>Not Found</h1>
The requested URL /favicon.ico was not found on this server.
</body></html>
[192.168.245.128 : 56678] client was disconnected
```

이후 56678 client의 favicon.ico이 자동으로 들어온다. child process는 MISS임을 판단했다. origin web server로 연결하고 write() 통해 보낸 요청과 read()를 통해 받은 응답 버퍼를 cmd에 출력해봤다. 해당 응답은 개별 캐시 파일에 저장된다. 그러나 MISS 판별 결과는 로그에 저장되지 않는다. favicon.ico 요청 같은 경우, 웹 서버에서 응답을 받기까지 시간이 다소 걸리기에 바로 다음 요청을 입력하거나 강제 종료 등을 하면 프로그램 동작에 차질이 생긴다. 대기 후 입력했다. 이후 web server와의 연결이 끊기고 client 연결과 child process가 종료된다.

```
[192.168.245.128 : 49482] client was connected
------
Request from [192.168.245.128 : 49482]
GET http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html HTTP/1.1
Host: gaia.cs.umass.edu
User-Agent: Mozilla/5.0 (X11; Ubuntu; Linux x86 64; rv:136.0) Gecko/20100101 Firefox/136.0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
Connection: keep-alive
Upgrade-Insecure-Requests: 1
If-Modified-Since: Wed, 21 May 2025 05:59:02 GMT
If-None-Match: "1194-6359f0fd92c74"
Priority: u=0, i
______
======= HIT =======
[192.168.245.128 : 49482] client was disconnected
[192.168.245.128 : 49496] client was connected
......
Request from [192.168.245.128 : 49496]
GET http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html HTTP/1.1
Host: gaia.cs.umass.edu
User-Agent: Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:136.0) Gecko/20100101 Firefox/136.0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
Connection: keep-alive
Upgrade-Insecure-Requests: 1
If-Modified-Since: Wed, 21 May 2025 05:59:02 GMT
If-None-Match: "1194-6359f0fd92c74"
Priority: u=0, i
======== HIT ========
[192.168.245.128 : 49496] client was disconnected
```

http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html 을 브라우저 상에 두 번 재입력했다.

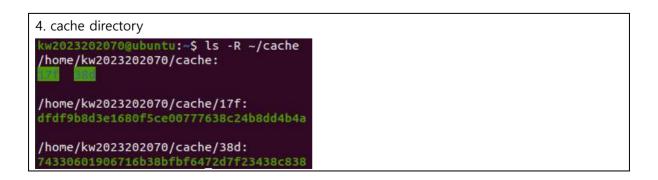
각각 49482, 49496 client가 요청을 보내오고 child process는 HIT임을 판단했다. 이후 child process는 캐시 파일에 저장되어 있는 response를 client(browser)로 보낸다. client 연결과 child process가 종료된다.

이후 ctrl+c를 입력하여 프로그램을 종료했다. 종료 로그가 기록된다.

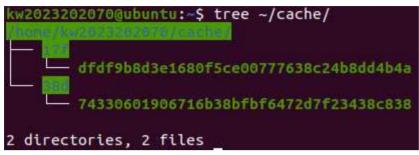


3.logfile ^Ckw2023202070@ubuntu:-cat ~/logfile/logfile.txt [MISS]http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html-[2025/05/21, 13:38:54] [HIT]17f/dfdf9b8d3e1680f5ce00777638c24b8dd4b4a-[2025/05/21, 13:39:10] [HIT]http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html [HIT]17f/dfdf9b8d3e1680f5ce00777638c24b8dd4b4a-[2025/05/21, 13:39:12] [HIT]http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html **SERVER** [Terminated] run time: 22 sec. #sub process: 4 http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html url에 대한 한 child process에

서 MISS 1번, 또다른 두 개의 child process에서 HIT 2번일 때의 로그 파일 출력 결과를 확인했다. ctirl+c 입력 후 종료 로그는 프로그램 총 실행 시간과 생성된 총 자식 프로세스 개수를 나타낸다. 위에 서술한 MISS 1번, HIT 2번 외에도 favicon.ico 자동 요청(MISS)을 처리한 자식 프로세스 1개도 포함하여 총 4개인 것을 확인했다.



\$ Is -R ~/cache: Is 명령어를 -R 옵션을 통해 재귀적으로 실행하여 ~/cache 디렉토리와 그 하위 디렉토리까지 출력하여 캐시 디렉토리와 파일이 생성됨을 확인했다.



\$ tree ~/cache/: ~/cache 구조 확인 결과, SHA1 해시된 url의 앞 3글자를 이름으로 하여 디렉토리가 생성되었다. 그 디렉토리의 하위에는 나머지 37글자 이름으로 파일이 생성되었다.

입력한 http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html url에 대한 MISS는 로 그파일과 캐시에 모두 기록한 반면(17f로 시작), http://gaia.cs.umass.edu/favicon.ico 자동 요청에 대한 MISS는 캐시에만 기록했다(38d로 시작).

```
dfdf9b8d3e1680f5ce00777638c24<u>b8dd4b4a</u>
       Open ▼ 🗊
     1 HTTP/1.1 200 OK
2 Date: Wed, 14 May 2025 14:47:28 GMT
     2 Date: Wed, 14 May 2025 14:47:28 GMT
3 Server: Apache/2.4.6 (cento5) OpenSsL/1.0.2k-fips PHP/7.4.33 mod_perl/2.0.11 Perl/v5.16.3
4 Last-Modified: Wed, 14 May 2025 05:59:01 GMT
5 ETag: "1194-635123eee37eb"
6 Accept-Ranges: bytes
7 Content-Length: 4500
8 Keep-Alive: timeout-5, max=100
9 Connection: Keep-Alive
    10 Content-Type: text/html; charset=UTF-8
    12 <html><head>
    13 <title>Historical Documents: THE BILL OF RIGHTS</title></head>
    16 <body bgcolor="#ffffff" link="#330000" vlink="#666633">
    17 <br>
    18 
    10 
10 </pr>
10 </pr>
10 </pr>
10 
10 
10 
11 
11 
12 
12 
13 
13 
14 
  22
23 23 27 
23 24 The Conventions of a number of the States having, at the time of adopting 24 the Constitution, expressed a desire, in order to prevent misconstruction 25 or abuse of its powers, that further declaratory and restrictive clauses 26 should be added, and as extending the ground of public confidence in the 27 Government will best insure the beneficent ends of its institution; 
28 States of America, in Congress assembled, two-thirds of both Houses concurring, 29 that the following articles be proposed to the Legislatures of the several 30 States, as amendments to the Constitution of the United States; all or any 31 of which articles, when ratified by three-fourths of the said Legislatures, 32 to be valid to all intents and purposes as part of the said Constitution, 33 namely: 
<a href="mailto:america">(p)</a>
<a href="mailto:america">(p)
                                                                                                                                                                                               Resolved, by the Senate and House of Representatives of the United
   34
35 ---congress shall make no law respecting an establishment of 36 religion, or prohibiting the free exercise thereof; or 37 abridging the freedom of speech, or of the press; or the 38 right of the people peaceably to assemble, and to petition 39 the government for a redress of grievances.
    40
41 <a name="2"><strong><h3>Amendment II</h3></strong></a>
43 
43 
43 
43 
44 of a free state, the right of the people to keep and bear 45 arms, shall not be infringed.
  47 <a name="3"><strong><h3>Amendment III</h3></strong></a>
 48
49 No soldier shall, in time of peace be quartered in any house, 50 without the consent of the owner, nor in time of war, but 51 in a manner to be prescribed by law.
  53 <a name="4"><strong><h3>Amendment IV</h3></strong></a>
 55 The right of the people to be secure in their persons, houses,
 63
64 >>p>No person shall be held to answer for a capital, or otherwise 65 infamous crime, unless on a presentment or indictment of a grand 66 jury, except in cases arising in the land or naval forces, 67 or in the militia, when in actual service in time of war 68 or public danger; nor shall any person be subject for the 69 same offense to be twice put in jeopardy of life or limb; 70 nor shall be compelled in any criminal case to be a witness 71 against himself, nor be deprived of life, liberty, or property, 72 without due process of law; nor shall private property be 73 taken for public use, without just compensation.
  75 <a name="6"><strong><h3>Amendment VI</h3></strong></a>
  77 In all criminal prosecutions, the accused shall enjoy the right
 86 <a name="7"><strong><h3>Amendment VII</h3></strong></a>
 88   In suits at common law, where the value in controversy shall 89 exceed twenty dollars, the right of trial by jury shall be
  90 preserved, and no fact tried by a jury, shall be otherwise 91 reexamined in any court of the United States, than according 92 to the rules of the common law.
    94 <a name="8"><strong><h3>Amendment VIII</h3></strong></a>
  96 Excessive bail shall not be required, nor excessive fines 97 imposed, nor cruel and unusual punishments inflicted.
    99 <a name="9"><strong><h3>Amendment IX</h3></strong></a>
101Tolor positive to deny or disparage others retained by the people.
104 <a name="10"><strong><h3>Amendment X</h3></strong></a>
105
 106
```

고찰

이번 Proxy 2-4 과제를 수행하면서, 프록시 서버가 수신하는 요청 중 어떤 정보를 로그와 캐시로 남길지 선택적으로 처리하는 방식에 대해 고민하고 구현할 수 있었다. 이번 구현에서는 사용자가 브라우저에 직접 입력한 요청만을 판별해 로그에 기록하고, 자동으로 발생하는 요청(favicon, css 등)은 캐시에는 저장하되 로그에서는 제외하는 방식으로 개선하였다. 이러한 방식은 실제 시스템 의 로그 관리 효율성을 고려해, 필요한 정보만을 남기는 설계 경험으로 이어졌다

또한 제안서에서의 예시 형식은 MISS 개수를 #sub process에 출력한 것을 확인했다. 예시는 참고일 뿐이니 반드시 그 형식을 고수해야 하는 것은 아니라 판단하였고 조교님께도 질문했기 때문에, 종료 로그인 **SERVER** [Terminated] 줄에서는 전체 실행 시간과 함께 생성된 자식 프로세스의 개수를 자동 요청까지 포함하여 정확히 기록하도록 구현하였다. 이를 통해 로그가 선택적으로 간결하게 구성되면서도, 시스템 전체의 실행 흐름은 마지막 로그 한 줄로도 정확히 파악할 수 있게하였다.

실행 중 가장 까다로웠던 점 중 하나는 자동 요청이 들어온 직후, 아직 오리진 서버로부터 응답을 완전히 받아오기 전 상태에서 사용자가 브라우저에 새로운 입력을 넣으면 프로그램 흐름이 중 단되는 현상이었다. 예를 들어 favicon 요청이 오리진 서버로 전송된 뒤 응답을 받는 데 시간이 몇 초 소요되면, 해당 자식 프로세스가 종료되지 않은 상태에서 새로운 요청이 이어지면 전체 흐름이 꼬이는 문제가 발생했다. 이를 해결하기 위해, 자동 요청 처리 응답이 모두 끝날 때까지 (cmd 출력으로 확인했다.) 잠시 대기한 뒤 브라우저 입력을 다시 시도해야 정상적인 순서로 흐름이 유지되었다. 이 과정을 통해 비동기적 요청 환경에서의 응답 지연과 사용자 입력 타이밍의 관계를 직접 체감할 수 있었고, 실제 서버 프로그램에서는 동기화와 응답 처리 타이밍 조절이 매우 중요한 요소임을 깊이 인식할 수 있었다.

Reference

시스템프로그래밍 이론 및 실습 자료 참고하였습니다.