캡스톤 디자인

최종 보고서 – 도비 프로젝트

December 11 / 2024

2 분반

Contents

프로젝트 소개 **1**

1 프로젝트 배경 1

2 도비 프로젝트 5

프로젝트 설계 **6**

1 설계 목표 6

2 기능 설계 7

3 기술 설계 8

데스크톱 애플리케이션 8

메인 서버 9

악성 도메인 탐지 AI 모델 11

4 시스템 설계 13

프로젝트 책임 영역 **15**

1 프로젝트 책임 영역 15

프로젝트 기대 효과 **20**

1 프로젝트 기대 효과 20

프로젝트 소개

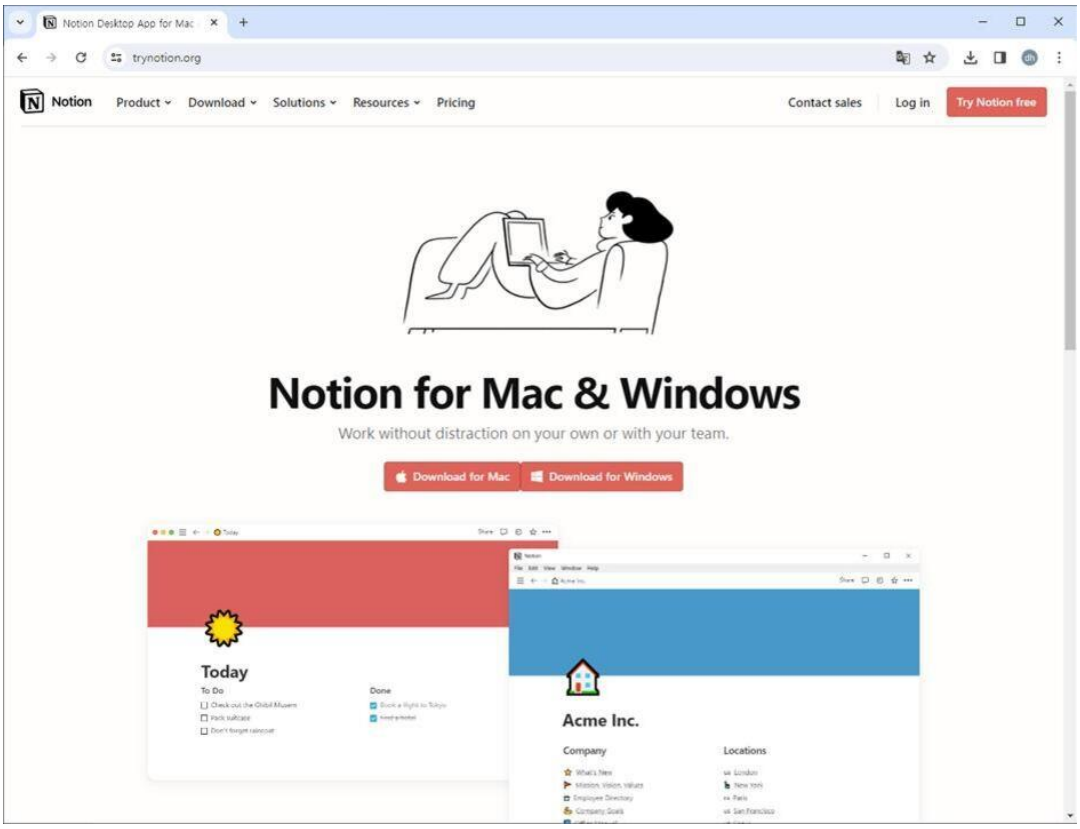
1. 프로젝트 배경

2020 년 초, 코로나 19 팬데믹은 전 세계적으로 IT 시장의 성장을 가속화했습니다. 많은 사람들의 IT 의존도가 증가하면서 온라인 기반 여가 활동, 재택근무, 방역 및 안전 서비스가 활성화되었습니다. 하지만, 이러한 성장과 함께 악성 도메인의 수가 급격히 증가하였고, 이를 신속하게 탐지하고 대응해야 할 필요성이 커졌습니다.

악성 도메인은 사용자를 속이거나 악의적인 목적으로 만들어진 웹사이트로, 주로 피싱, 악성 코드 배포, 개인정보 탈취 등의 공격에 사용됩니다. 사용자는 습관적으로 웹사이트에 접속하는 행동만으로도 큰 피해를 입을 수 있습니다. 이로 인해 악성 도메인 탐지와 대응의 중요성이 대두되고 있습니다.

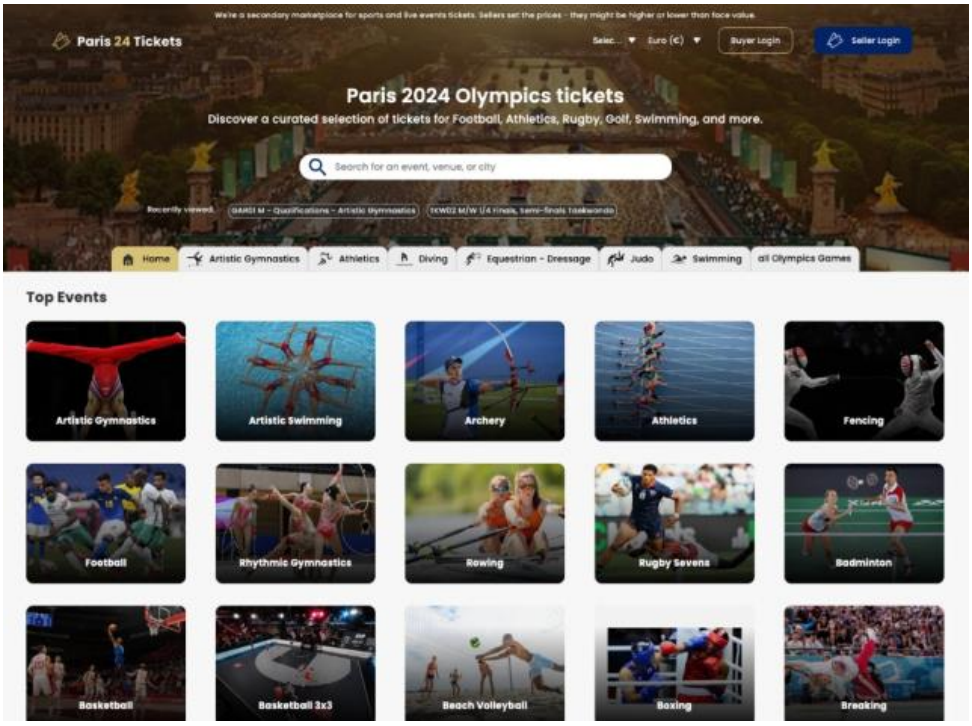
유명 사이트를 사칭한 악성 도메인 사례

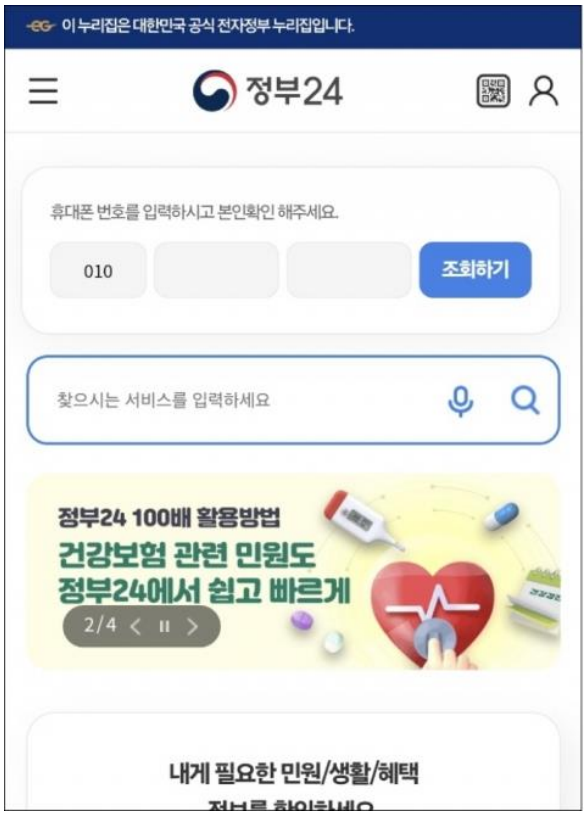
최근 들어 '넷플릭스', '노션'과 같은 글로벌 서비스 사이트를 사칭한 악성 도메인이 급증하고 있습니다. 2024 년 3 월, '노션'을 사칭한 악성 도메인 사이트는 많은 사용자의 민감한 정보를 탈취하거나 악성 코드를 배포해 큰 피해를 일으켰습니다. 넷플릭스를 사칭한 사이트도 계정 정보 및 금융 정보를 탈취하려는 시도가 있었으며, 많은 사용자들이 피해를 입었습니다.



정부 및 공공기관 사칭 사례

‘국민건강보험공단’, ‘정부 24’, ‘파리올림픽 티켓 구매 사이트’와 같은 정부 및 공공기관을 사칭한 악성 도메인이 자주 등장하고 있습니다. 이러한 도메인들은 신뢰할 수 있는 기관으로 가장하여 개인정보를 탈취하거나 금전적 피해를 입히려 합니다. 특히 국제적인 이벤트를 겨냥한 파리올림픽 티켓 사칭 사이트는 대규모 피해를 초래할 수 있는 대표적인 사례입니다.

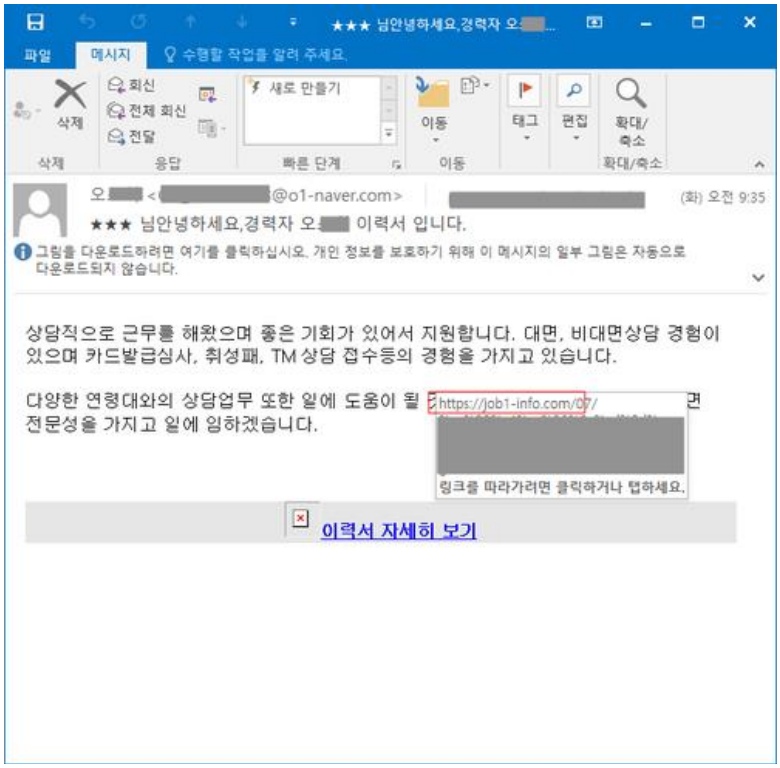


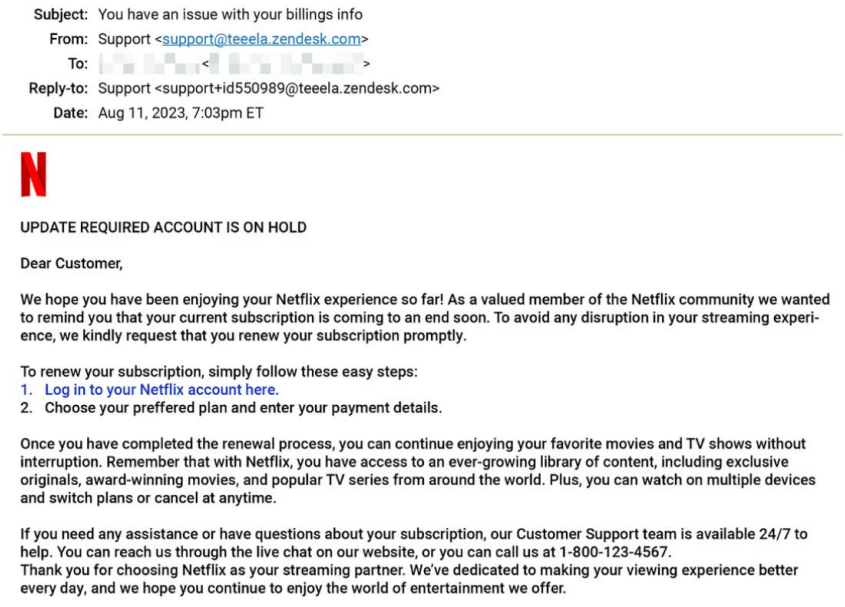


생성형 AI 를 통한 악성 도메인이 포함된 이메일 공격 사례

2024 년, 과학기술정보통신부는 “생성형 AI 를 악용한 사이버 범죄 가능성 증가”를 사이버 보안 위협의 주요 전망 중 하나로 지목하였습니다. 생성형 AI 는 공격자가 쉽게 악성코드를 제작하거나, 피싱 이메일을 전문가가 작성한 것처럼 꾸며 공격의 성공률을 높일 수 있습니다.

실제 사례로, 생성형 AI 로 작성된 이력서 이메일에 악성 도메인이 포함되어 피해를 입힌 사례가 보고된 바 있습니다. 또한, ‘넷플릭스’ 고객 서비스 직원을 사칭하여 구독 만료를 이유로 구독 갱신을 유도하고 악성 도메인으로 연결되는 링크를 포함한 이메일 공격도 발생했습니다. 이처럼 정교하게 설계된 공격은 피해자가 의심 없이 악성 도메인에 접속하게 하여 심각한 피해를 초래할 수 있습니다.





Subject: You have an issue with your billings info
From: Support <support@teeela.zendesk.com>
To: <>
Reply-to: Support <support+id550989@teeela.zendesk.com>
Date: Aug 11, 2023, 7:03pm ET

N

UPDATE REQUIRED ACCOUNT IS ON HOLD

Dear Customer,

We hope you have been enjoying your Netflix experience so far! As a valued member of the Netflix community we wanted to remind you that your current subscription is coming to an end soon. To avoid any disruption in your streaming experience, we kindly request that you renew your subscription promptly.

To renew your subscription, simply follow these easy steps:
1. Log in to your Netflix account here.
2. Choose your preffered plan and enter your payment details.

Once you have completed the renewal process, you can continue enjoying your favorite movies and TV shows without interruption. Remember that with Netflix, you have access to an ever-growing library of content, including exclusive originals, award-winning movies, and popular TV series from around the world. Plus, you can watch on multiple devices and switch plans or cancel at anytime.

If you need any assistance or have questions about your subscription, our Customer Support team is available 24/7 to help. You can reach us through the live chat on our website, or you can call us at 1-800-123-4567.
Thank you for choosing Netflix as your streaming partner. We've dedicated to making your viewing experience better every day, and we hope you continue to enjoy the world of entertainment we offer.

악성 도메인 탐지 기술의 필요성

최근 공격자들은 DGA(Domain Generation Algorithm)를 사용해 수천 개의 악성 도메인을 자동으로 생성하여 DNS 차단을 우회하고, 악성 서버와의 통신을 유지하려는 시도를 지속하고 있습니다. 이 공격 기법은 일회성 도메인이 아닌, 일정한 주기로 새로운 도메인을 생성하여 탐지 회피를 더욱 어렵게 만듭니다. DGA 로 생성된 도메인은 대부분의 정상 도메인과는 다른 패턴을 가지며, 이런 특징은 머신러닝 및 딥러닝 기반의 탐지 모델을 활용해 효과적으로 탐지할 수 있습니다.

DGA 로 생성된 도메인은 특정 어휘적 및 구조적 특징을 가지는데, 이는 정규 도메인과 명확히 구분될 수 있습니다. 아래 표는 DGA 도메인에서 주로 나타나는 특징을 설명하고 있으며, 이 정보를 기반으로 악성 도메인 탐지와 대응 방안을 마련하는 것이 '도비' 프로젝트의 핵심 목표입니다.

DGA 로 생성된 악성 도메인의 주요 특징

분류	특징
길이	URL, hostName, Path, Query 길이
횟수	“.”, “-”, “@”, “_”, “%”, “&”, “#”, 숫자 개수, hostName 의 “-” 개수, subdomain Level, path Level, Query 개수
존재 여부	“~”가 존재하는가, https 인가, ipaddress 형태인가, “//”가 존재하는가

DGA 로 생성된 도메인은 이런 특성을 통해 정상 도메인과 구별될 수 있으며, 머신러닝과 딥러닝 모델을 활용하여 이를 자동으로 탐지할 수 있습니다. 예를 들어, 머신러닝 모델은 URL 의 길이나 특수 문자의 빈도 등을 분석하여 비정상적인 도메인을 식별하고, 딥러닝 모델은 더 복잡한 패턴을 학습해 정확한 탐지를 할 수 있습니다. 이를 바탕으로 '도비' 프로젝트는 DGA 로 생성된 악성 도메인을 신속하게 탐지하고 대응하는 체계를 구축하는 것을 목표로 삼고, 프로젝트를 시작하게 되었습니다.

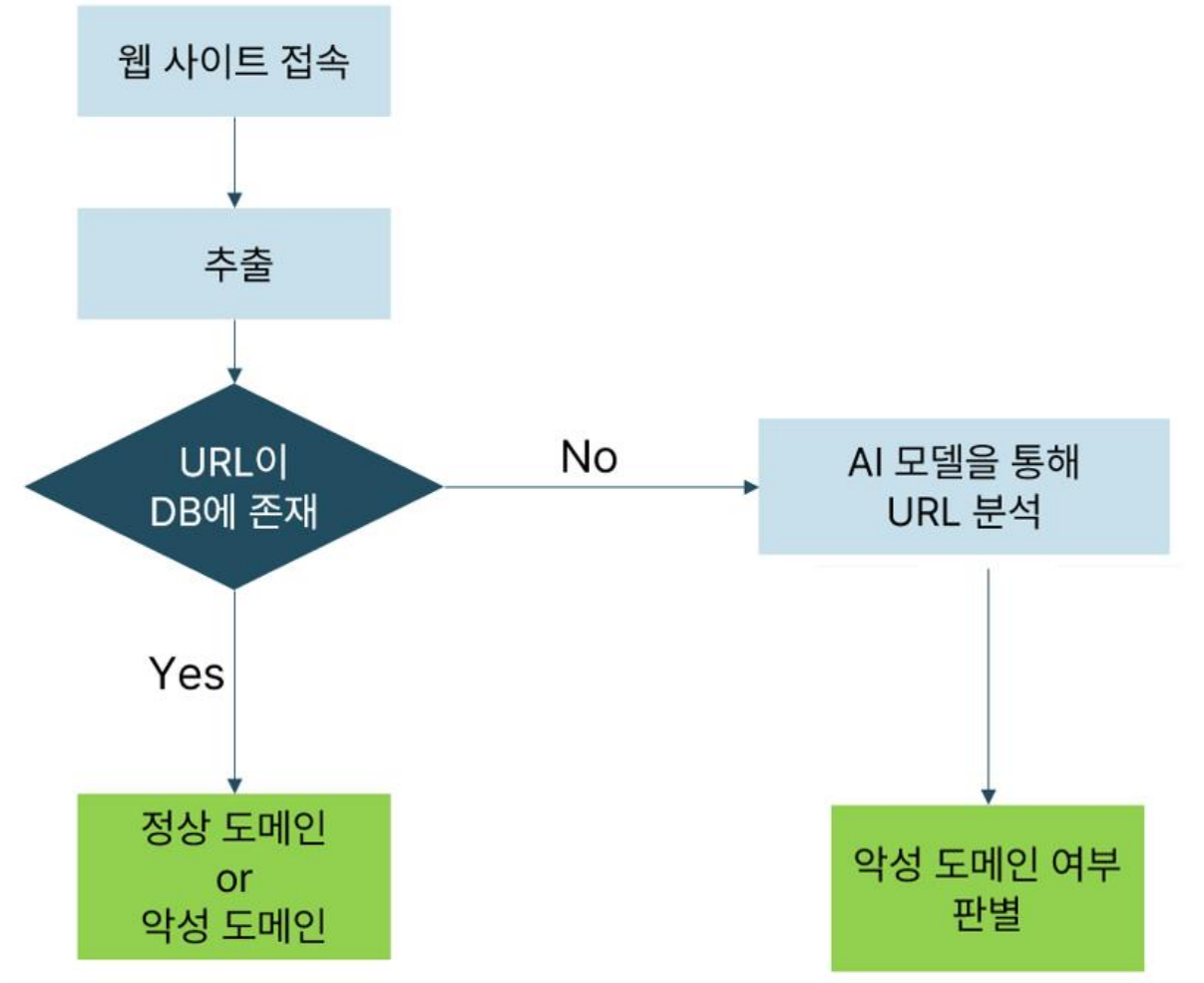
2. 도비 프로젝트

도비 프로젝트는 "악성 **도**메인 탐지 서**비**스" 또는 "**도**메인 **비**서"라는 개념에서 출발하여, 악성 도메인을 빠르고 정확하게 판별하고 탐지하는 경량화된 보안 솔루션입니다. 프로젝트의 첫 단계로 탐지 대상인 악성 도메인의 범위는 다음과 같이 정의하였습니다.

악성 도메인
: 악성 도메인은 인터넷상에서 사용자를 속이거나 악의적인 목적을 달성하기 위한 도메인

도비 프로젝트는 원격 서버에서 연산을 처리하는 방식을 통해 경량화된 보안을 제공합니다. 이 방식은 로컬 시스템의 컴퓨팅 자원 사용을 최소화하면서도, 서버 측에서 복잡한 탐지 파이프라인을 수행하여 빠르고 정확한 악성 도메인 탐지가 가능합니다. 또한, 악성 도메인의 패턴을 학습한 딥러닝 모델을 통해 실시간 탐지가 이루어지며, 이를 통해 네트워크 및 컴퓨팅 성능 저하 없이 안정적인 보안을 유지할 수 있습니다.

악성 도메인 판별 과정



프로젝트 설계

1. 설계 목표

도비 프로젝트를 기획하고 설계하면서 세운 목표는 다음과 같습니다.

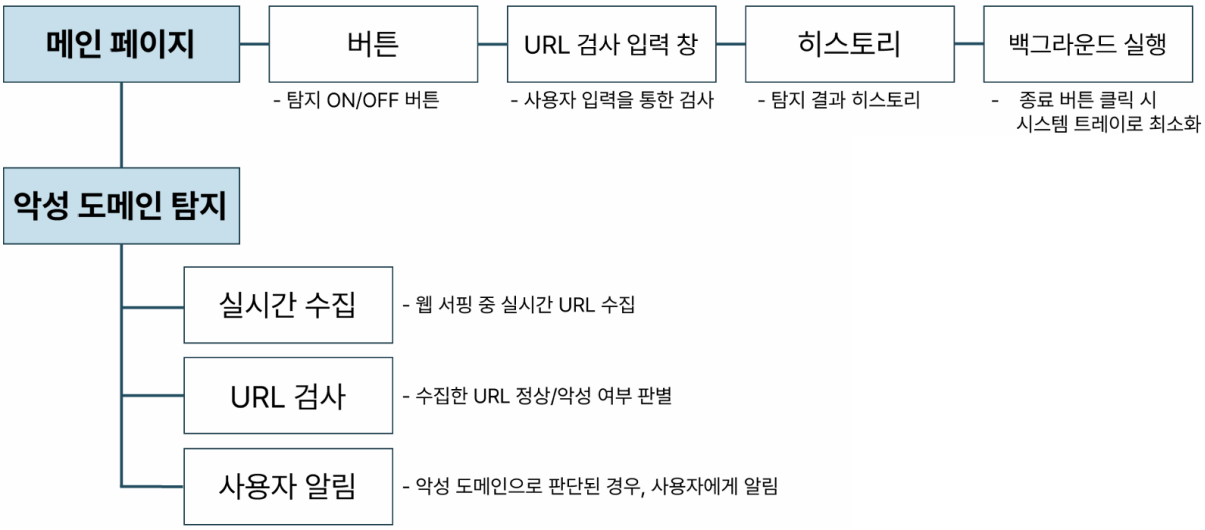
악성 도메인 패턴 학습 및 탐지
딥러닝 기법을 활용해 악성 도메인과 정상 도메인의 구분이 가능한 패턴을 학습하고, 이를 기반으로 악성 도메인을 탐지하는 모델 개발

간편한 접근성
무거운 기존의 상용 보안 솔루션과 달리, 사용자가 쉽게 접근하고 사용할 수 있는 가벼운 서비스 형태로 제공

다양한 플랫폼 확장
차후 데스크탑 환경뿐만 아니라 모바일 앱, 웹 확장 기능 등 다양한 플랫폼으로의 확장을 고려한 설계

2. 기능 설계

도비 프로젝트의 기능 설계는 악성 도메인 탐지 과정의 각 단계를 효과적으로 처리하기 위해 여러 기능의 모듈로 구성되어 있습니다.



주요 기능

- 메인 페이지:

  소개: 도비 서비스의 목적과 주요 기능 소개

  약관: 도비 서비스 사용에 필요한 약관 동의 안내

  사용 방법: 도비 서비스 사용 방법 안내

  업데이트 정보: 최신 버전 업데이트 내용 및 다운로드 링크

  탐지 기능 온/오프: 악성 도메인 탐지 기능 온/오프 스위치

- 악성 도메인 탐지 기능:

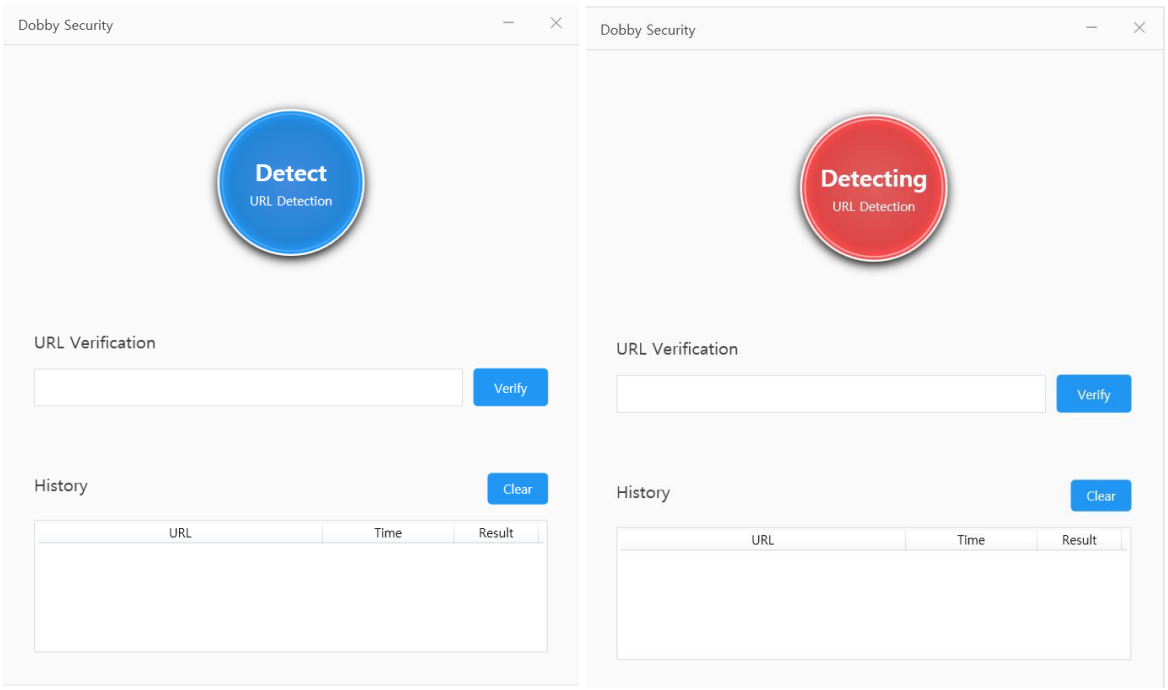
  실시간 수집: 웹 서핑 중 도메인 실시간 수집

  도메인 검사: 수집한 도메인을 검사하여 정상 여부 판별

  사용자 알림: 악성 도메인 판단 시 사용자에게 실시간 알림

3. 기술 설계

데스크톱 애플리케이션



1. URL 캡처 기능

URL 캡처 기능은 사용자가 접속하는 웹사이트의 URL 을 실시간으로 캡처하여 서버로 전송하는 중요한 역할을 수행합니다. 이는 웹페이지가 악성 도메인인지 판단하기 위한 첫 번째 단계로, 보안 시스템의 시작점이 됩니다.

해당 기능은 UI 자동화와 윈도우 이벤트 후킹을 통해 구현되었습니다. UI 자동화는 MS Windows 에서 대부분의 사용자 인터페이스(UI) 요소에 접근할 수 있는 권한을 제공하며, 이는 화면 읽기 프로그램과 같은 보조 기술 제품들이 UI 정보를 제공하고 조작할 수 있게 합니다. 이를 활용해 브라우저의 주소창에 접근하고, URL 이 변경되는 이벤트가 발생하면 이벤트 후킹 메커니즘이 이를 감지하여 해당 URL 문자열을 추출합니다.

윈도우 이벤트 후킹은 특정 윈도우 이벤트가 발생할 때 이를 감지하고 처리하는 강력한 기법으로, 주소창의 URL 변화를 실시간으로 추적할 수 있습니다. 이를 통해 사용자는 안전하게 웹을 이용할 수 있습니다.

2. 서버와 통신 기능

서버와의 통신 기능은 캡처된 URL 데이터를 서버로 전송하고, 서버로부터 악성 도메인 여부를 판단하는 결과를 수신하는 중요한 역할을 합니다. 이를 통해 클라이언트는 실시간으로 도메인의 안전성을 평가받고 필요한 보호 조치를 취할 수 있습니다.

이 기능은 REST API 를 통해 Spring Boot 서버와 통신하며, 클라이언트가 캡처한 URL 데이터를 JSON 형식으로 서버에 전송합니다. 서버는 이를 분석한 후, 도메인 분석 결과를 JSON 형식으로 반환하여 클라이언트에게 도메인의 상태를 전달합니다. 이를 통해 사용자는 실시간으로 악성 도메인 여부를 확인할 수 있습니다.

3. 사용자 인터페이스 (UI)

사용자 인터페이스(UI)는 사용자가 악성 도메인 탐지 기능을 쉽게 활성화하거나 비활성화할 수 있도록 설계되었습니다. 또한, 악성 도메인 탐지가 이루어졌을 때 경고 메시지를 사용자에게 즉시 알리는 기능을 제공합니다. 부가적으로 사용자가 도메인에 접속하기 전에 애플리케이션에 직접 URL 을 입력하여 해당 URL 이 악성인지 판단하는 기능을 제공합니다. 또한 사용자가 Dobby 창을 닫은 상태로도 악성 도메인 탐지를 할 수 있도록 시스템 트레이에 남아 동작하도록 하였습니다. URL 검사 기록은 History 표에 저장됩니다.

UI 는 탐지 기능을 ON/OFF 하는 버튼과 부가 기능을 위한 URL 입력란, URL 검사 결과 기록표, 악성 도메인 탐지 시 경고 메시지를 표시하는 컴포넌트로 구성됩니다. UI 로직과 디자인은 C# WPF 로 구현되었습니다.

메인 서버

1. 데스크톱 애플리케이션과의 통신

메인 서버는 데스크톱 애플리케이션에서 전달받은 URL 을 분석하여, 해당 도메인이 악성인지 여부를 판단합니다. 만약 분석 결과 악성 도메인으로 확인되면, 서버는 데스크톱 애플리케이션에 경고 신호를 보내고, 애플리케이션은 사용자에게 경고 메시지를 표시합니다.

이 과정은 REST API 를 통해 이루어지며, 서버는 수신된 URL 을 처리한 후, 결과를 boolean 값으로 반환합니다. 이 통신 과정의 안정성과 효율성을 높이기 위해 RESTful API 설계를 최적화할 예정입니다.

2. DB 조회 기능

메인 서버는 데스크톱 애플리케이션에서 수신된 URL 을 데이터베이스에서 조회하여 해당 도메인이 정상인지 여부를 확인합니다. 시스템의 조회 성능을 극대화하기 위해 Bloom Filter 와 Redis DBMS 가 함께 사용됩니다.

Bloom Filter 는 URL 이 DB 에 존재할 가능성을 빠르게 확인하는 역할을 수행합니다. 이는 전체 DB 를 조회하지 않고도 URL 이 존재할 가능성을 미리 추정할 수 있어, 불필요한 조회 과정을 줄이고 성능을 크게 향상시킵니다. 만약 Bloom Filter 가 해당 URL 이 DB 에 존재할 가능성이 있다고 판단하면, 서버는 Redis 에서 실제 데이터를 조회하여 도메인이 정상인지, 악성인지를 판별합니다.

Redis 는 메모리 기반의 비관계형 데이터베이스로, 매우 빠른 조회 속도를 자랑합니다. Key-Value 구조를 활용해 대량의 데이터 조회도 고성능으로 처리할 수 있으며, 인메모리 방식 덕분에 실시간 성능이 중요한 악성 도메인 탐지 시스템에서 매우 효과적입니다.

만약 Bloom Filter 가 URL 이 DB 에 존재하지 않는다고 판단할 경우, 메인 서버는 해당 URL 을 AI 모델에 전달하여 분석을 진행합니다. 분석 결과는 Redis 에 캐싱되어, 이후 동일한 URL 이 조회될 때 더욱 빠르게 처리할 수 있습니다. 이를 통해 시스템의 전체적인 성능과 응답 시간을 최적화할 수 있습니다.

3. AI 모델과의 통신

메인 서버는 gRPC 를 통해 AI 모델과 통신하여 URL 의 악성 여부를 분석합니다. AI 모델은 Python 으로 구현되어 있으며, 메인 서버는 Java 기반의 gRPC 클라이언트를 사용하여 AI 모델과 연결됩니다.

Spring Boot 에서 구현한 gRPC 클라이언트를 통해, 서버는 URL 을 AI 모델에 전달하고, 분석 결과를 boolean 값으로 반환받습니다. AI 모델은 URL 의 특성을 분석하여, 해당 URL 이 정상인지 또는 악성인지 판단합니다.

악성 도메인 탐지 AI 모델

도비에서 인공지능은 악성 도메인을 탐지하는 중요한 역할을 담당합니다. 딥러닝을 이용한 도메인 탐지 인공지능의 학습은 다음과 같은 단계를 거쳐 이루어졌습니다:

1. 데이터 탐색

도메인 탐지를 위한 초기 데이터셋은 Kaggle 에서 제공된 오픈소스 데이터셋을 사용했습니다. 이 데이터셋에는 총 651,191 개의 URL 이 포함되어 있으며, 정상 도메인과 악성 도메인을 구분하는 데 필요한 다양한 정보가 포함되어 있습니다.

특징 추출 단계에서는 URL 의 구조적 요소를 분석하여, 도메인의 문자 구성, 특수 문자, 숫자, 도메인 길이 등을 기반으로 주요 특징들을 정량화 하였습니다. 이를 통해 딥러닝 모델이 학습할 수 있는 형태로 데이터를 가공하였습니다.

getLength	Url의 길이
specialCount	특수문자의 개수
abnormalUrl	Url이 해당 도메인을 포함하는지
httpSecure	Https / Http 유무
digitCount	숫자의 개수
digitLetter	문자의 개수
shorteningService	단축 Url 유무 확인
ipAddress	ip 주소 확인

2. 딥러닝 학습 및 평가

초기 데이터를 활용하여 딥러닝 접근 방식을 도입하였습니다. CNN(Convolutional Neural Network)과 RNN(Recurrent Neural Network) 아키텍처를 테스트한 결과, CNN 아키텍처가 더 우수한 성능을 보였으며, 특히 CNN1 모델이 빠른 학습 속도와 높은 정확도를 달성했습니다.

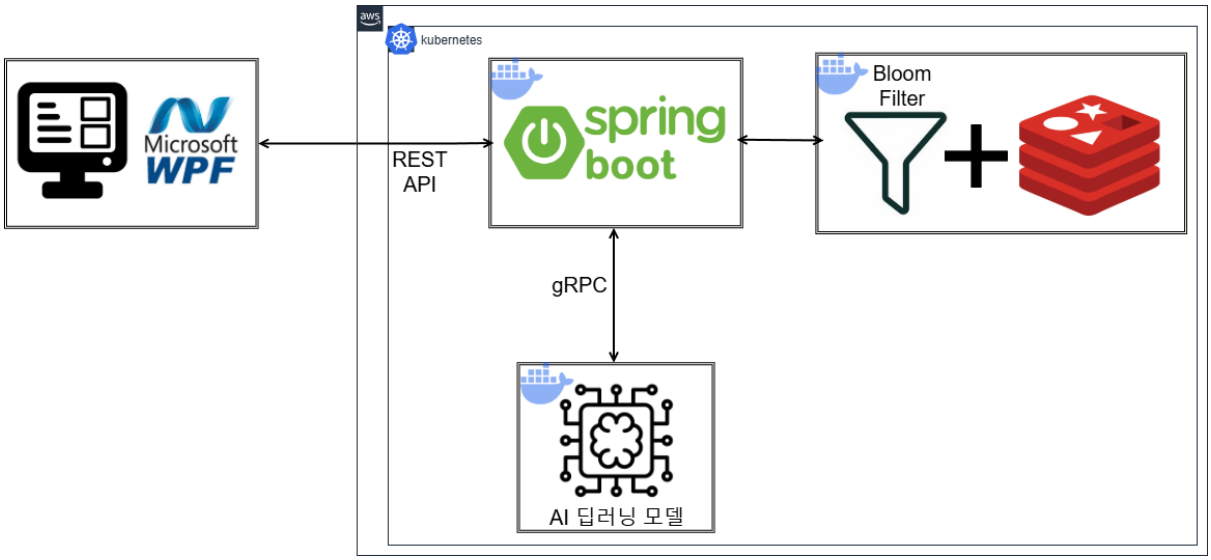
Arch.	Batch Size	Train Acc.	Val. Acc.	Test Acc.	Avg. Train Time (sec)
RNN	3000	0.962	0.962	0.963	709.39
BRNN	Cannot be performed (due to need of huge GPU memory)				
CNN	7000	**0.973**	**0.976**	**0.975**	**10.75**
ANN	7000	0.911	0.909	0.910	11.48
ATT	Cannot be performed (due to need of huge GPU memory)				

Arch.	Val. Acc.	Val Loss	Test Acc.	Test Loss	Avg. Train Time (sec)
CNN1	**0.981**	**0.051**	**0.980**	**0.051**	**10.9**
CNN2	0.976	0.064	0.976	0.063	22.0
CNN3	0.977	0.086	0.977	0.085	17.6

최종적으로, 학습에 사용된 데이터셋은 PhishTank 에서 수집된 약 2,300,000 개의 악성 도메인과, 검색 엔진을 통해 수집한 2,800,000 개의 정상 도메인으로 구성되었습니다. 이러한 대규모 데이터셋을 기반으로 딥러닝 모델은 더욱 정교하게 악성 도메인을 탐지할 수 있게 되었으며, 이를 통해 성능이 크게 향상되었습니다.

4. 시스템 설계

도비 프로젝트는 사용자의 웹 브라우징을 안전하게 보호하기 위해 설계된 보안 솔루션입니다. 이 시스템은 데스크톱 애플리케이션과 메인 서버, 그리고 AI 모델로 구성되어 있으며, 실시간으로 악성 도메인을 탐지하고 사용자에게 경고를 제공합니다.



1. 데스크톱 애플리케이션 데이터 수집

- 사용자가 웹사이트에 접속하면, 데스크톱 애플리케이션이 실시간으로 해당 사이트의 URL 을 캡처합니다.
- 캡처된 URL 데이터는 REST API 를 통해 메인 서버로 전송됩니다.

2. 메인 서버 데이터 처리

- 메인 서버는 수신된 URL 을 먼저 내부 DB 에서 조회하여, 해당 도메인의 정상/악성 여부를 확인합니다.
- 만약 DB 에서 URL 이 조회되지 않거나 불확실한 경우, URL 데이터를 AI 모델에 전송하여 심층 분석을 요청합니다.

3. AI 모델 분석

- AI 모델은 메인 서버에서 전송된 URL 데이터를 딥러닝 기반 알고리즘을 통해 실시간으로 분석합니다.
- 모델은 악성 도메인 탐지를 위한 다양한 특징을 학습한 상태이며, URL 의 악성 여부를 판단하여 boolean 값(True/False) 형태로 메인 서버에 결과를 반환합니다.

4. 메인 서버 최종 판단

- 메인 서버는 AI 모델로부터 반환된 분석 결과를 종합하여 해당 도메인의 최종 악성 여부를 판단합니다.
- 만약 도메인이 악성으로 판단될 경우, 메인 서버는 즉시 데스크톱 애플리케이션으로 경고 신호를 전송하여 사용자에게 경고 메시지를 표시합니다.
- 사용자는 실시간으로 악성 도메인 경고를 확인하며, 이를 통해 보안 위협으로부터 보호받을 수 있습니다.
- AI 모델로부터 반환된 결과는 DB 에 캐싱되어, 이후 동일한 도메인이 요청될 때 더 빠르게 응답할 수 있도록 처리됩니다. 이를 통해 AI 모델에 대한 추가 요청을 줄이고, 시스템 성능을 최적화합니다.

프로젝트 책임 영역

1. 프로젝트 책임 영역

도비 프로젝트에서 저의 주요 역할은 팀장으로서 프로젝트의 전반적인 방향을 설정하고, 팀원들이 각자의 역할을 효율적으로 수행할 수 있도록 협업을 조율하는 것이었습니다. 프로젝트 초기에는 제가 직접 악성 도메인 탐지라는 아이디어를 제안하고, 이를 기반으로 팀을 구성하였습니다. 그 이후, 기획, 백엔드 아키텍처 설계 및 개발, CI/CD 환경 설정 등 프로젝트의 핵심적인 부분을 총괄하며, 프로젝트의 성공적인 진행을 위해 다음과 같은 역할을 수행했습니다.

Ⅰ. 기획

악성 도메인 탐지라는 주제를 기반으로 프로젝트의 기획을 담당하였으며, 이를 통해 프로젝트의 방향성을 설정하였습니다. 프로젝트 제안, 요구사항 정의 등을 구성하고 문서화하며 프로젝트의 기본 틀을 잡았고, 이를 바탕으로 팀원들과의 협업이 원활하게 이루어질 수 있도록 기술적 접근 방식을 정의했습니다. 팀원들과의 지속적인 소통을 통해 각 구성원의 역할 분배를 체계적으로 수행하였고, 목표와 일정을 명확히 설정하여 프로젝트의 성공적인 진행을 도모했습니다.

Ⅱ. 백엔드 아키텍처 설계 및 개발

도비 프로젝트에서 핵심적인 역할은 백엔드 아키텍처의 설계 및 구축이었습니다. 이중 서버 구조를 채택하여, 주 서버는 Spring Boot 로 구축하고, 보조 서버는 gRPC 를 통해 각기 다른 기능을 담당하도록 하였습니다. 이러한 구조는 마이크로서비스 아키텍처의 장점을 최대한 활용할 수 있도록 설계되었으며, 확장성과 유지보수가 용이하게 구현되었습니다.

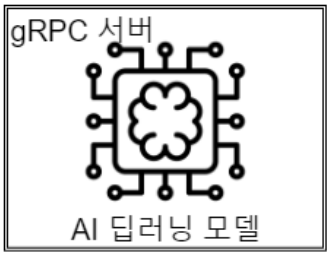
특히, Bloom Filter 와 Redis DBMS 를 추가하여 데이터베이스 조회 성능을 극대화했습니다. Bloom Filter 는 특정 URL 이 DB 에 존재할 가능성을 매우 빠르게 확인해, 불필요한 전체 조회를 줄이고 성능을 최적화하였습니다. 이를 통해 DB 에 부하를 주지 않으면서 빠른 조회가 가능해졌습니다.

1. Spring Boot 서버

Spring Boot 는 자바 기반의 프레임워크로, 빠르고 간편한 설정을 통해 백엔드 서비스를 효율적으로 개발할 수 있습니다. 도비 프로젝트에서 Spring Boot 를 사용한 이유는 다음과 같습니다.

- 간편한 설정: Spring Boot 는 복잡한 XML 설정을 최소화하고, 필요한 의존성을 자동으로 구성하는 방식으로 개발자가 빠르게 서버를 설정할 수 있도록 지원합니다. 도비 프로젝트에서는 개발 속도를 높이는 데 이 기능이 매우 유용했습니다.

- 확장성과 모듈화: Spring Boot 는 마이크로 서비스 아키텍처에 적합하며, 각 기능을 모듈화하여 쉽게 확장할 수 있는 유연성을 제공합니다. 이점 덕분에 도비 프로젝트에서는 다양한 기능을 개별 모듈로 분리하여 관리함으로써, 서비스 확장 시 유연하게 대응할 수 있었습니다.



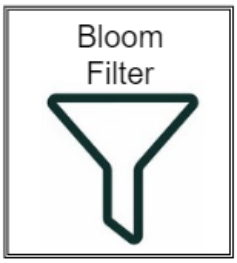
2. gRPC 서버

gRPC 는 구글에서 개발한 고성능 원격 프로시저 호출(Remote Procedure Call) 프레임워크로, 여러 프로그래밍 언어 간의 통신을 효율적으로 지원합니다. 도비 프로젝트에서 gRPC 를 사용한 이유는 다음과 같습니다.

- 고성능: gRPC 는 HTTP/2 프로토콜을 기반으로 하여 효율적인 네트워크 자원 사용을 가능하게 합니다. 특히 병렬 스트리밍이 가능해, 서버와 클라이언트 간 대용량 데이터를 빠르게 처리할 수 있습니다. 도비 프로젝트에서는 실시간 악성 도메인 탐지를 위해 빠른 데이터 전송이 필수적이었기 때문에 gRPC 의 성능적 이점이 크게 작용했습니다.

- 다양한 언어 지원: gRPC 는 여러 프로그래밍 언어 간의 호환성을 지원합니다. 도비 프로젝트의 AI 모델은 Python 으로 구축되었고, 메인 서버는 Java 로 개발되었기 때문에 gRPC 는 이 두 언어 간의 원활한 통신을 가능하게 했습니다.

- 프로토콜 버퍼: gRPC 는 데이터 직렬화에 Protocol Buffers 를 사용하여, JSON 이나 XML 보다 훨씬 가볍고 빠르게 데이터를 처리할 수 있습니다. 이를 통해 대규모 데이터 전송에도 효율적인 처리 속도를 제공할 수 있었습니다.



3. Bloom Filter

Bloom Filter 는 확률적 자료 구조로, 메모리 효율성을 극대화하면서도 특정 데이터가 존재할 가능성을 빠르게 판별할 수 있습니다. 이를 통해 DB 조회의 성능을 최적화하였습니다. 도비 프로젝트에서 Bloom Filter 를 사용한 이유는 다음과 같습니다.

- 빠른 조회 가능성 판단: Bloom Filter 는 특정 URL 이 데이터베이스에 존재할 가능성을 매우 빠르게 확인할 수 있습니다. 이를 통해 DB 전체를 조회할 필요 없이, 존재하지 않을 가능성이 큰 경우 추가적인 조회를 피하고 성능을 크게 향상시킬 수 있습니다.

- 메모리 효율성: Bloom Filter 는 매우 적은 메모리만 사용하여 대규모 데이터를 처리할 수 있습니다. 이는 데이터가 많을수록 조회 성능에 부담을 줄일 수 있어, 도비 프로젝트에서 대규모 악성 도메인 데이터 처리를 가능하게 했습니다.

4. DBMS

Redis DBMS 는 인메모리 데이터베이스로, 도비 프로젝트에서 실시간으로 URL 조회 및 캐시 데이터를 저장하는 데 사용되었습니다. Redis 를 사용한 이유는 다음과 같습니다.

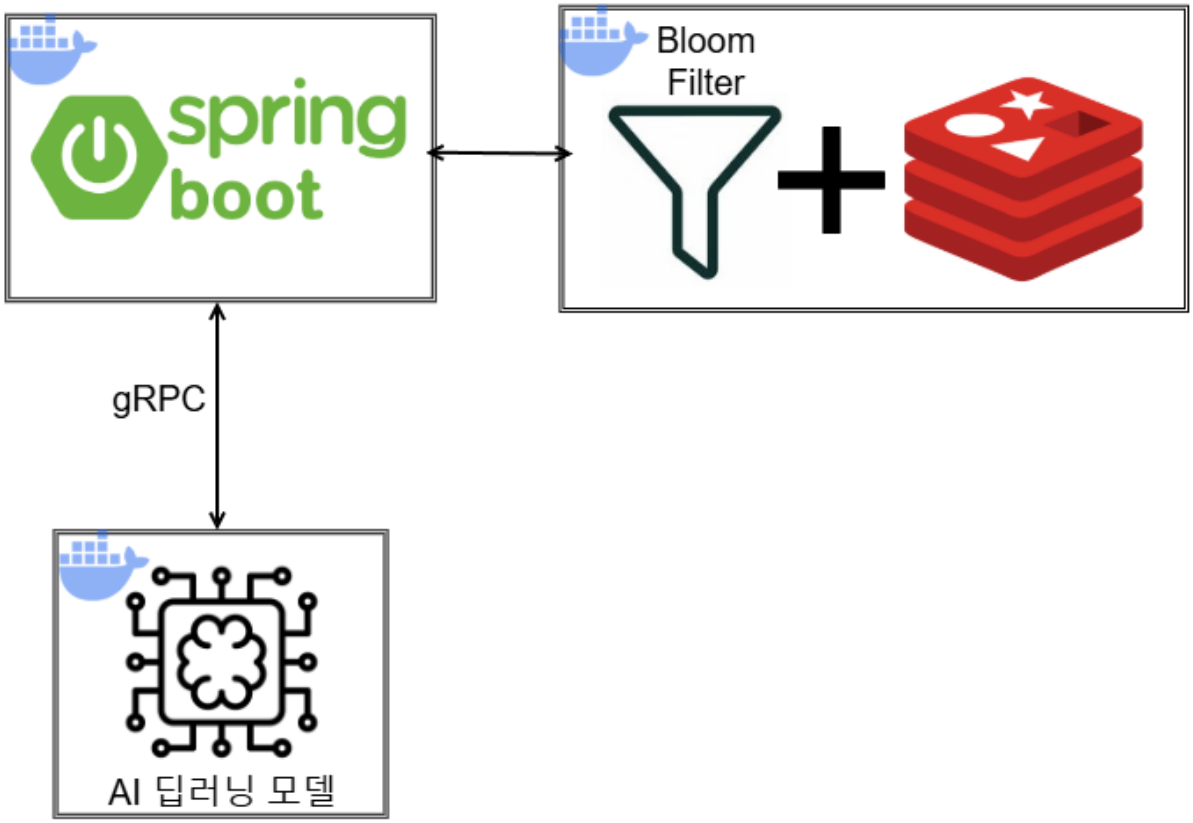
- 빠른 성능: Redis 는 데이터를 메모리 상에서 처리하기 때문에, 읽기/쓰기 성능이 매우 빠릅니다. 이는 도비 프로젝트에서 실시간으로 악성 도메인을 탐지하고 빠르게 응답해야 하는 요구 사항에 부합했습니다.

- 캐시 기능: Redis 는 캐싱 시스템으로서도 매우 강력한 기능을 제공합니다. 악성 도메인 탐지 결과를 캐시해 두면, 이후 동일한 URL 을 다시 조회할 때 Redis 에서 바로 결과를 반환할 수 있어, AI 모델 호출 빈도를 줄이고 성능을 향상시킬 수 있었습니다.

- 데이터 구조의 유연성: Redis 는 단순한 키-값 저장뿐만 아니라, 다양한 데이터 구조(리스트, 셋, 해시 등)를 지원하여 복잡한 데이터 쿼리를 효율적으로 처리할 수 있었습니다. 이를 통해 대규모 URL 데이터에 대한 효율적인 조회 및 관리가 가능했습니다.

III. CI/CD 환경 설정

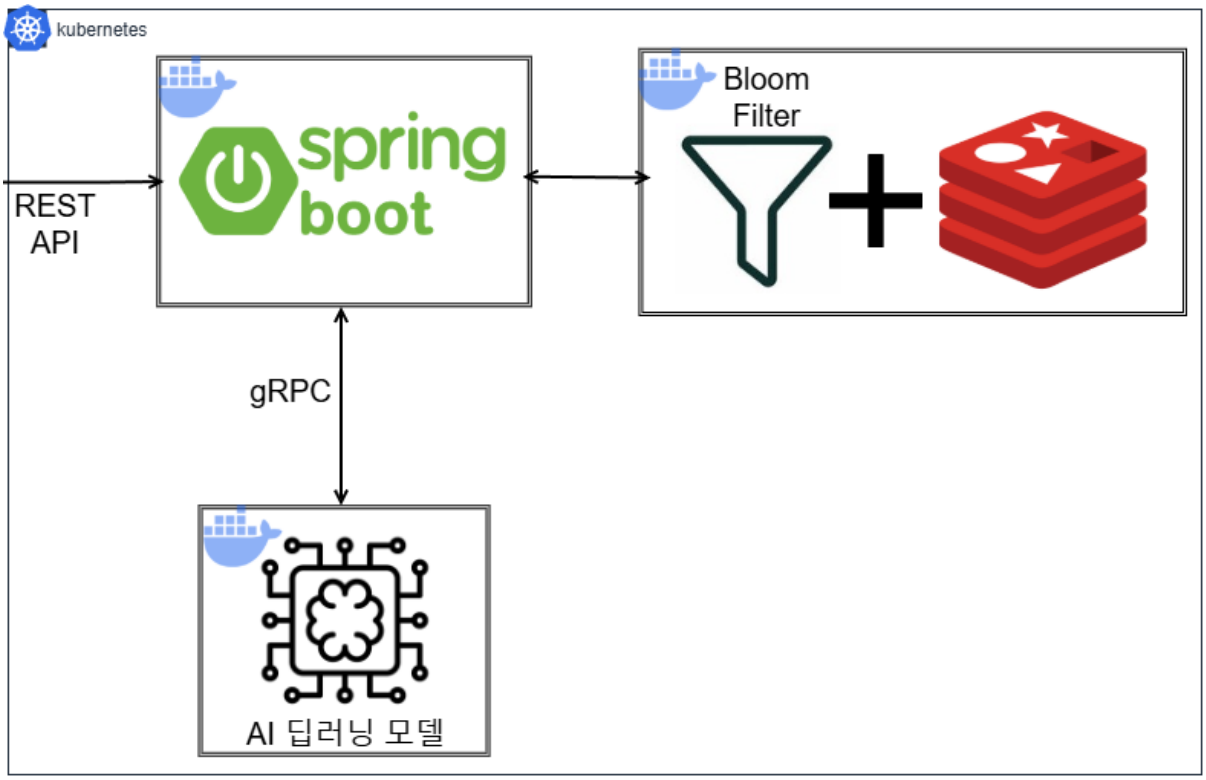
도비 프로젝트의 지속적인 통합 및 배포를 자동화하기 위해 Docker 와 Kubernetes 를 활용한 CI/CD 환경을 구축하였습니다. 이를 통해 개발, 테스트, 배포 과정을 효율적으로 관리하고, 팀원들이 일관된 환경에서 작업할 수 있도록 보장하였습니다.



1. Docker

Docker 는 애플리케이션을 컨테이너화하여 운영 환경 간의 차이를 최소화하고, 일관성 있는 배포를 가능하게 하는 도구입니다. 도비 프로젝트에서는 개발과 배포 과정에서 일관성 유지를 위해 Docker 를 사용하였습니다. Docker 를 사용한 이유는 다음과 같습니다.

- 이식성: Docker 컨테이너는 애플리케이션과 그 실행 환경을 패키징하여, 어디서나 동일한 환경에서 실행될 수 있게 합니다. 이 덕분에 도비 프로젝트는 로컬 개발 환경과 서버 배포 환경 간의 차이를 최소화할 수 있었습니다.
- 자원 격리: 컨테이너는 각기 독립된 환경에서 실행되므로, 하나의 서버에서 여러 애플리케이션이 독립적으로 실행되도록 할 수 있습니다. 이를 통해 도비 프로젝트에서 여러 서버 인스턴스를 효율적으로 관리할 수 있었습니다.



2. Kubernetes

Kubernetes 는 컨테이너 오케스트레이션 도구로, 대규모 애플리케이션의 배포와 관리를 자동화하는 역할을 합니다. 도비 프로젝트에서는 Kubernetes 를 통해 여러 컨테이너를 효율적으로 관리하고, 자동 스케일링 및 부하 분산을 통해 안정적인 운영을 보장하였습니다. Kubernetes 를 사용한 이유는 다음과 같습니다.

- 자동 스케일링: Kubernetes 는 트래픽에 따라 자동으로 컨테이너 수를 조정할 수 있어, 사용량이 많은 경우에는 자동으로 리소스를 확장하고, 트래픽이 적을 때는 축소할 수 있습니다. 이는 도비 프로젝트에서 변동하는 트래픽에 효과적으로 대응할 수 있도록 도와주었습니다.

- 자체 복구 기능: Kubernetes 는 장애가 발생한 컨테이너를 자동으로 감지하고 복구하며, 서비스 중단 없이 애플리케이션을 유지합니다. 이로 인해 도비 프로젝트에서 안정적인 운영을 보장할 수 있었습니다.

- 부하 분산: Kubernetes 는 여러 노드에 걸쳐 부하를 분산하여 리소스 사용의 최적화를 지원합니다. 이를 통해 도비 프로젝트는 서버 자원을 효율적으로 사용하면서도 고가용성을 유지할 수 있었습니다.

프로젝트 기대 효과

1. 프로젝트 기대 효과

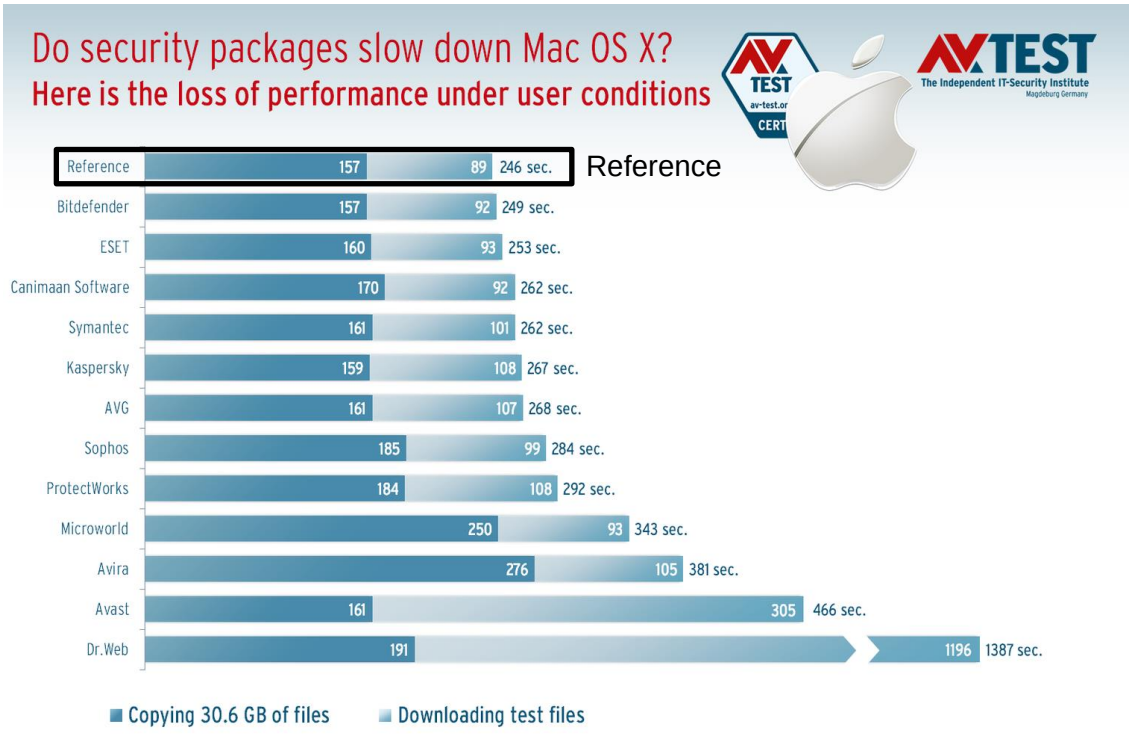
도비 프로젝트는 악성 도메인 탐지를 위한 경량화된 보안 솔루션을 제공하여 사용자들에게 안전한 웹 환경을 조성하는 데 기여할 것으로 기대됩니다. 도비 프로젝트가 달성하고자 하는 주요 기대 효과는 다음과 같습니다.

실시간 악성 도메인 탐지 기능 강화

도비 프로젝트는 실시간으로 악성 도메인을 탐지하고 사용자에게 경고를 제공하여, 피싱이나 악성 코드 배포와 같은 사이버 공격으로부터 사용자를 보호할 수 있습니다. 이 기능은 Bloom Filter 와 Redis DBMS 를 활용한 빠른 데이터 조회를 통해 신속한 탐지를 가능하게 하고, AI 모델의 학습 결과를 기반으로 도메인의 상태를 판단하여 정확도를 높입니다. 이를 통해 악성 도메인에 대한 탐지 속도와 정확성을 향상시켜 보안 위협에 대한 즉각적인 대응을 가능하게 합니다.

네트워크 및 시스템 성능 유지

기존의 상용 보안 솔루션은 시스템 자원을 과도하게 소비하여 네트워크 성능이나 컴퓨팅 성능을 저하시키는 문제가 있었습니다. 아래는 바이러스 백신 및 보안 제품군 소프트웨어를 평가하고 등급을 매기는 독립 기관 AV-TEST 에서 측정한 안티 바이러스 프로그램별 속도 저하 수치 통계입니다.



도비 프로젝트는 서버에서 주요 연산을 처리하는 구조로 설계되었으며, 경량화된 방식으로 로컬 시스템에 최소한의 부하만 주도록 최적화되었습니다. 특히, Redis 캐시와 AI 모델 기반의 데이터 분석을 활용해 빠르고 효율적인 탐지를 가능하게 함으로써 네트워크 성능을 유지하면서도 높은 수준의 보안을 제공합니다.

다양한 플랫폼 확장 가능성

도비 프로젝트는 모듈화된 구조를 기반으로 설계되었기 때문에, 향후 모바일 앱, 웹 브라우저 확장 프로그램 등 다양한 플랫폼으로 확장될 수 있는 가능성을 갖추고 있습니다. 이를 통해 데스크톱뿐만 아니라 다양한 디바이스 환경에서도 동일한 보안 서비스를 제공할 수 있습니다. 또한, 확장성과 유지보수의 용이함을 갖춘 마이크로서비스 아키텍처 덕분에 서비스 확장이 유연하고 효율적으로 이루어질 수 있습니다.

보안 의식 향상

도비 프로젝트는 사용자가 일상적으로 접속하는 웹사이트에서 발생할 수 있는 위험성을 경고함으로써, 사용자의 보안 의식을 높일 수 있습니다. 사용자는 도메인 탐지 기능을 통해 악성 도메인에 대한 경각심을 가지게 되고, 웹 서핑 중 보안 위협에 대한 실시간 알림을 통해 더 안전하게 인터넷을 사용할 수 있습니다.

사이버 범죄 예방 기여

도비 프로젝트는 악성 도메인 탐지를 통해 사이버 범죄 예방에 중요한 역할을 할 것으로 기대됩니다. 최근 급증하고 있는 피싱 및 악성 코드 배포와 같은 공격에서 사용자를 보호하며, 특히 DGA(Domain Generation Algorithm)로 생성된 악성 도메인과 같은 고급 공격 기법에 대해 딥러닝 모델을 활용해 보다 정교한 탐지를 수행합니다. 이를 통해 사회적, 경제적 피해를 예방하고, 기업 및 개인의 정보 보안을 강화하는 데 기여할 수 있습니다.