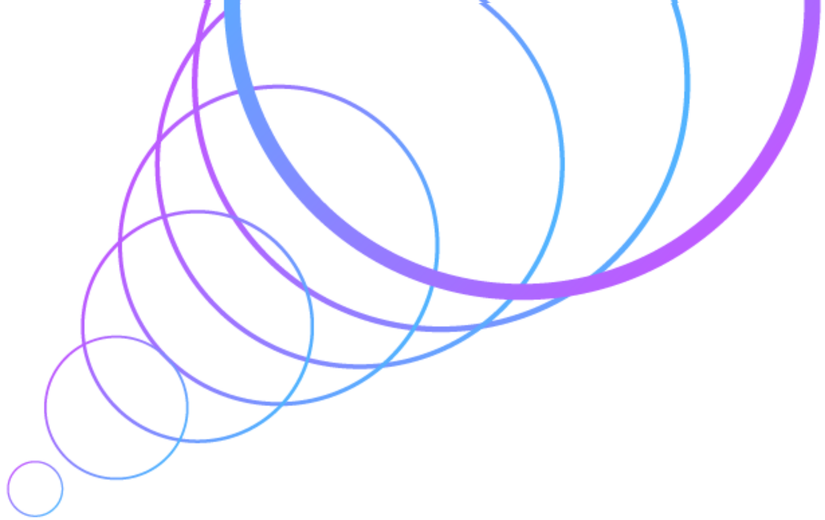


Whale 피싱과 전쟁 중

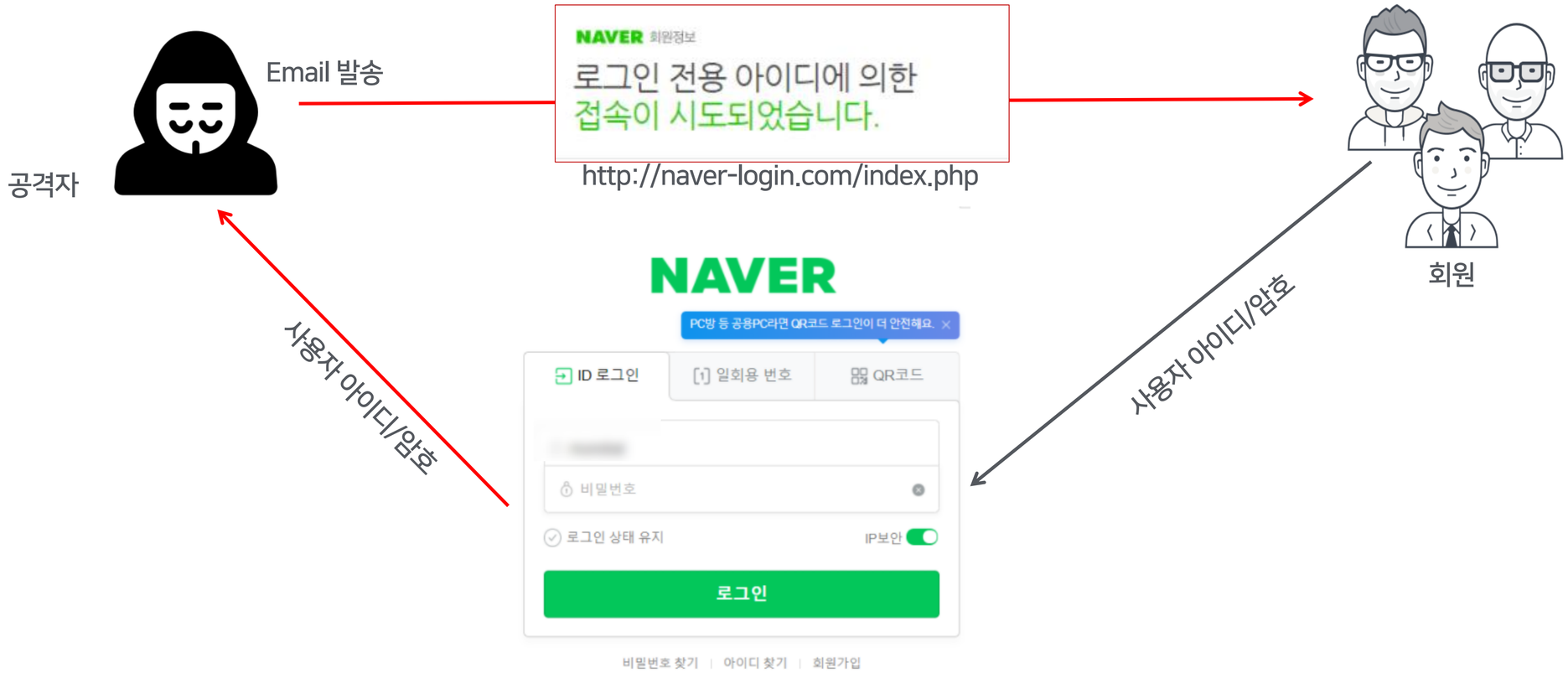
이민수, 최병운 NAVER Security & Whale

CONTENTS

1.피싱과 Whale

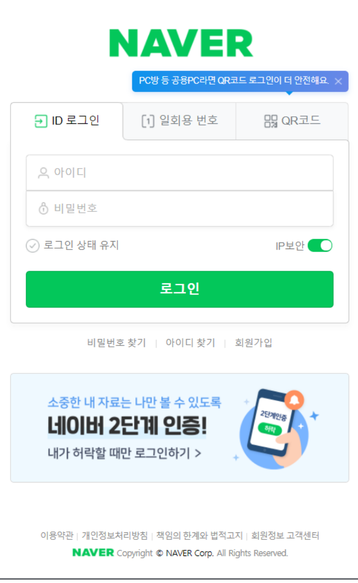
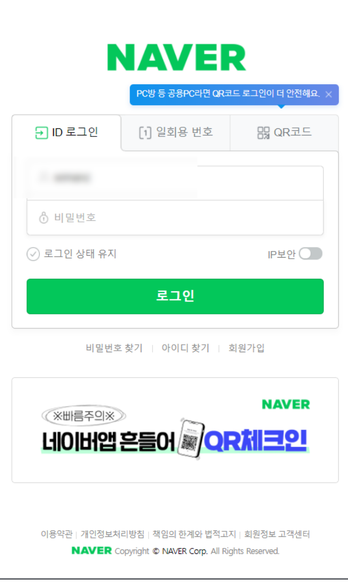
1.1 피싱이란?

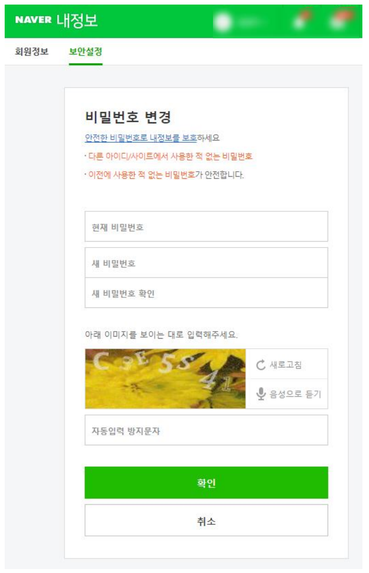
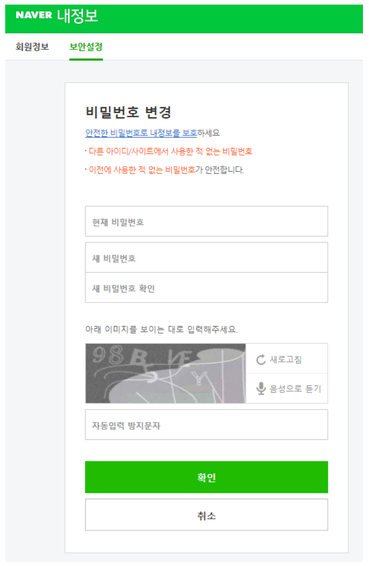
특정 서비스와 유사한 웹 페이지를 만들어 사용자에게 배포함으로서
중요 정보를 훔쳐가는 사기 수법



1.2 피싱 특성은?

사용자가 정상 서비스를 이용하는 착각을 일으키도록 정상 사이트와 유사한 URL, 동일한 UI 및 컨텐츠를 사용

정상 사이트	피싱 사이트
https://nid.naver.com/nid..	https://nid.naevear.com/nid...
	

정상 사이트	피싱 사이트
https://nid.naver.com/user2...	https://nid.naverhelp.com.co/...
	

※일부 피싱의 경우 이러한 규칙이 적용되지 않는 사례들도 있음

1.3 발전하는 피싱

더 정교하게 (정성스레)

- 피싱 전용 도메인 사용 - 도메인 구매, 서버 인증서 등록
- 최신 UI 반영 – 네이버 로그인 디자인 변경 후 바로 적용
- 자동화된 수집 방지 기술 적용

공격 대상을 좁혀서

- 불특정 다수 공격에서 특정 일부인원 대상 공격

짧은 시간만 공격 시도

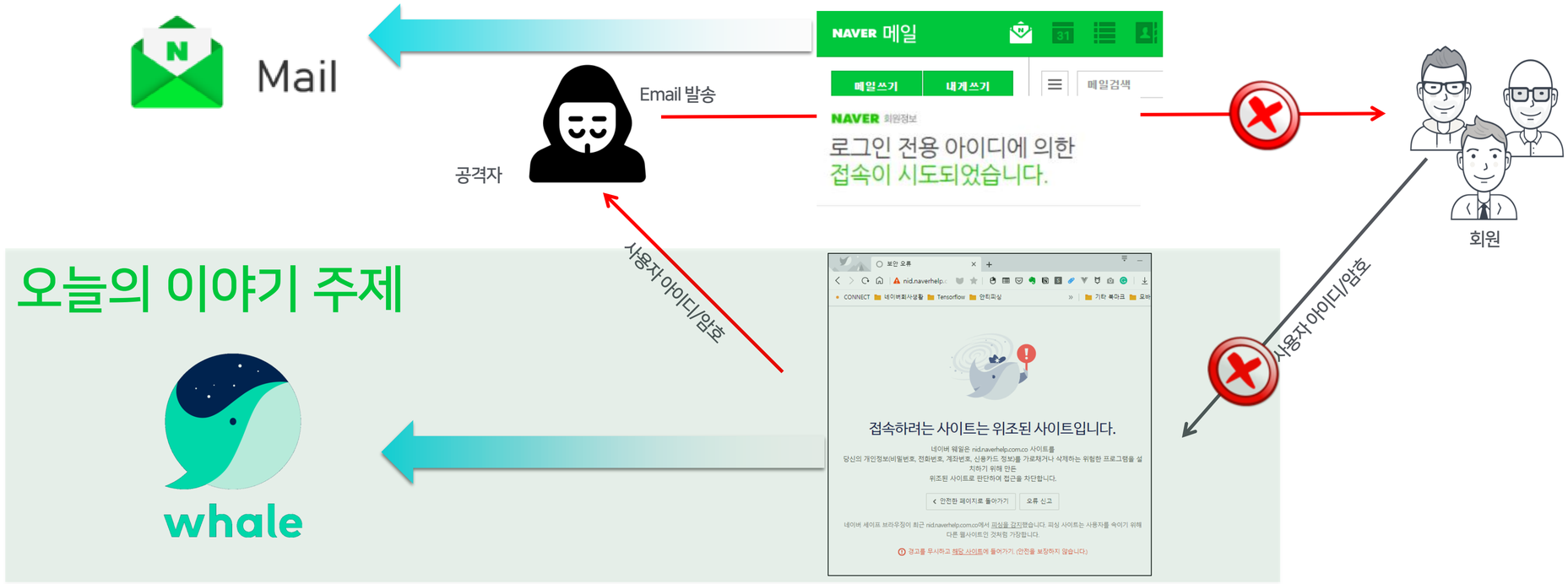
- 피싱 URL에 접근 가능 시간을 제한

탐지 기술 우회

1.3 피싱 대응 가능 지점

피싱 대응에 적합한 시점(지점)은 공격 시나리오에서 2곳

- 피싱 메일을 즉시 스팸 처리 – 메일 시스템
- 피싱 페이지에 접속 차단 – 웹 브라우저



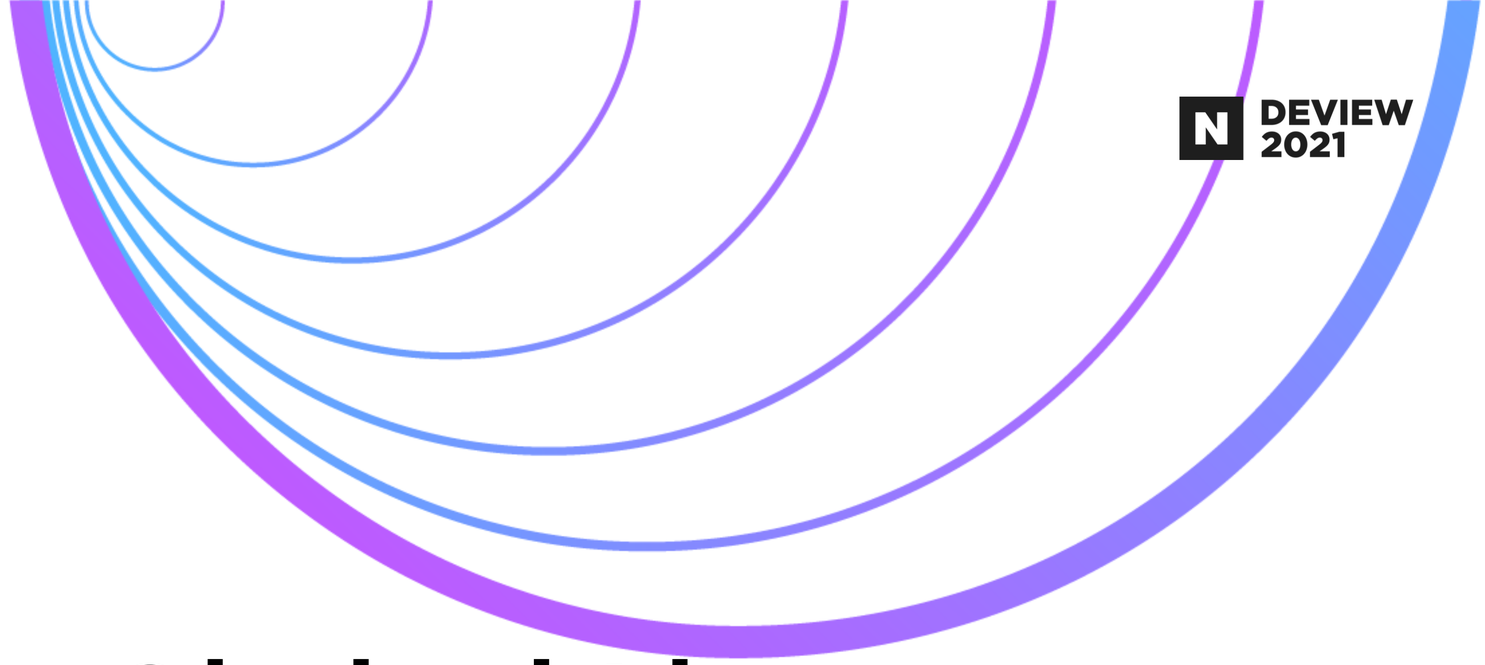
1.4 Whale은 피싱 방화벽(??)

웹 브라우저가 피싱 탐지에 좋은 이유

- 피싱이 사용자에게 전달되는 관문으로 반드시 지나게 되는 통로
- 사용자 도구로서 행동 패턴 범위에 있어 피싱 정보 수집에 용이
- 사용자 측에서 대응 가능해 탐지에 필요한 자원 소모가 적음

Whale은 여러가지 피싱 탐지 기술들을 탑재한 피싱 방화벽

- 외부 제공 피싱 DB 탑재 (예, Open Phish, APWG 등)
- Naver 피싱 DB 탑재 : 자체 수집 피싱 데이터
- Zero-day 피싱 탐지 가능한 Whale CSD 모듈 탑재

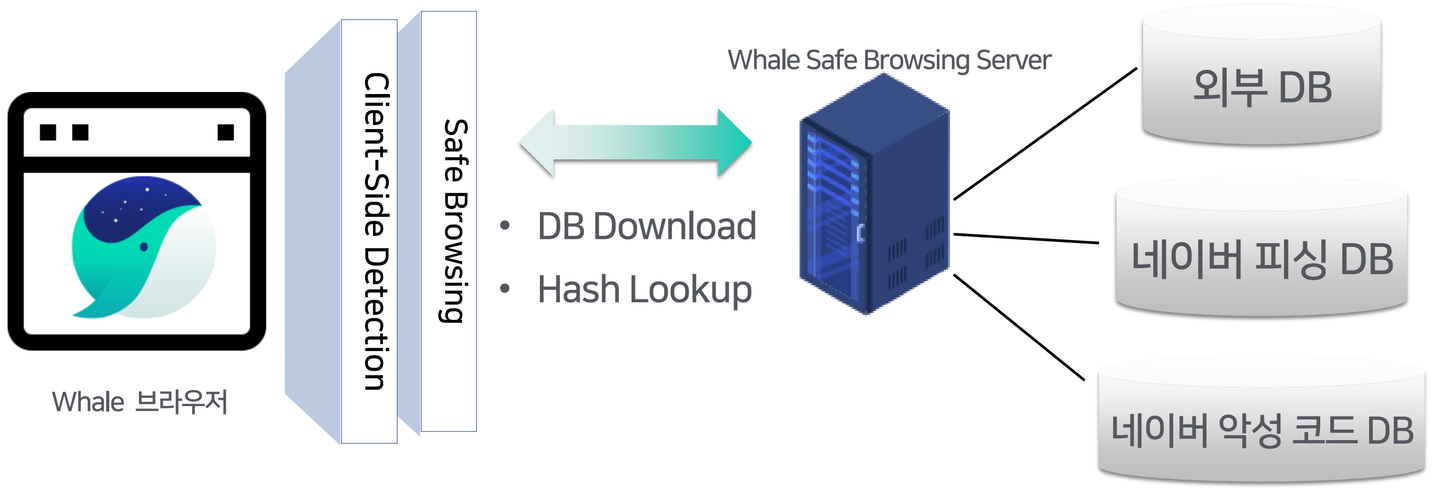


2. Whale 안티 피싱 구조

2.1 전체 시스템 구조

Whale은 피싱 대응은 2가지 다른 형태의 모듈을 사용

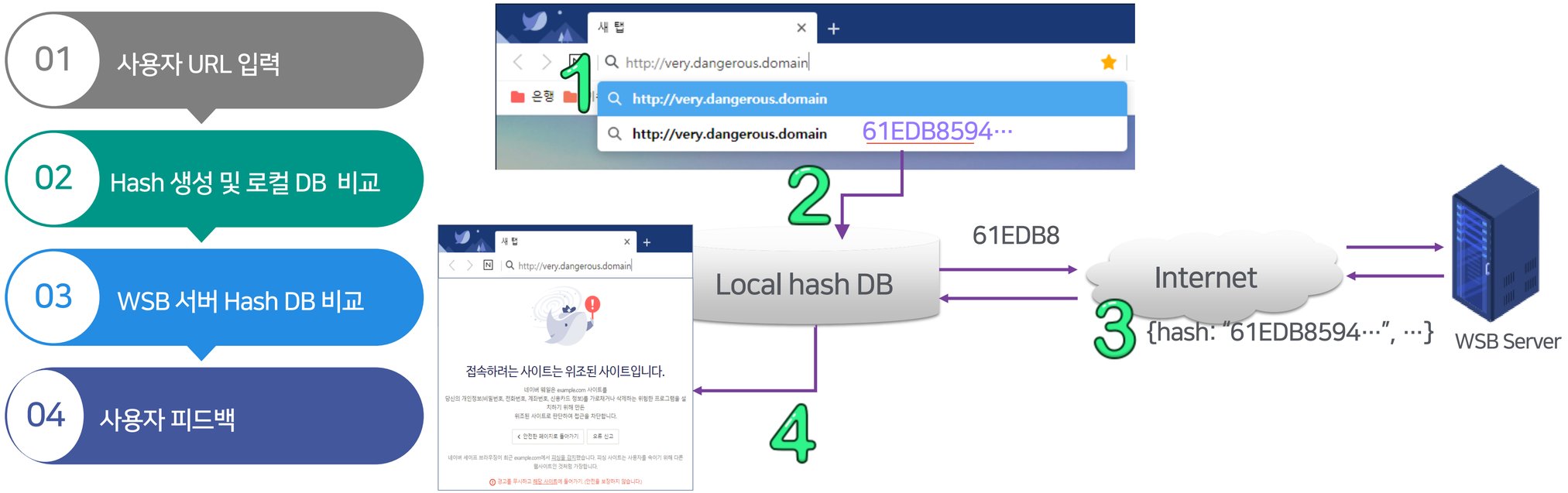
- 1) Whale Safe Browsing 모듈 - DB 기반 탐지
- 2) Whale Client-Side Detection 모듈 - 페이지 분석을 통한 선탐지



2.2 Whale Safe Browsing(WSB) 모듈

피싱을 포함한 악성 URL 데이터베이스 기반의 대응 엔진

- 외부 제공 피싱 데이터와 네이버, 라인 자체 수집 데이터 적용
- URL이 데이터 베이스에 포함되었는지 확인 후 사용자에게 피드백 수행



2.3 Client-Side Detection(CSD) 모듈

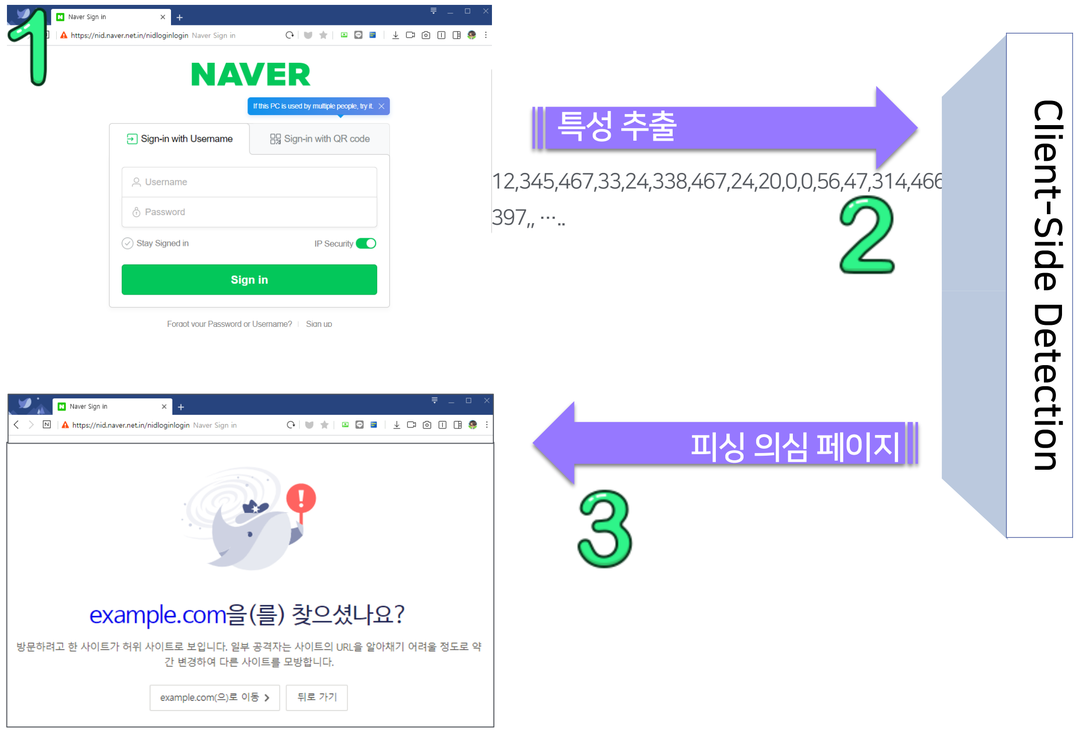
WSB 업데이트 전 zero-day 피싱 대응을 위한 탐지 엔진

- 사전 확보 DB 없이 웹 페이지 혹은 URL 특성 정보를 이용한 탐지 기술

01 사용자 URL 입력

02 특성 추출 및 피싱 여부 판단

03 사용자 피드백



3.Whale Safe Browsing

3.1 WSB 배경 설명

외부 피싱 데이터베이스 의존성으로 인한 한계성

- 외부 데이터는 주로 해외 피싱 데이터만 포함하여 국내 피싱 대응이 어려움
- 외부 데이터에서 발생하는 오탐에 대한 대응이 느리며, 최신 피싱 차단에 한계가 있음
- 적용 가능 DB 예시) Open Phish, Phish Tank, APWG 데이터, GSB 등

출발점 : 외부 데이터 의존으로 인한 한계를 어떻게 해결할까?

추가 DB를 구축하여 외부 데이터 의존 한계성을 극복하자

※ *WSB : Whale Safe Browsing*

3.2 WSB 개발 고려사항 (1/2)

1. 외부 피싱 데이터의 중요성 유지

- 글로벌 피싱 데이터를 포함한 외부 데이터는 여전히 중요함
- 해외 피싱 데이터 서비스를 위해 유지

2. 설계 상 고려 사항

- 다양한 소스로부터 수집한 데이터 호환성을 고려한 설계
- 오탐에 대한 빠른 대응 방법 고려
- 최신 피싱의 빠른 차단이 가능한 구조 설계

3.2 WSB 개발 고려사항 (2/2)

3. 자체 데이터베이스 유지 전략

- 자체 DB는 국내 피싱에 집중하여 데이터 축적
- 자체 독립 서버 구축하여 DB 관리

4. 개인 정보 보호 고려사항

- 개인 방문한 url의 정보를 서버가 알 수 없게 설계

3.3 WSB 의 동작

Local Hash DB를 통해 url 전송없이 피싱을 탐지

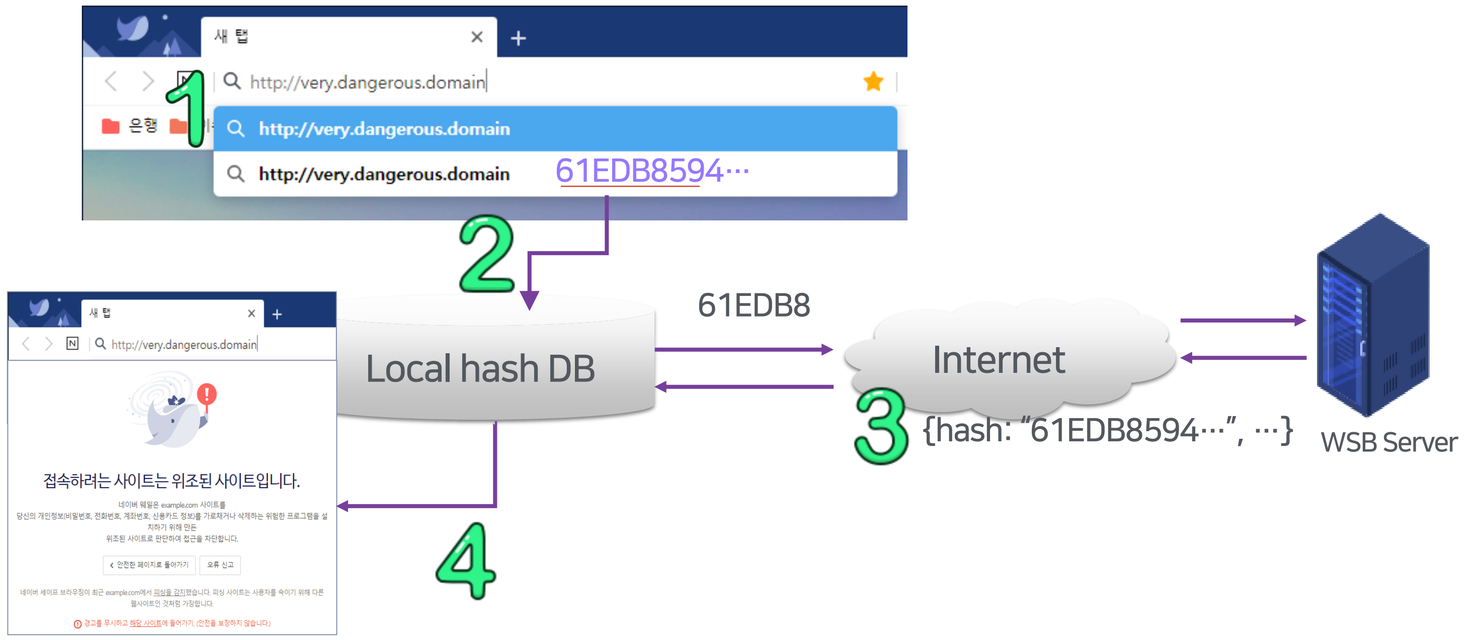
- 서버에는 Hash의 일부 값만 전송하여 개인정보 전송을 방지

01 사용자 URL 입력

02 Hash 생성 및 로컬 DB 비교

03 WSB 서버 Hash DB 비교

04 사용자 피드백



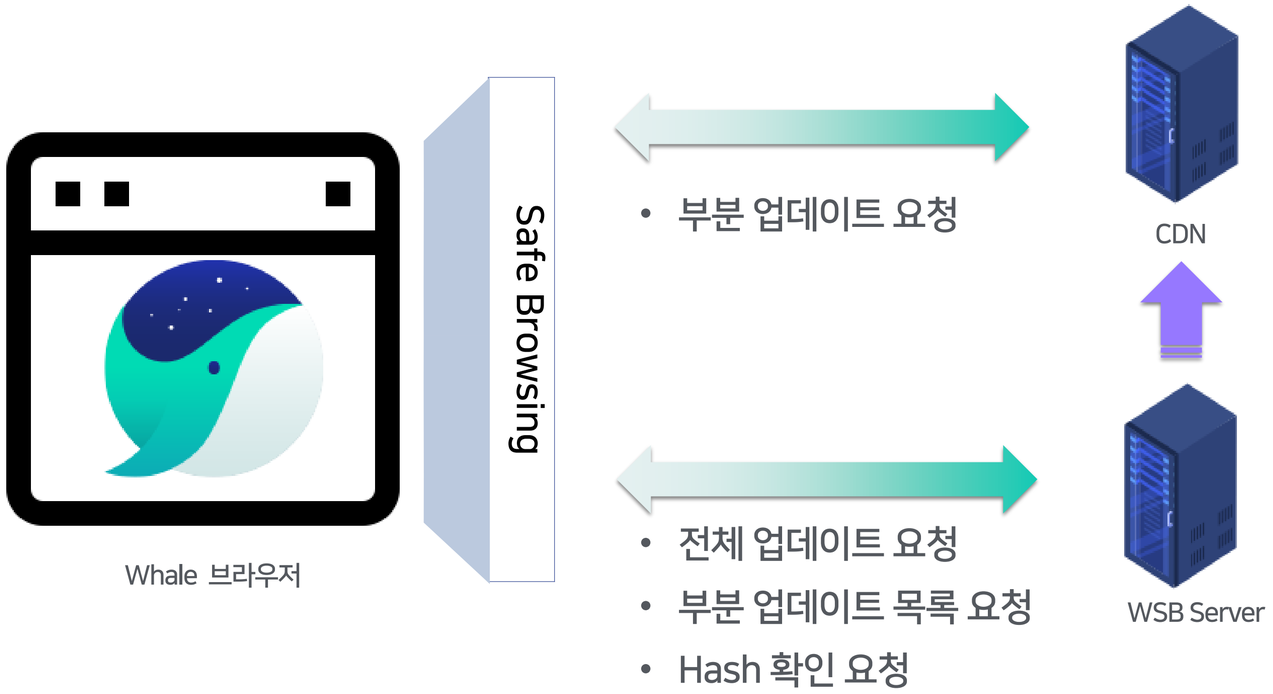
3.4 WSB 운영 고려사항

Whale과 서버간 통신양의 최소화 필요

- 데이터 베이스 업데이트 및 전송으로 인한 많은 네트워크 비용이 발생함
- 사용자 보호를 위한 “적절한 횟수”의 부분 업데이트는 필수사항 임

해결 방안

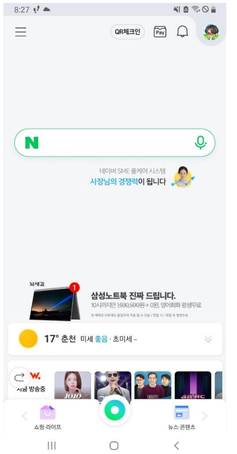
- CDN을 이용하여 트래픽 감축
- 네트워크 비용을 고려한 전체 업데이트 시간 분산 (월요일 아침 9시)



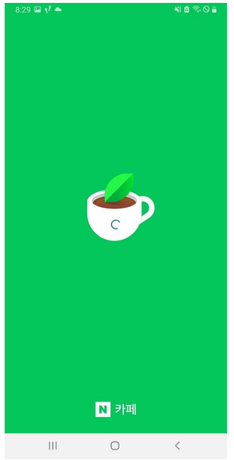
3.5 WSB 적용 대상

Whale 엔진(XWhale)을 이용하는 제품

- 네이버 앱
- 카페 앱(적용 검토 중)



네이버 앱



카페 앱

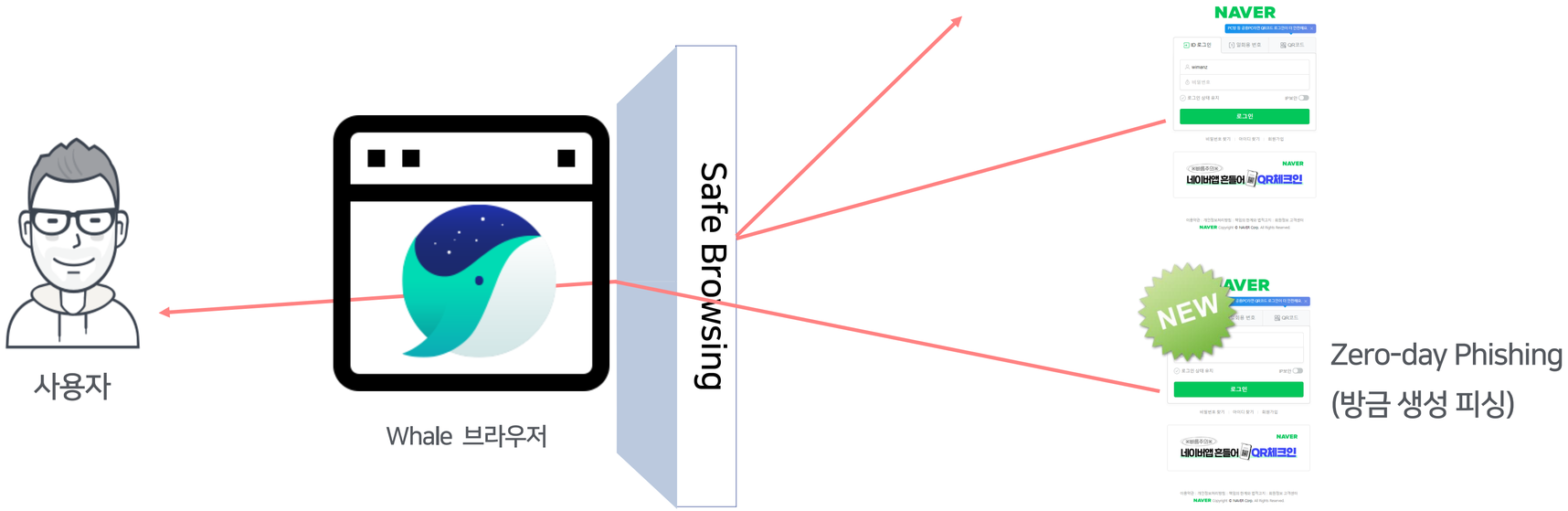
네이버 서비스

- 네이버 QR 코드
- 단축 URL

3.6 WSB 의 한계

진화하는 피싱에 대응하기에 역부족

- DB에 수집되지 않은 피싱에 대한 대응이 어려움
- 사용자 증가와 DB 증가 시 네트워크 트래픽이 증가



4.Whale CSD 적용 검토 기술

4.1 CSD 개요

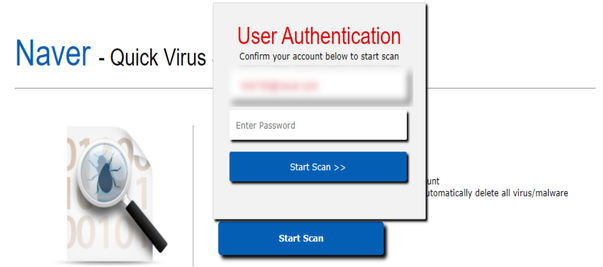
목표 : 피싱 DB에 없는 zero-day 피싱 탐지 (오탐없이)

- 새로 생성된 피싱도 탐지 가능해야 함
- 피싱 탐지 결과에 오탐이 없어야 함

CSD의 공통 가정 사항

피싱은 대상 사이트의 외향적인 요소를 모방한다.

↳ 외향적인 요소를 고려하지 않은 피싱은 사용자를 속이기 어렵다.



다른 외형의 네이버 피싱 사례

※ *CSD : Client-Side Detection engine*

4.2 기술 개발 진행 방향

DEVIEW 2021

3가지의 기술적인 접근 진행

접근 1. URL 기반 탐지

- URL을 구성하는 keyword들을 보고 판단
- 필요한 정보가 적어 빠른 탐지 가능

접근 2. HTML 컨텐츠 기반

- html 내의 컨텐츠들을 보고 판단
- 컨텐츠 기반으로 URL 기반 탐지 보다 정확한 탐지가 가능

접근 3. 시각적 특성 기반

- 사용자에게 보여지는 시각적 특성을 보고 판단
- 분석 불가한 스크립트, 이미지 사용에 영향없이 탐지

4.3 URL 기반 탐지 - 개요

URL 내에 특성 정보를 이용하여 피싱인지 여부를 판단하는 방식

- URL에 포함된 도메인, path 는 유의미한 데이터를 포함하는 경우가 많음
- 기존에 URL 특성을 이용한 피싱 혹은 악성 코드 URL 탐지 시도가 있었음

네이버 대상 피싱의 경우도 일부 공통 특성들을 포함

- 네이버 서비스와 유사한 URL 구성을 포함

네이버 피싱 URL 예시

https://nid.naevear.com/nidlogin.login?mode=form&url=https%3A%2F%2Fwww.naver.com&locale=ko_KR&svctype=1

http://navcorpmanager.website/?cfyyFGMj62v0rdawSHq2Bs1NUvuM75Tv%20yyFGMj62v0rdawSHq2Bs1NUvuM75Tv&ref=ZGNwNjEz

https://nidlogin.naversky.com/nidloginlogin?mode=form&url=https%3A%2F%2Fmail.naversky.com%2F&email=qwertyabcd&nhntype=2020

...

4.3 URL 기반 탐지 - 기존 기술 접근 방식

악성 URL의 일반 URL 과 다른 특성들을 기반으로 탐지 시도

- 매우 긴 URL, 특수 문자 사용 빈도 등 일반 URL에서 볼 수 없는 패턴들을 사용

피싱 URL은 일반 URL 패턴과 다르다.

최근 정교하게 만들어지는 피싱 페이지의 경우 탐지하기 어려움

기존 사용 URL 특성 예시

- IP 주소 사용
- 도메인 영역 내 '-' 사용
- 구두점 빈도 : '!' '#' ..
- 긴 호스트 이름
- ...

4.3 URL 기반 탐지 - 접근 방법

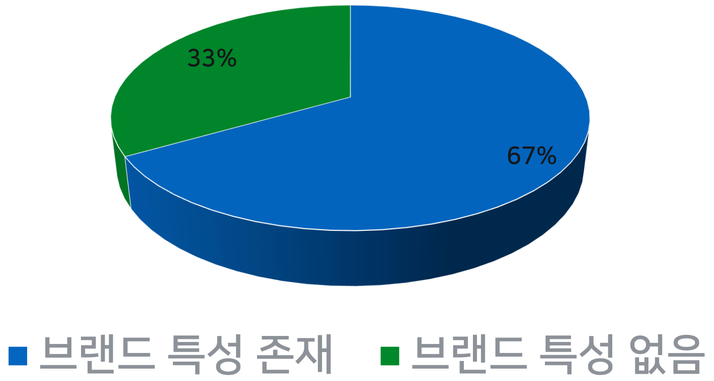
피싱 대상 서비스 URL과의 유사성을 기반으로 한 방식으로 접근

- 브랜드 이름, 서비스 이름, 페이지 이름 등이 피싱 URL 내에 포함되는지 여부 등
- 브랜드 및 서비스 이름과 유사한 정보를 고려하여 평가 진행

유사성을 고려한 특성 정보 예시

- 사이트의 브랜드 이름 혹은 서비스 이름
- 변형된 브랜드 이름 혹은 서비스 이름
- 피싱 대상 페이지의 path 유사 키워드
- 일반적인 피싱에서 공통적으로 발생하는 키워드
- …

최근 3개월 피싱 URL 브랜드 포함 비율



※ 실제 사용 파라미터 및 키워드는 정보 유출의 이유로 설명에서 제외함

4.3 URL 기반 탐지 - 평가 방법

입력 URL에 대한 URL 피싱 지수를 계산하고, 임계 값에 따라 피싱 후보군 선정

- 총 5개의 추출 정보를 선정하였고, 개별 특성 별로 가중치 파라미터를 적용
- 정상 사이트의 패턴을 특성에 추가하여 true negative 발생 보완
- URL 피싱 지수가 미리 정의한 임계 값 범위 인 경우 대응

$$URL\ Phishing\ Score = \sum_{i=1}^{5} (F_i \times W_i)$$

※ 실제 사용 파라미터 및 키워드는 정보 유출의 이유로 설명에서 제외함

4.3 URL 기반 탐지 – 고민해본 문제

DEVIEW 2021

변형된 브랜드 이름을 검출 할 수 있는 방법이 있을까?

- 일부 letter를 변경한 유사 키워드 사용 사례, 다른 단어와 합쳐서 사용하는 복합어 패턴
- 유사 키워드 예 : never, naever, naaver 등
- 복합어 패턴 예 : naverlogin.com, navermail.com 등

유사 패턴

naever

다수의 창의적인 유사 패턴 존재

유사 키워드 검출 방법 필요

- 편집 거리 알고리즘 적용

복합어 패턴

naverlogin

고정된 단어 중심의 복합어 구성

간단한 문자열 비교로 처리 가능

4.3 URL 기반 탐지 - 유사 패턴 검출

Levenshtein distance [1] 적용 – 편집 거리 알고리즘

- 두 문자열을 동일한 문자열로 변경할 때 수행해야 하는 삽입, 치환, 삭제, 추가 에 대한 비용 계산
- kitten -> sitten (substitution of “s” for “k”) : 1
- sitten -> sittin (substitution of “I” for “e”) : 1
- sittin -> sitting (insertion of “g” at the end) : 1

편집 거리 점수 계산 후 임계 값 이상인 경우 유사 키워드 판정

1 Levenshtein distance, https://en.wikipedia.org/wiki/Levenshtein_distance, last updated: 2021-09-20, visited : 2021-09-30

4.4 HTML 컨텐츠 기반 - 개요

DEVIEW 2021

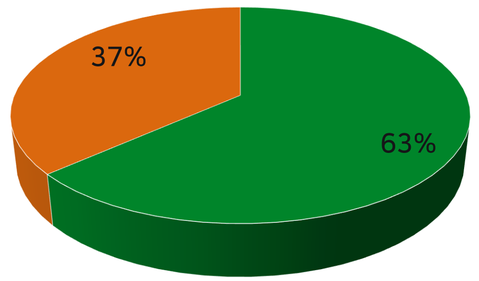
URL을 유사하게 쓰지 않는 피싱의 경우 URL 기반 탐지 기법으로 탐지하기 어려움

- 네이버 피싱의 경우 37% 정도가 URL 내 유사 패턴이 존재하지 않았음

피싱 탐지 방안 필요

컨텐츠 기반 탐지

피싱 URL 내 네이버 관련 브랜드 포함 URL 비율

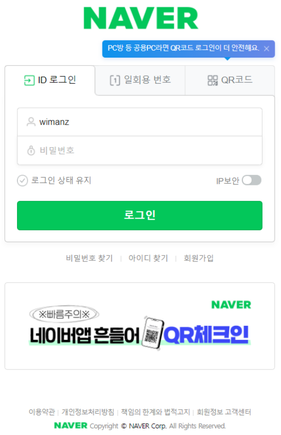


최근 4년 , 최근 3개월 네이버 피싱 URL 분석 결과

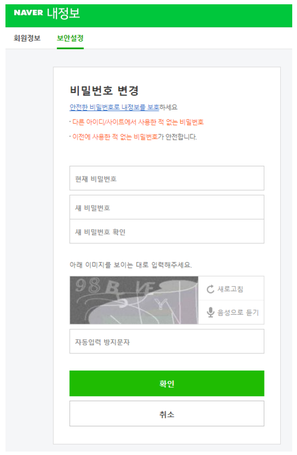
4.4 HTML 컨텐츠 기반 - 배경 지식

피싱의 모방 페이지는 모방 대상 사이트 별로 한정적

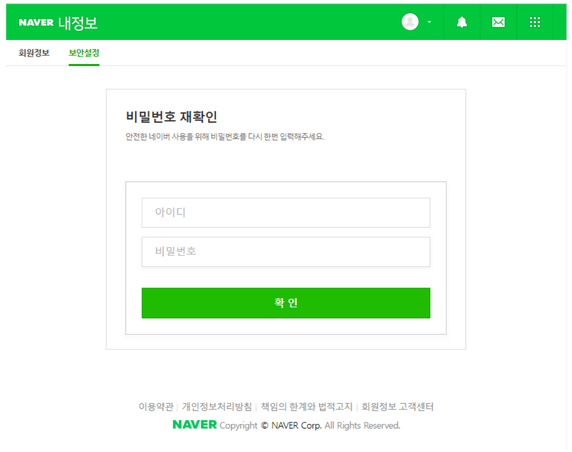
- 피싱 모방 대상은 주로 로그인, 비밀번호 입력 페이지로 한정됨
- 네이버 피싱의 경우 3가지 모방 대상 페이지로 한정되어 있음을 확인



네이버 피싱 타입 A



네이버 피싱 타입 B



네이버 피싱 타입 C

네이버 피싱의 3가지 모방 페이지

4.4 HTML 컨텐츠 기반 – 기본 접근 방식

피싱 페이지 내에 원본 페이지에서 추출한 특정 키워드들이 존재 하는지 여부를 기준으로 평가

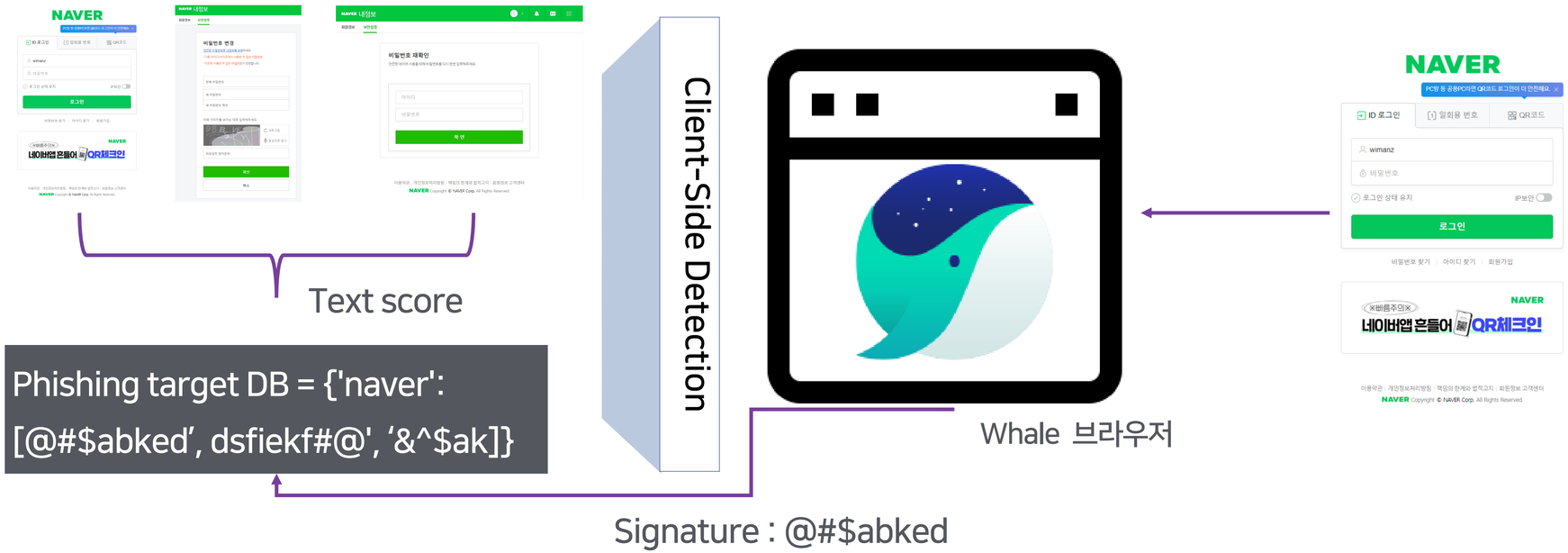
- 키워드란 피싱 타깃 페이지에 포함된 대표적인 단어들을 의미
- 정상 사이트를 피싱으로 잘못 탐지하는 사례가 많이 발생

기본 방식의 정확도(오탐 줄이기)를 높이는 방식 필요

4.4 HTML 컨텐츠 기반 - 접근 방법

사용자가 사이트 접속 시 페이지 signature를 추출하여 비교

- 3가지 타입의 페이지의 고유 signature를 생성하여 각 페이지와 비교
- Signature 생성을 위한 알고리즘 개발 및 적용 – 예) 적용 가능 Alg. : TF/IDF 등



4.4 HTML 컨텐츠 기반 - 적용 시나리오

DEVIEW 2021

URL 기반의 탐지 방식을 보완하는 방식으로 사용 가능

- URL 기반 탐지 기술에서 후보군으로 식별된 페이지에 대한 검증 용도

독립적인 방식으로 적용 가능

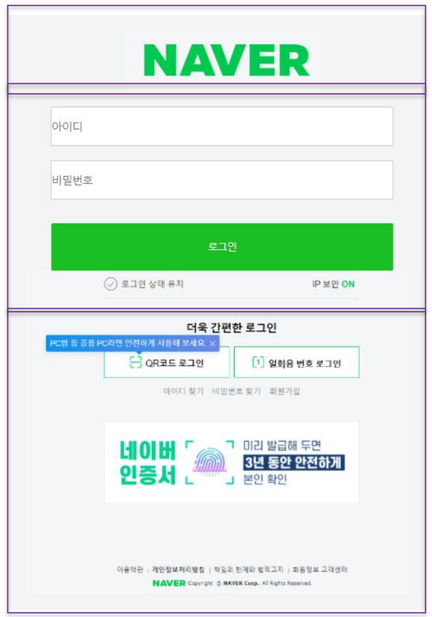
- 폼 태그 인식 등과 같은 민감한 정보 입력 기능이 포함되는 모든 웹 페이지에 대해서 평가하여 피싱 검출

4.5 시각적 특성 기반 - 개요

피싱 탐지 우회를 위해 이미지 기반의 피싱 탐지 공격 사례가 발생

- 이미지 기반 피싱은 제한적인 html 데이터를 제공하여 기존 탐지 기술(html 컨텐츠 기반)을 우회 가능

실제 이미지 기반 피싱 사례



이미지 영역 1

폼 영역 2

이미지 영역 1

4.5 시각적 특성 기반 - 접근 방법

DEVIEW 2021

탐지 우회 기술 영향을 줄이기 위해 피싱의 시각적 특성을 이용한 탐지 방안 고려

- 개별 사이트의 서비스는 고유 디자인 영역을 포함

브랜드 서비스 별 시각적인 특성

4.5 시각적 특성 기반 - 디자인 영역

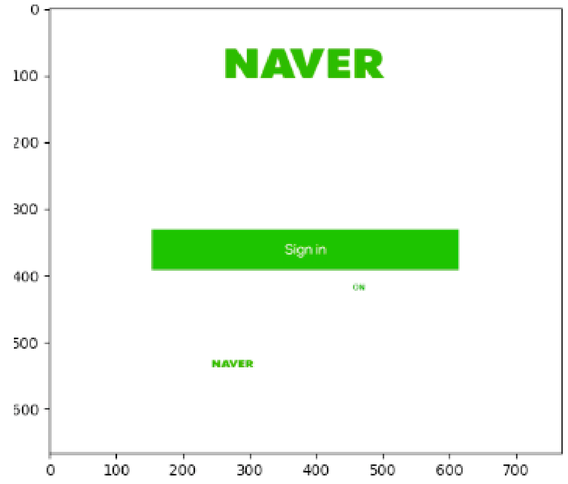
DEVIEW 2021

각 사이트를 대표하는 색, 혹은 로고 등 사람들이 인식하는 디자인적인 요소가 포함된 영역

- 개별 사이트의 고유 디자인 영역의 오브젝트들 추출하여 비교

네이버 로그인 페이지

디자인 영역 추출



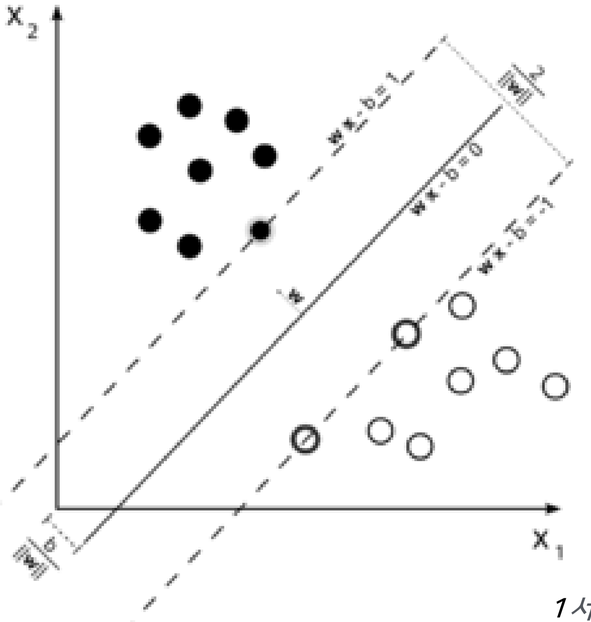
네이버 로그인 내 디자인 영역

각 사이트 별 디자인 영역을 정의하여 활용

4.5 시각적 특성 기반 - 영역 비교 방안 검토

여러가지 영역 비교 방안을 고려함

- 이미지 오브젝트 표현 알고리즘(SITF 등)기반 단순 비교 – 정확도가 낮음
- Deep learning 기반 기법 적용 검토 – 적은 데이터 셋으로 적용 어려움
- 적은 데이터로도 분류 가능한 Support Vector Machine 적용



Wikipedia 설명[1]

서포트 벡터 머신(support vector machine, SVM[1].[2])은 기계 학습의 분야 중 하나로 패턴 인식, 자료 분석을 위한 지도 학습 모델이며, 주로 분류와 회귀 분석을 위해 사용한다. 두 카테고리 중 어느 하나에 속한 데이터의 집합이 주어졌을 때, SVM 알고리즘은 주어진 데이터 집합을 바탕으로 하여 새로운 데이터가 어느 카테고리에 속할지 판단하는 비확률적 이진 선형 분류 모델을 만든다.

1 서포트 벡터 머신, https://ko.wikipedia.org/wiki/서포트_벡터_머신, last updated: 2020-07-10, visited : 2021-09-30

4.5 시각적 특성 기반 - SVM 적용 개요

DEVIEW 2021

SVM 적용을 위해 데이터 수집 및 전처리 작업 필요

- 4단계의 머신 러닝 수행 과정 필요

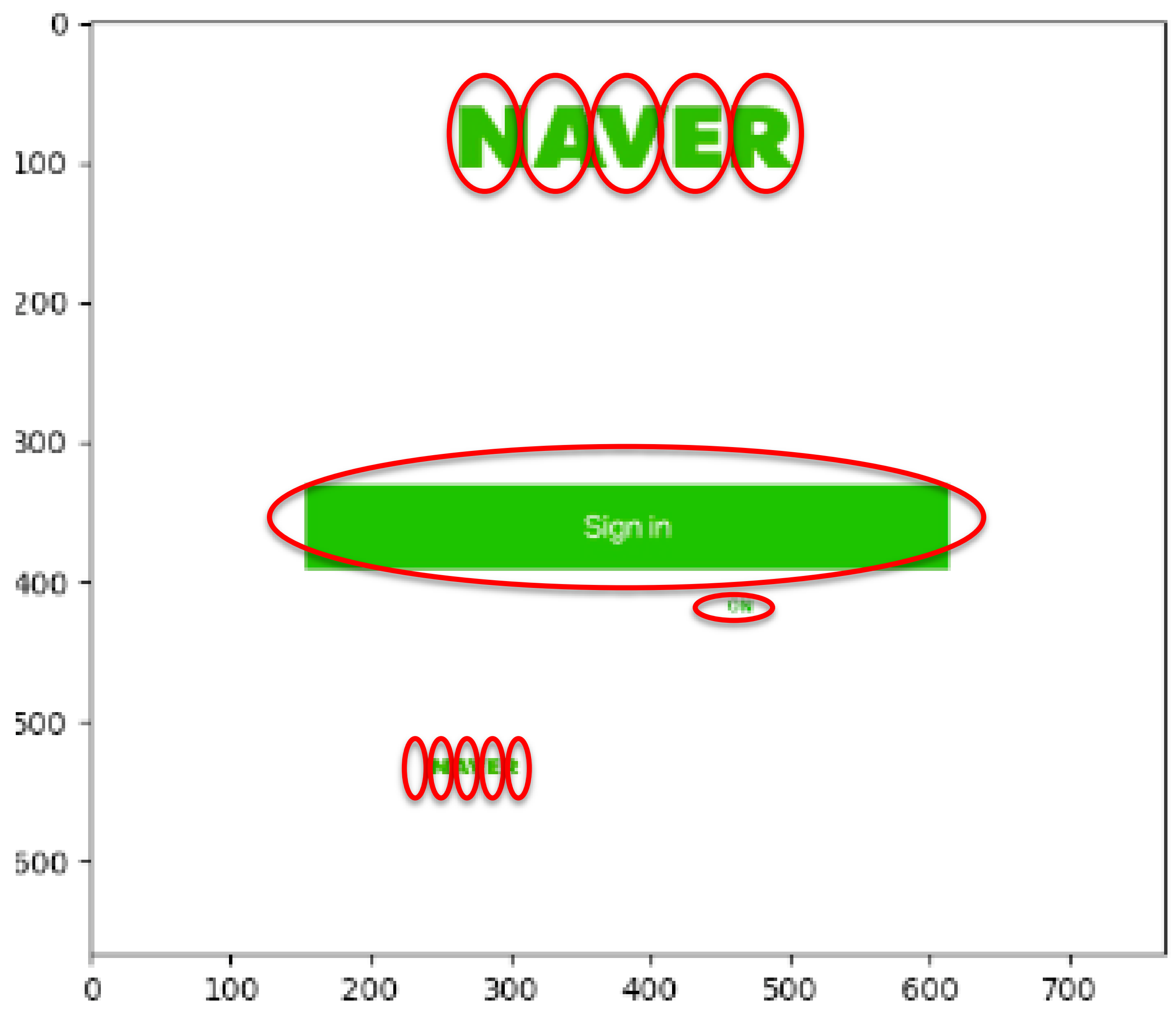
01	데이터 수집	주어진 문제를 해결하기 위해서 어떤 데이터를 사용할지 선택
02	데이터 전처리	데이터를 적용하는 알고리즘에 적합하게 조정하는 작업
03	데이터 학습	데이터 학습을 위해서 어떤 알고리즘에 어떤 파라미터를 적용할 지 선택
04	모델 평가	학습 완료된 모델에 대해서 정확도를 검토하고 보완하는 작업

4.5 시각적 특성 기반 – 데이터 전처리

DEVIEW 2021

디자인 영역 내의 오브젝트를 표현할 수 있는 값들을 선정

- 오브젝트 수
- 개별 오브젝트 별 위치, 면적, 둘레 길이



Feature extraction

Training data set

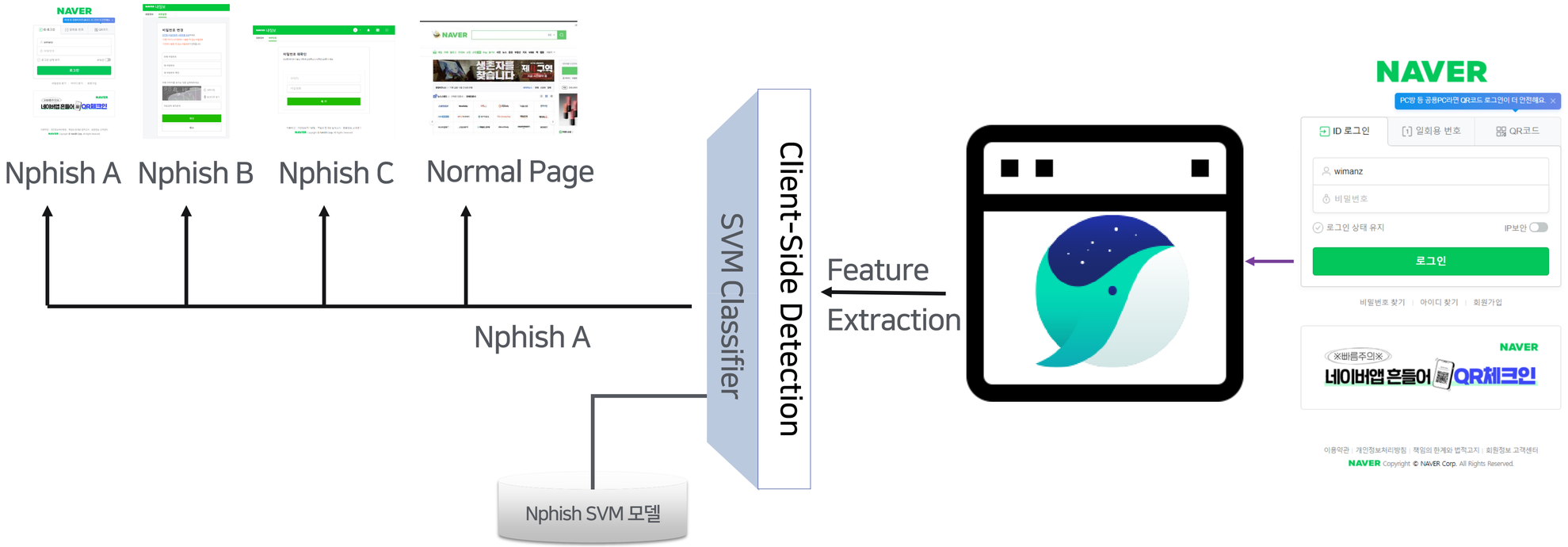
12,345,467,33,24,338,467,24,20,0,0,56,47,314,466,36,28,447,397,12,18,446,397,4,22,0,0,10282,627,454,196,558,115,427,195,434,125,402,195,429,124,370,199,519,105,339,195,595,144,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
...

Train 데이터 셋 생성 시 오브젝트 위치 조정한 임의 데이터도 포함

SVM 모델 생성 > 대략 95% 정도 정확도 확인

4.5 시각적 특성 기반 - 적용 시나리오

Whale 내에 미리 학습된 모델 파일 배포 후 웹 사이트 접근 시 분류



5. 진화하는 피싱 대응

5.1 피싱 진화에 대한 대응

세이프 브라우징 DB 업데이트

피싱 발생 시 빠르게 데이터를 수집하고 세이프 브라우징 DB에 반영할 수 있는 대응 시스템 필요

CSD 파라미터 업데이트

진화하는 URL 및 피싱 페이지의 트랜드를 분석하여 CSD 파라미터 및 모델에 반영 필요

피싱 데이터를 빠르게 많이 수집 할 수 있는 방안 마련 필요

5.2 피싱 데이터 확보 전략

피싱 공격의 96%는 e-mail을 통해서 전파됨[1]

- 피싱의 시작점은 메일 서비스임
- 네이버 메일 내 권한 확보가 가능한 범위 내에서 정보 취득 검토

메일은 개인 정보보호와 관련된 민감한 정보이지만, 사용자가 신고한 스팸의 경우 보안 목적으로 제한적인 내용만 확인

- 사용자 신고 스팸 메일 내에서 피싱 URL을 수집하여 대응

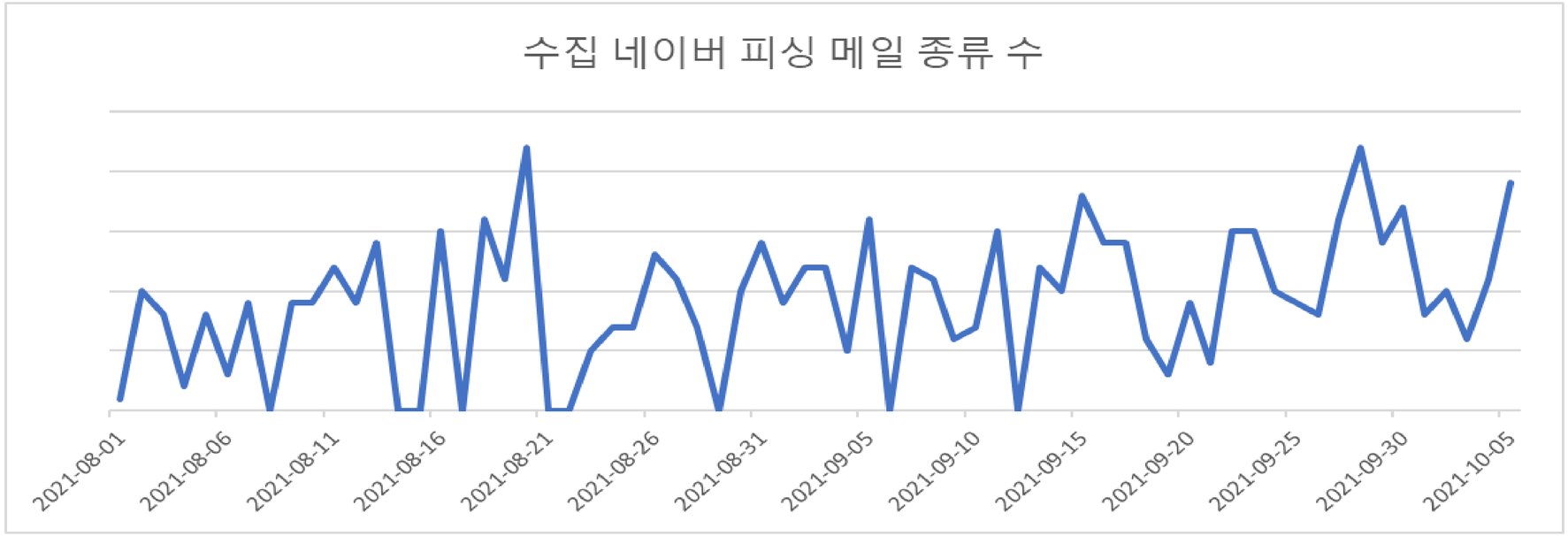
1. Must-Know Phishing Statistics, https://www.tessian.com/blog/phishing-statistics-2020,last updated: 2021-09-16, visited : 2021-09-30

5.2 피싱 데이터 확보 전략 - 스팸 내 피싱 수

DEVIEW 2021

네이버 피싱 URL이 포함된 다수의 스팸 메일 식별

- 매일 일정양의 네이버 피싱 메일 검출 가능
- 신규 피싱 URL을 확인



※ 1개의 스팸 종은 동일한 패턴의 스팸 메일을 의미함 (동 제목, 동 피싱URL패턴)

5.3 진화하는 탐지 모듈 - WSB

네이버 자체 피싱 DB 구축 및 피싱 데이터 확보 노력

- 네이버 스팸 메일 시스템 기반으로 최신 피싱 데이터 수집
- NAPS(Naver Anti-Phishing System) 선제적인 피싱 URL 수집
  - 회원 피싱 탐지 모듈, CT 탐지 모듈, Whale CSD 탐지 URL, HUMINT기반

국내 주요 사이트들에 대한 피싱 URL 수집 및 적용 예정

- Whale 사용자들 사용 국내 주요 사이트에 대한 피싱 URL 데이터 확보 기술도 확대 적용 예정

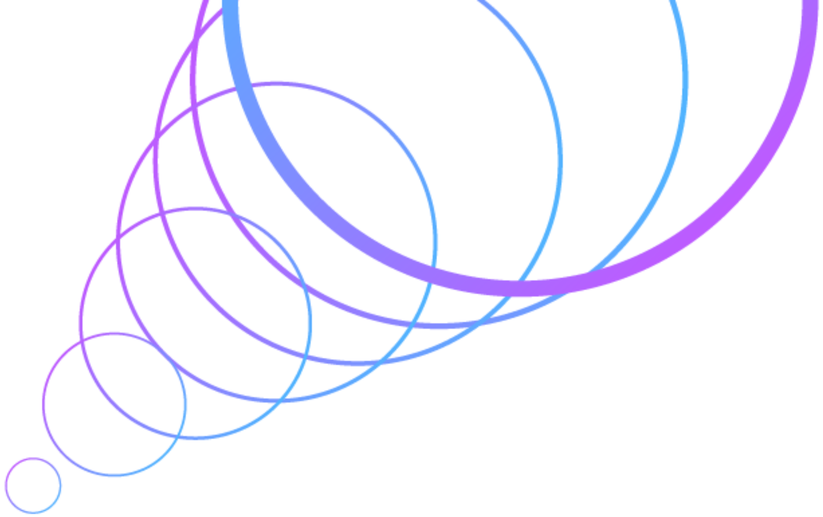
5.3 진화하는 탐지 모듈 - CSD

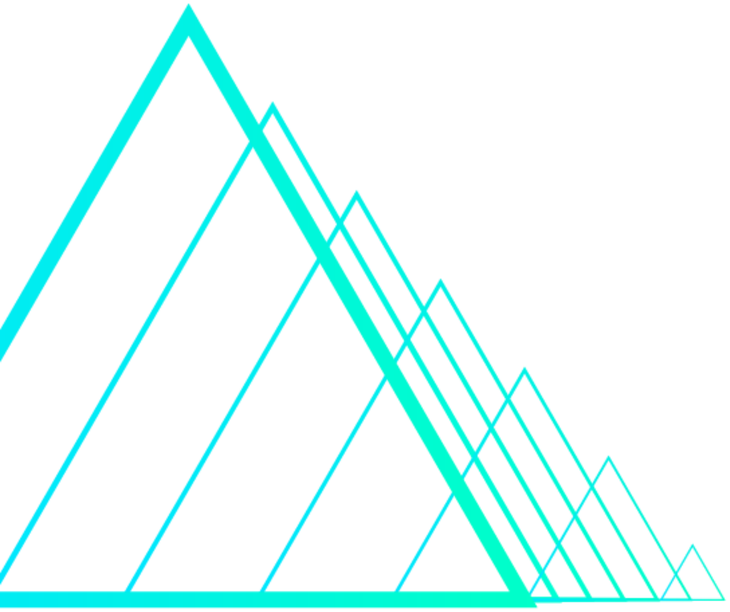
피싱 탐지 알고리즘 파라미터, 모델에 대한 업데이트 수행

- 네이버 스팸 메일 시스템 기반으로 수집한 최신 피싱 데이터를 활용
- CSD 적용 파라미터 및 평가 모델의 주기적인 업데이트 수행
- CSD 적용 결과를 주기적으로 분석하여 알고리즘 개선 수행

국내 주요 사이트 들에 대한 파라미터 및 signature 추가 예정

- Whale 사용자들이 사용하는 주요 사이트 대응 파라미터, signature 등을 추가할 예정





Q & A

Thank You