

# 무분별한 스크래핑 어떻게 대응하고 계신가요

안티스크래핑 솔루션의 필요성과 사례



# 1. 배경

## 다양한 스크래핑 서비스

서비스 사용자에게 스크래핑을 이용한 다양한 서비스는 편리함을 제공하며 급성장세를 보이며 토스, 삼짬삼, 핀다 등 대표적인 서비스 이외에도 다양한 방면에서 사용되고 있습니다.



- + 사용현황  
월간 활성 사용자수 : 1,900만 이상  
서비스 수 : 70개 이상

- + 주요 서비스  
예금 조회/송금 등 은행업무  
신용등급 조회 서비스  
전자서명인증서비스 / 인증서서비스  
대출이자 등 비교서비스



- + 사용현황  
가입 고객수 : 2,000만 이상  
월간 활성 사용자수 : 660만 이상

- + 주요 서비스  
종합소득세 조회/환급  
종합소득세 납부 서비스  
개인사업자 부가가치세 신고



- + 사용현황  
다운로드 수 : 600만 이상  
누적 대출관리액 : 139조 이상

- + 주요 서비스  
대출중개 서비스  
대출 비교 및 대환 서비스  
주식담보대출, 전세대출, 오토론 확장  
시장상권 분석서비스

# 1. 배경

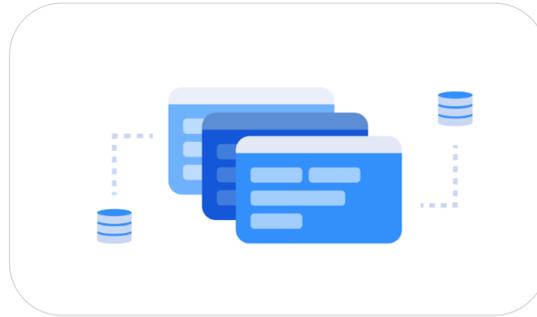
## 스크래핑의 형태

고도화되는 공격이슈에 따라 다양해지는 모바일 보안 대응 솔루션의 기술적용과 운영관리 과정에서의 발생하는 불편함은 모바일 서비스를 기획함에 있어 큰 걸림돌로 작용하고 있습니다.



### 크롤링(Crawling)

- + 방대한 데이터 수집을 목적으로 조건부에 의하여 자동으로 탐색/수집/분류하는 작업
- + 특징  
하나의 요소에 대한 심층 분석에 유리, 실시간 정보수집을 하여 데이터의 변화에 대한 반응이 유연



### 스크래핑(Scraping)

- + 원하는 정확한 정보를 특정 웹사이트나 페이지를 통해 요청하여 정보를 추출하는 작업
- + 특징  
원하는 정보/시간/출처도 특정하는 형태로 리소스를 최소화하고 데이터를 추출하는 데에 최적화



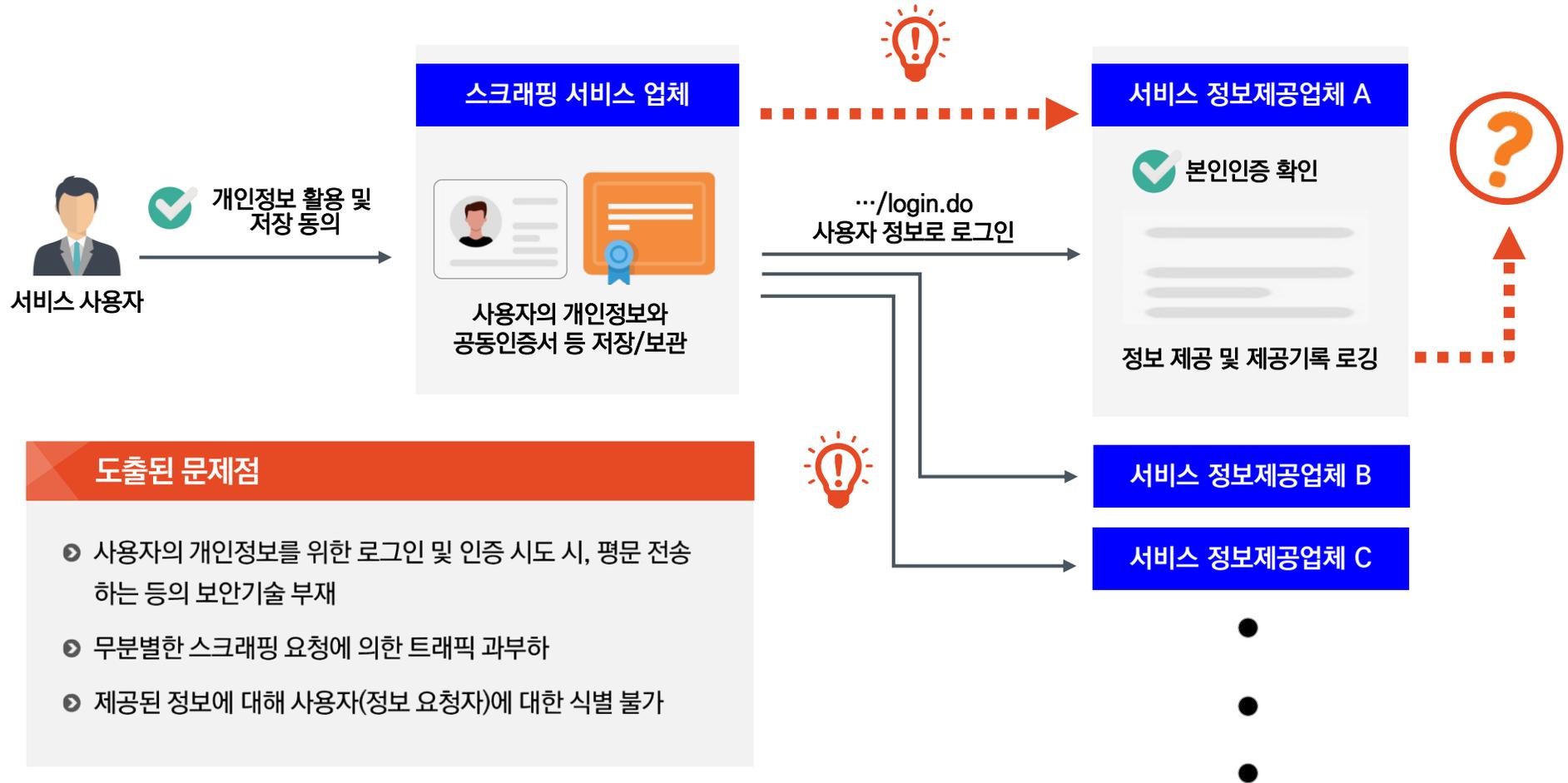
### RPA(Robotic Process Automation)

- + 사람이 해오던 반복적인 업무에 대해 자동화를 위하여 프로세스를 정립하고 데이터를 수집하거나 업무를 처리
- + 특징  
사용자의 업무 프로세스를 정립화하여 사용하는 형태로 개인별 Customizing이 확실하고 반복적 패턴을 보임

# 1. 배경

## 스크래핑의 문제점

편리하지만 무분별하게 늘어나는 스크래핑 서비스에 대한 문제점에 대해 다양한 각계 분야에서 경고하고 있습니다.



### 도출된 문제점

- ▶ 사용자의 개인정보를 위한 로그인 및 인증 시도 시, 평문 전송 하는 등의 보안기술 부재
- ▶ 무분별한 스크래핑 요청에 의한 트래픽 과부하
- ▶ 제공된 정보에 대해 사용자(정보 요청자)에 대한 식별 불가

# 1. 배경

## 시장 배경사항

편리하지만 무분별하게 늘어나는 스크래핑 서비스에 대한 문제점에 대해 다양한 각계 분야에서 경고하고 있습니다.

[파이낸셜뉴스] 가입자 1600만명을 돌파한 온라인 세금 환급 플랫폼 '삼점삼'이 개인정보 수집 등으로 국회 기획재정위원회 국정감사에서 도마 위에 올랐다.

김영선 국민의힘 의원은 지난 10일 기재위 국정감사에서 삼점삼을 지목해 홈택스 ID, PW, 건강 정보, 신용카드 정보 등 민감정보를 포함한 개인정보를 수집, 보유, 활용하는 상황에 대해 국제 청 차원의 단속, 시정조치나 감시체계가 없음을 지적했다.

개인정보보호위원회 행정처분 내용에 따르면 삼점삼 플랫폼 운영사 자비스앤빌런즈는 주민등록번호, 건강정보를 포함한 개인의 민감정보를 수집해 '최소수집원칙'을 위반했다.

- + 홈택스 ID, PW, 건강정보, 신용카드 정보 등 민감정보를 포함한 개인정보를 수집, 보유, 활용하는 상황에 대해 국제 청 차원의 단속, 시정조치나 감시체계가 없음을 지적

**정보 주체에 대한 식별 불가**  
(서비스 정보 제공업체)

- + 개인정보 수집과 이용, 파기에 관한 법정 고지사항과 구체적 정보를 알리지 않고 개인정보 처리 주체를 모호하게 적시하는 등 개인정보보호법 조항 다수를 위반

**서비스 이용자의 개인정보 보호**  
(스크래핑 서비스 업체)

- + OWASP - 스크래핑을 자동화된 위협으로 지정
- + KISA(개인정보보호 월간 동향 보고서) - 세계 12개 국가의 스크래핑 대응 관련 공동서명 관련 소식 게시
- + KISA(보안공지) - “자동 로그인 시도 차단 등 부정 로그인 차단 강화”

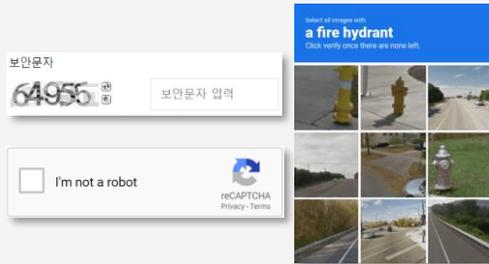
# 1. 배경

## 기존 대응방안

스크래핑 서비스들에 대항하여 국내외 서비스 제공업체에서는 아래와 같은 방안으로 대응하고 있으나 근본적인 해결책이 되진 못했습니다.

### Captcha

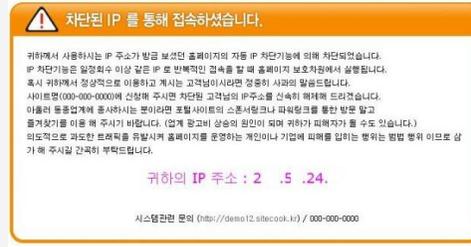
- 사람과 Bot 을 구분하기 위한 방법



- ▶ 이미 뚫려버린 입력형 캡차
- ▶ 캡차 농장(captcha Farm)의 등장
- ▶ 불편한 UI / UX
- ▶ Google 의 개인식별 이슈

### IP 차단

- 과도한 트래픽을 유발하는 IP 차단



- ▶ Proxy를 통한 간단한 IP 우회 가능
- ▶ 정상적인 사용자 차단 우려
- ▶ 사후 대처만 가능하다는 한계점

### HTML Randomize

- 데이터 추출을 어렵게 하기 위한 기술 (HTML class/id 등 을 randomize)

```
</div />
▶<div class="nH bKk nn" style="width: 1123px;">...</div>
▶<div class="nH bAw nn" style="width: 56px; height: 386px;">...</div>
</div>
<div class="dJ"></div>
```

- ▶ Randomize 범위에 대한 현실적 한계 (CSS, JS, Image 등 연계 제어)
- ▶ 적용을 위한 과도한 수정 범위
- ▶ 스크래핑 행위를 차단하는게 아닌 데이터로의 접근과 추출 난이도를 제고

## 2. 제안제품 소개

### 제품 개요

DynaPath는 스크래핑 서비스에 대한 강력한 보안성을 제공함과 동시에 간편하게 적용이 가능하고 실시간 정책적용 및 모니터링 기능을 통해 서비스 구축/운영 단계에서도 편의성을 보장합니다.



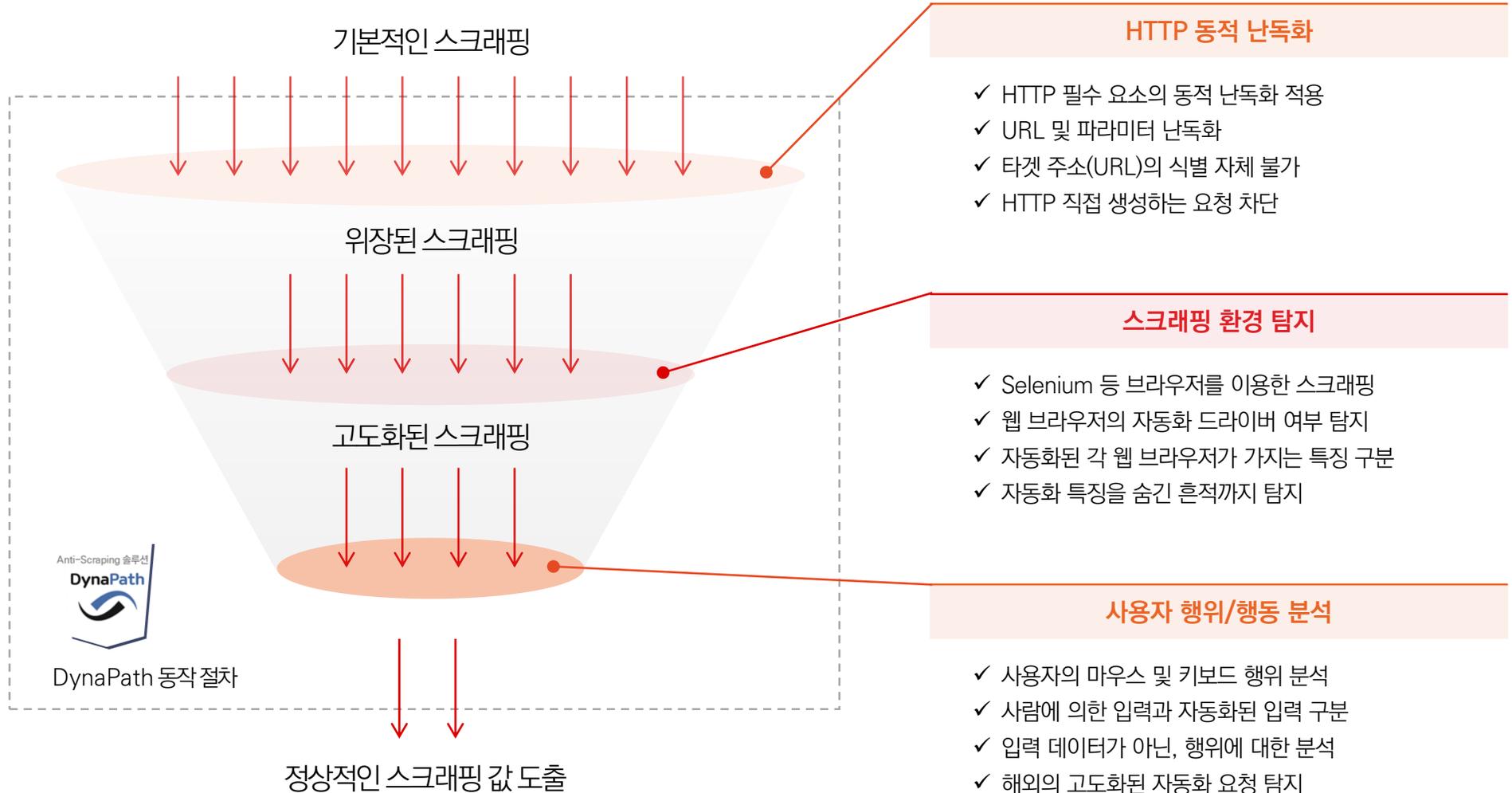
구분	내용
제품명	DynaPath(다이나패스)
제품 개요	<ul style="list-style-type: none"> <li>실시간 안티스크래핑 솔루션</li> <li>GS인증, 조달등록 대기중(6~7월경 예상)</li> <li>업무 WAS서버에 간편한 적용</li> <li>고도화되는 스크래핑 방식에 대한 지속적 유지관리</li> </ul>

주요기능	세부내용
실시간 스크래핑 탐지 / 차단	<ul style="list-style-type: none"> <li>스크래핑 툴을 이용한 요청의 탐지 및 차단</li> <li>Client 단말 환경 및 행위 분석을 통한 탐지</li> <li>JS 위/변조 차단 등 자기보호 기능</li> </ul>
인가 및 유입량 제어	<ul style="list-style-type: none"> <li>사전 협의된 스크래핑 요청 허용</li> <li>기본 정책은 차단, 선별적으로 허용 가능</li> <li>스크래핑 요청 유입량 제어 정책</li> </ul>
통계 및 실시간 모니터링	<ul style="list-style-type: none"> <li>실시간 스크래핑 요청 현황 모니터링</li> <li>인가된 스크래핑의 HTTP 요청 통계</li> <li>다양한 조건 별 통계 및 서버 상태 모니터링</li> <li>매니저 서버를 통한 재기동없이 정책 실시간 관리</li> </ul>

## 2. 제안제품 소개

### DynaPath 제품 개요

DynaPath는 3단계에 걸쳐 현존하는 다양한 스크래핑을 차단합니다.



#### HTTP 동적 난독화

- ✓ HTTP 필수 요소의 동적 난독화 적용
- ✓ URL 및 파라미터 난독화
- ✓ 타겟 주소(URL)의 식별 자체 불가
- ✓ HTTP 직접 생성하는 요청 차단

#### 스크래핑 환경 탐지

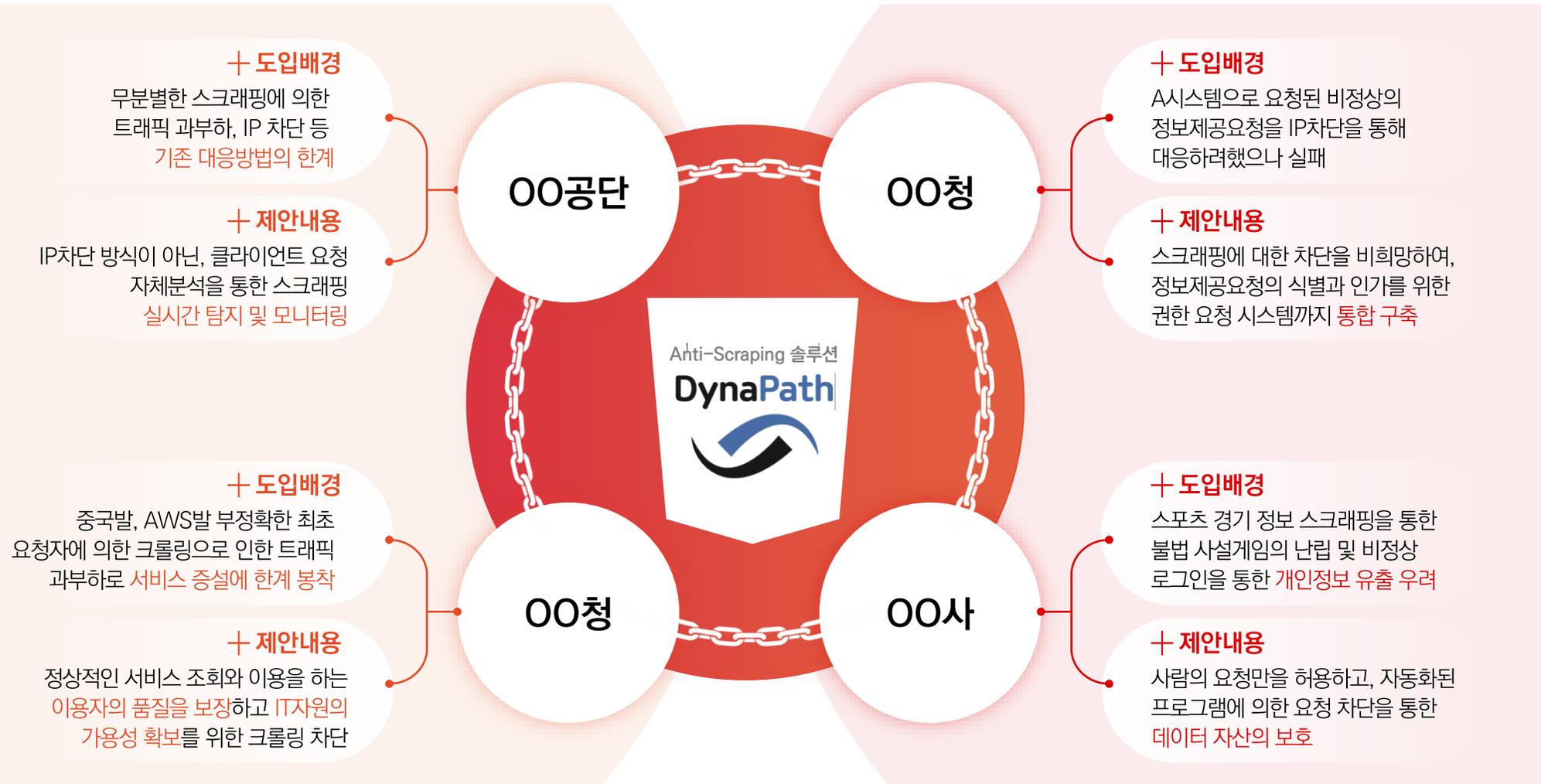
- ✓ Selenium 등 브라우저를 이용한 스크래핑
- ✓ 웹 브라우저의 자동화 드라이버 여부 탐지
- ✓ 자동화된 각 웹 브라우저가 가지는 특징 구분
- ✓ 자동화 특징을 숨긴 흔적까지 탐지

#### 사용자 행위/행동 분석

- ✓ 사용자의 마우스 및 키보드 행위 분석
- ✓ 사람에 의한 입력과 자동화된 입력 구분
- ✓ 입력 데이터가 아닌, 행위에 대한 분석
- ✓ 해외의 고도화된 자동화 요청 탐지

### 3. Reference

#### | DynaPath 도입 및 제품상담사례

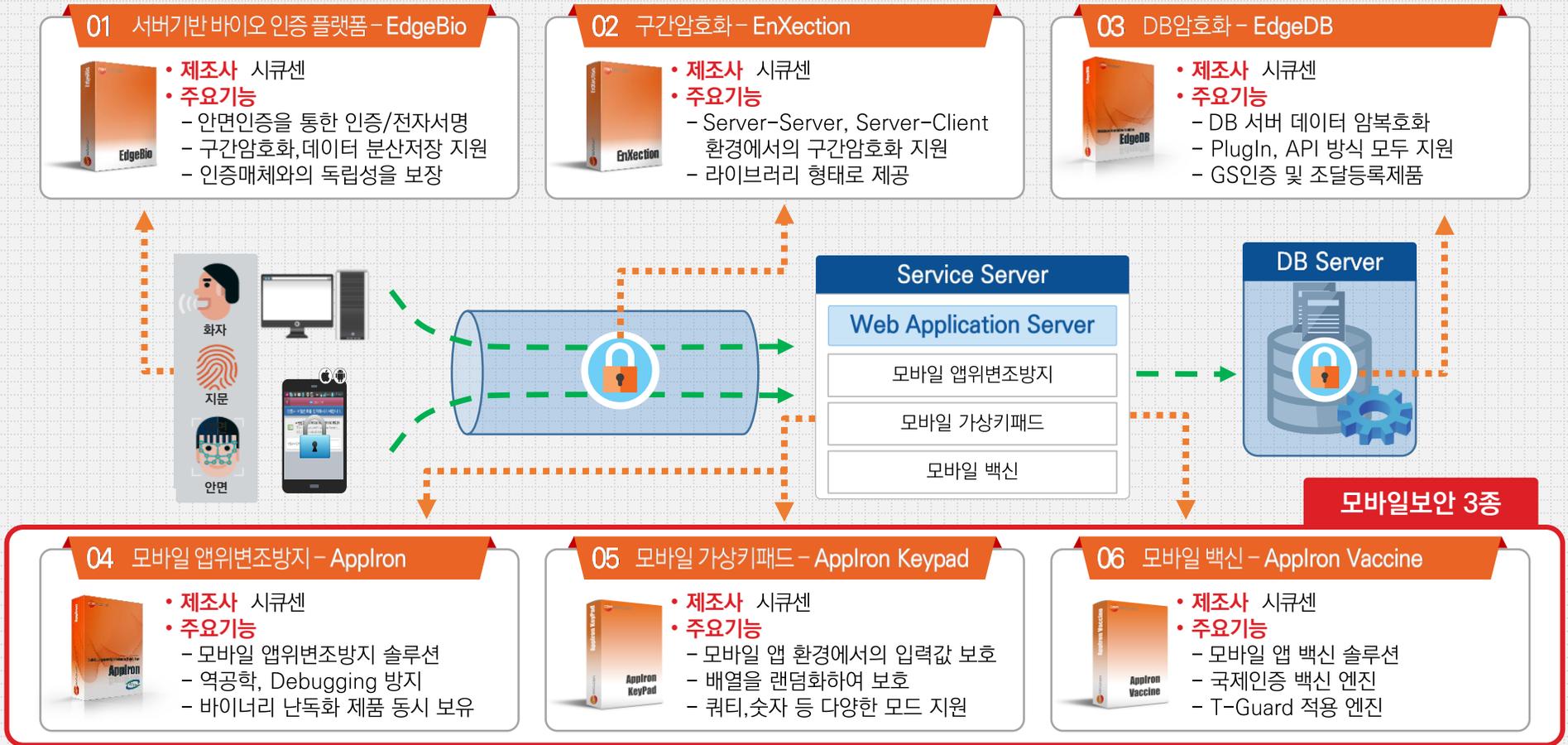




## 4. 회사 소개

### 제품 및 서비스 로드맵

시큐센은 PC 및 모바일 환경에서 사용자에게 다양한 인증 수단과 보안인프라 환경을 구축하기 위하여 정보보호 솔루션을 통하여 폭넓은 환경을 구현합니다.



서울특별시 영등포구 가마산로 343 콤텍빌딩  
343, Gamasan-ro, Yeongdeungpo-gu, Seoul, Republic of Korea  
TEL+82-2-3495-0700 FAX+82-2-521-6275

[WWW.SECUCEN.COM](http://WWW.SECUCEN.COM)

