생성형 AI의 다양한 공격 사례와 데이터 유출 관점의 효과적인 대응

Trellix Korea
허효승 이사

2023.09

Trellix

데이터의 관리는 계속 어려워지고 있습니다

데이터 증가

계속 더 많은 데이터 생성

2022년 생성된 데이터량

97ZB

1 ZB = 1 Billion TB

데이터 저장소증가

데이터는 점점 더 많은 저장소를 필요로 합니다

A leading global bank:

>600

Siloed data repositories

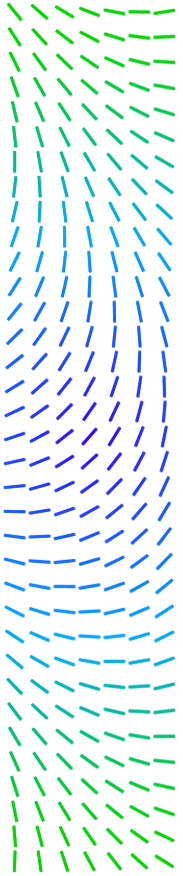
유출의 위험성 증가

내/외부의 위협으로 부터 보호

For a 1,000-person company:

15,000

External collaborators have access to company data



Trellix

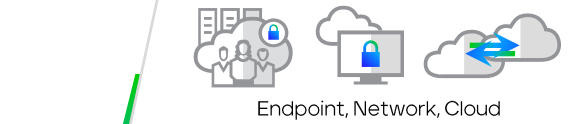
현재의 데이터 관리 현황

기존의 데이터 관리체계 이대로 괜찮을까?

실수

고의

침해
사고

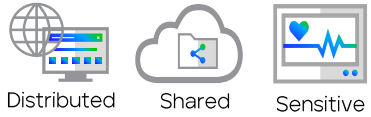


데이터의 분산

원격 근무

유출 위협

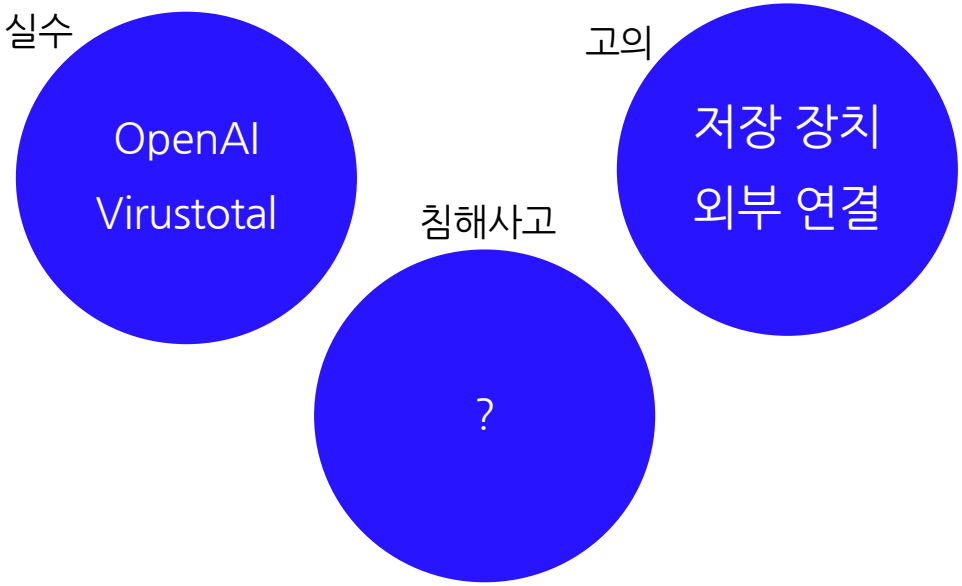
데이터의 복잡성

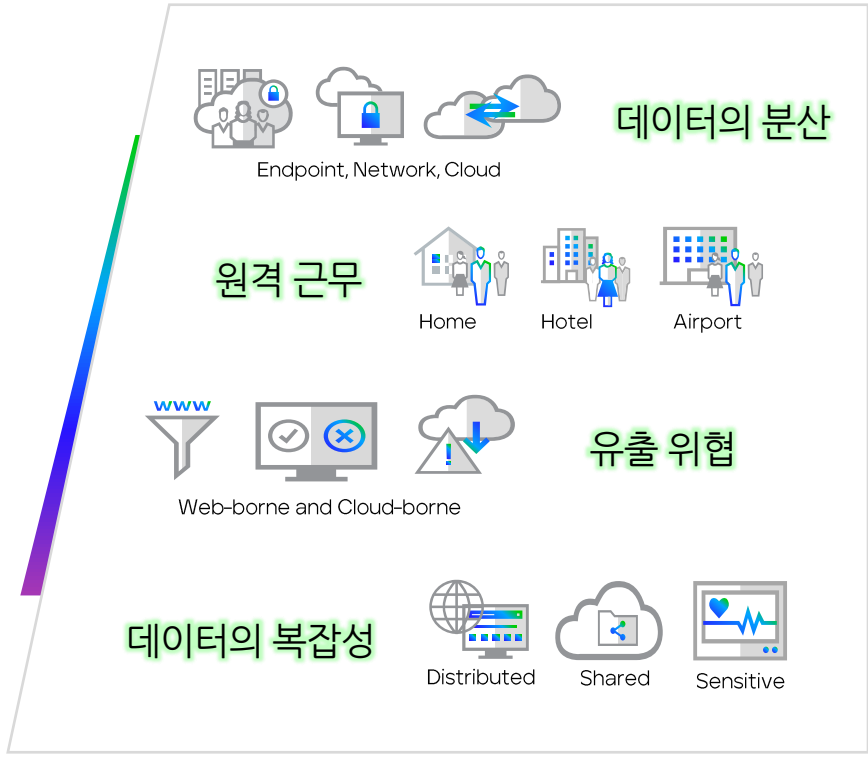


Trellix

현재의 데이터 관리 현황

기존의 데이터 관리체계 이대로 괜찮을까?





Trellix

OPEN AI 의 부정적 사례

구분	세부 내용
멀웨어 생성	• 초보 사이버범죄자도 ChatGPT로 손쉽게 악성코드 생성 가능 • 다만, 오류 및 논리적 결함으로 인해 효율성이 떨어질 수 있음
멀웨어 분석	• 카스퍼스키에서는 ChatGPT를 활용해 소스코드를 빠르게 해석하는 플러그인을 개발 • 일부 결함이 있을 수 있지만, 전반적으로 분석 효율을 높이는 도구로 활용 가능
취약점 검색	• 소스코드를 읽고 취약점이 있을 수 있는 위치를 식별 – 공격자와 방어자 모두에게 유용하게 활용될 수 있음
보안 컨설팅	• 설득력 있어 보이는 사이버보안에 대한 조언을 제공 가능 – 다만, 그 출처를 알 수 없어 잘못된 조언이 포함될 수 있음
피싱과 악성메일	• 챗봇을 활용하여 성공하는 피싱의 수가 증가할 것이며, 이메일 뿐만 아니라 소셜 네트워크나 메신저 등에도 활용될 수 있음

카스퍼스키 ChatGPT (공격활용사례), 출처 : KISA Insight (2023 Vol3)

3-2. 민감정보 유출과 결과물 오남용

| 무분별한 데이터 입력으로 인한 민감정보의 유출 가능성 존재

- 사용자가 **ChatGPT에 입력한 정보**는 사용자 컨텐츠로 **OpenAI 서버에 저장**되기 때문에, 개인정보나 회사 기밀정보 등 **민감정보를 ChatGPT에 입력하지 않도록 주의 필요**
  - ☞ OpenAI 社에 발생하는 **보안 사고**, 우회 질문을 통한 민감정보 유출, 인공지능 모델이나 서비스에 대한 **해킹 공격** 등으로 인해 **사용자가 입력한 정보**가 **유출될 수 있음**

저작권 침해 개발된 생성 AI 가

무분별하게 이용하는 사람들에 의해 2차 침해를 유발 !!

아주 빠른 속도로 우리 주변의 산업 분야로 적용되기 시작했다.

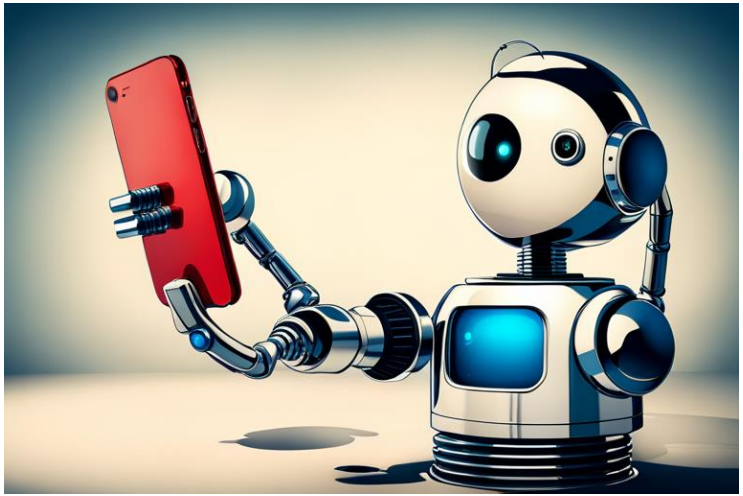
Trellix

OPEN AI 의 시대, 데이터 보안의 중요성

2022년 11월 30일 출시 + 5일 → **100만 사용자 돌파**

출시 2개월 후 (2023년 1월) → **1억명 돌파**

모든 신기술은 비즈니스 영역에서 혁신과 효율성에 초점을 맞추고 있습니다.



하지만 데이터는 어떻게 사용되며, 누가 데이터에 액세스 할 수 있고 어떻게 보안이 유지 될까요?

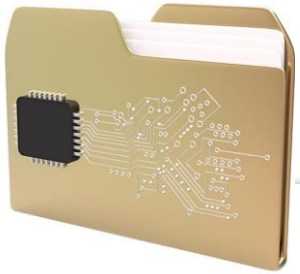
지금은 데이터 보안에 대하여 그 중요성이 더욱 높아졌습니다

Trellix

Data Flow Protection

1. 내부 데이터의 가시성 확보
2. 데이터의 흐름의 추적
3. 보안 데이터와의 결합

데이터 보안의 시작은?

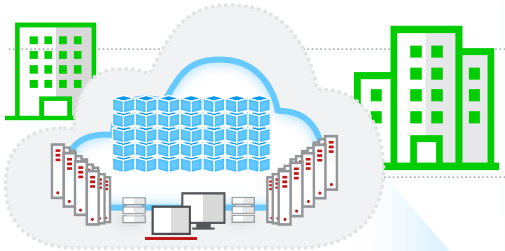


외부로 나가는 “ 모든 데이터” 의 실시간 기록

- ✓ 오래된 데이터 **탐지 정책의 개선**이 필요한데?
- ✓ 정책 **적용 전 테스트**가 필요하지 않을까?
- ✓ 나중에 **무엇인가 유출**되었는데 확인 방법은 없을까?

Trellix

데이터 가시성 확보 필요



알 수 없는 파일에 대한 식별

파일(애플리케이션) 의 상세 속성 정보

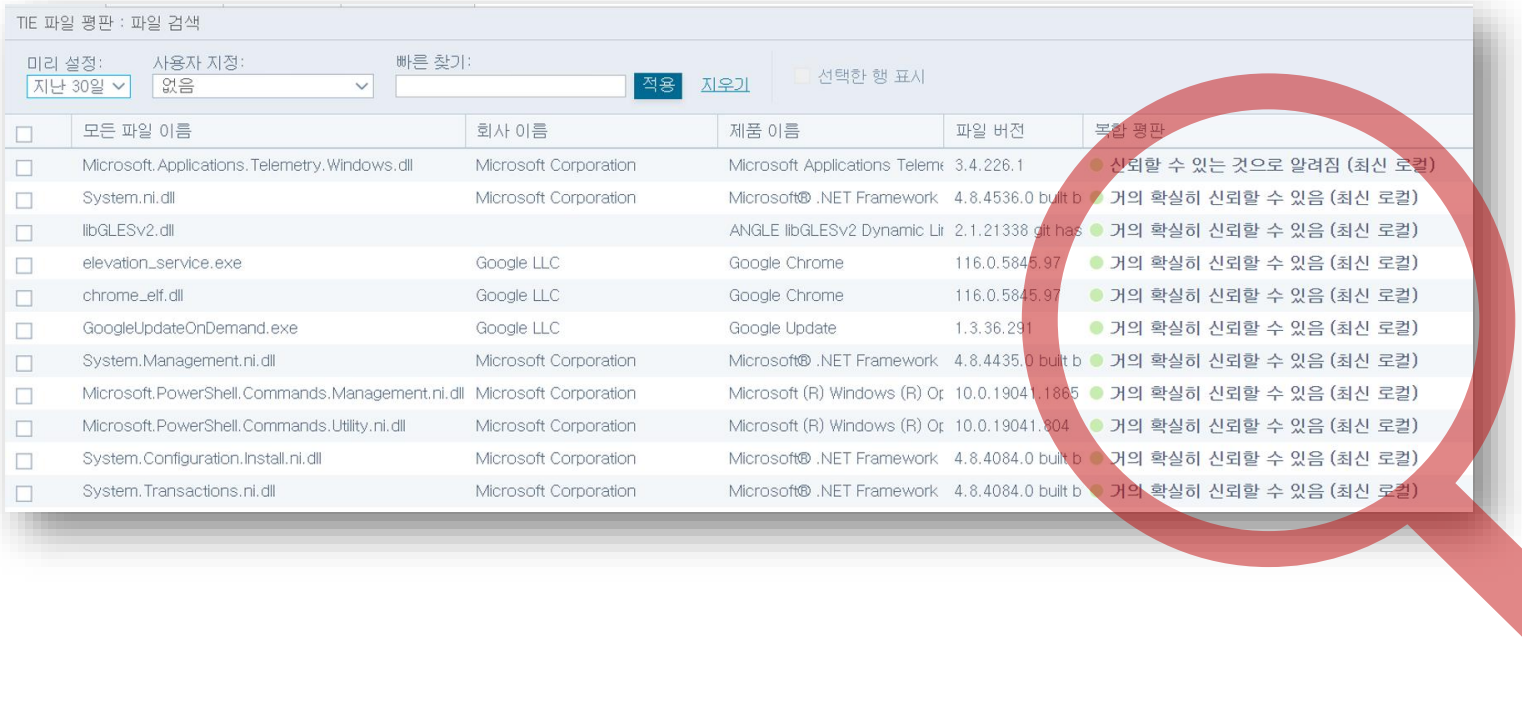
파일에 대한 자체 샌드박스 분석

단순 파일 평판이 아닌 인증서 정보 매핑

기업내 모든 시스템과 파일의 평판 정보 공유

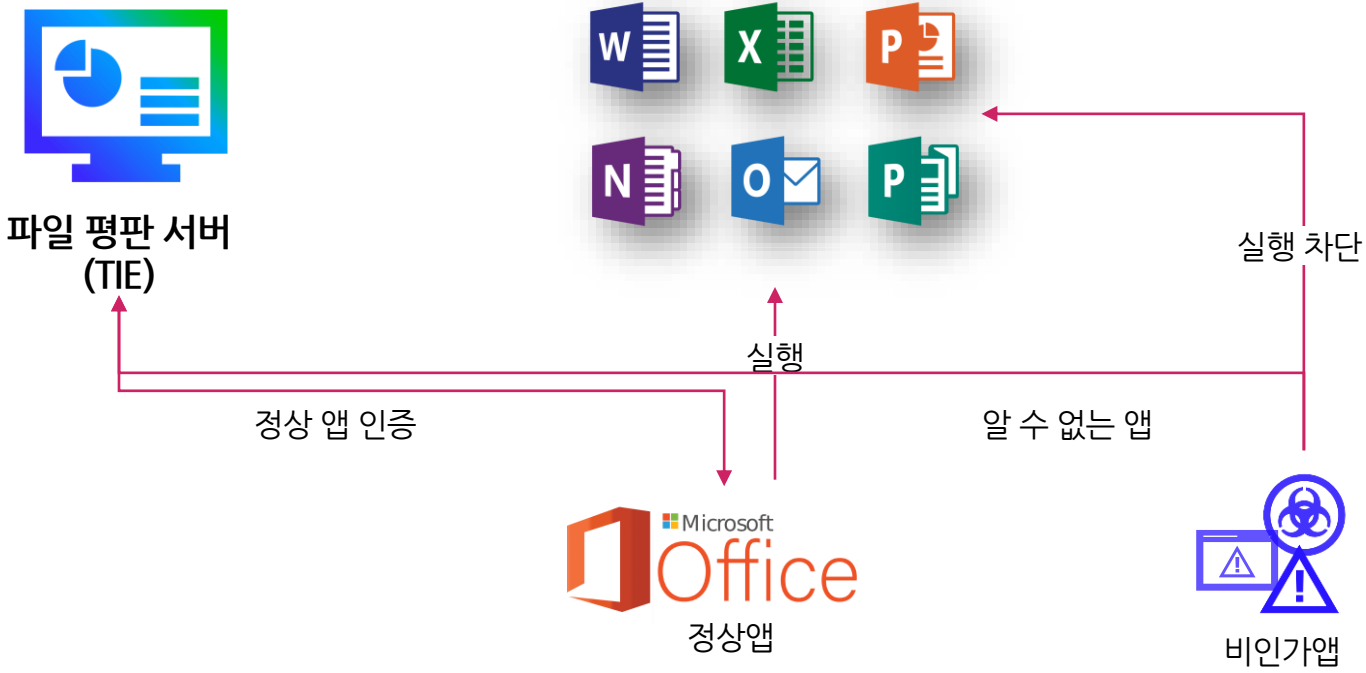
Trellix

내부 파일에 대한 검증



Trellix

비 정상적 접근 시도 탐지/차단



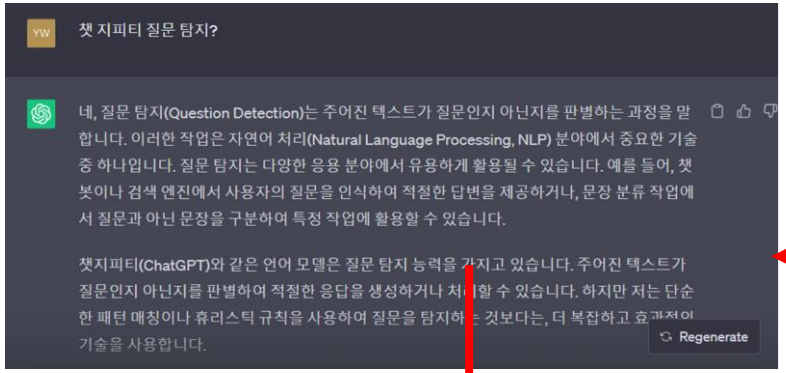
Trellix

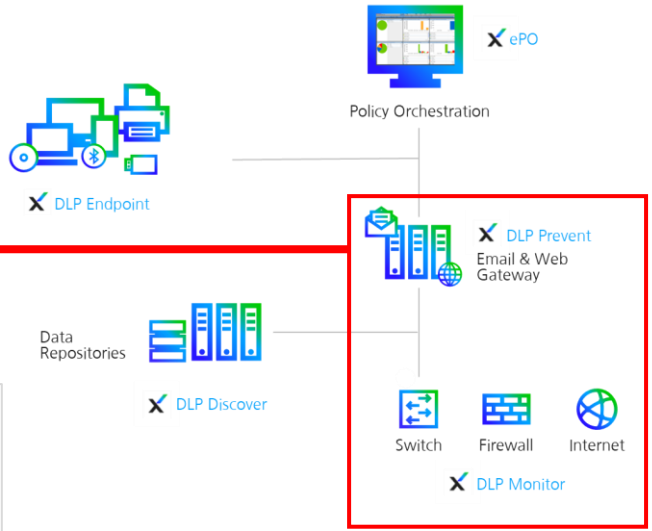
데이터 추적

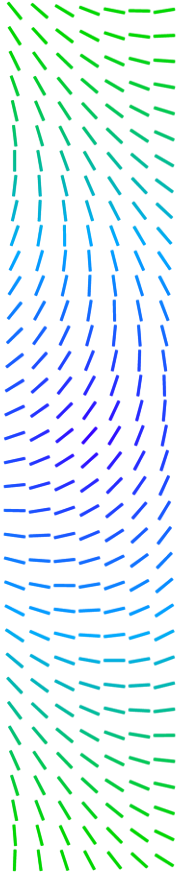
- ✓ 이전에 고려하지 않았던 **RISK**
- ✓ 트래픽 분류 , 인덱싱 및 저장
- ✓ **실시간** 규칙과 일치하는 정보
- ✓ 누가 사용하고 어디로 가는지

Trellix

네트워크 레벨에서 데이터 유출 감시 체계







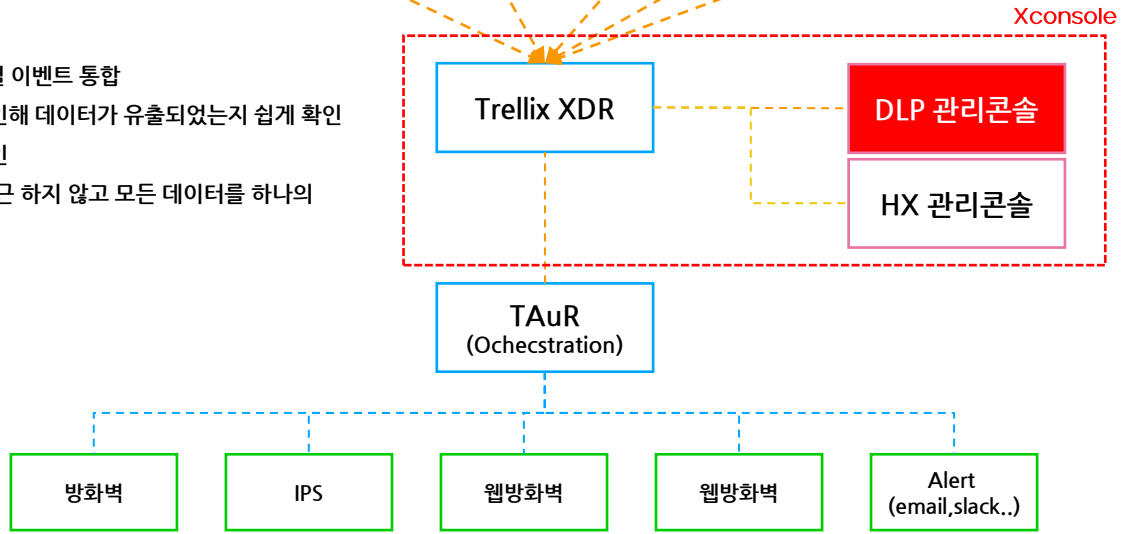
Trellix

단순 데이터 차단이 아닌 보안 관점에서 접근

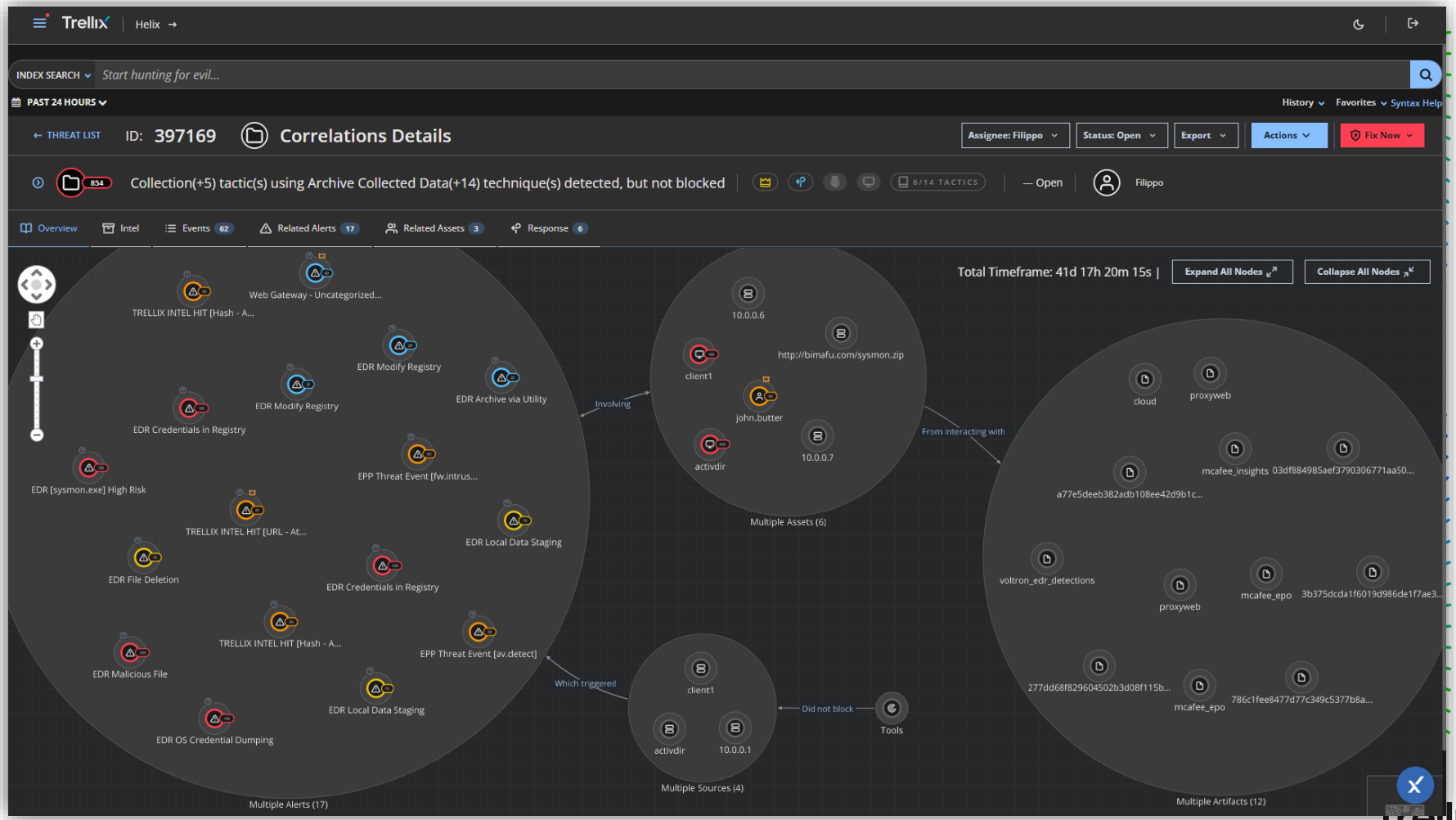
이메일 악성코드 유입 → 악성 URL 접속 → 악성코드 실행 → 감염 → 데이터 유출

ETP HX HX HX DLP

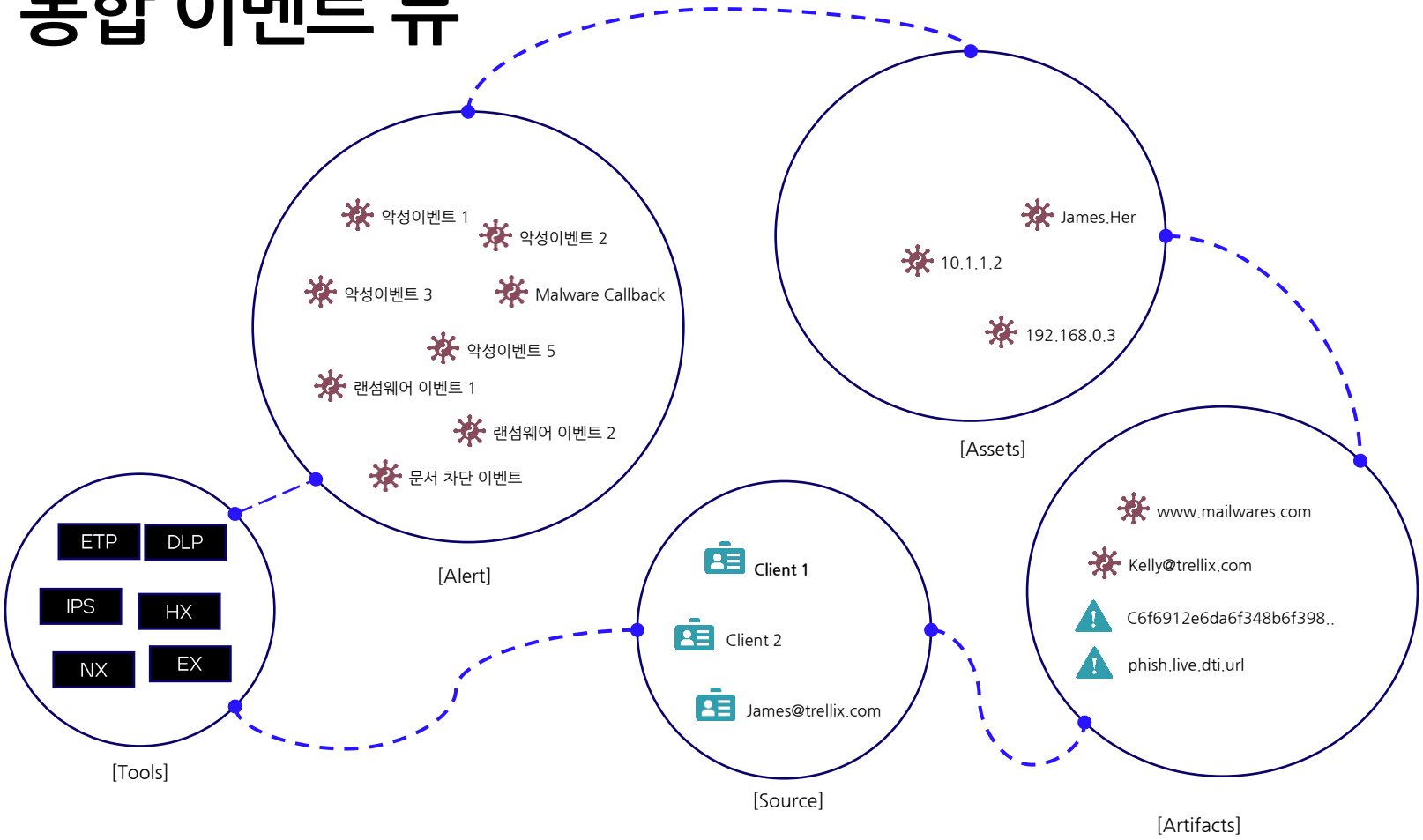
- ✓ ETP,HX,DLP 개별 이벤트 통합
- ✓ 어떠한 공격으로 인해 데이터가 유출되었는지 쉽게 확인
- ✓ 연관된 이벤트 확인
- ✓ 각 콘솔에 직접 접근 하지 않고 모든 데이터를 하나의 UI 에서 통합 관리



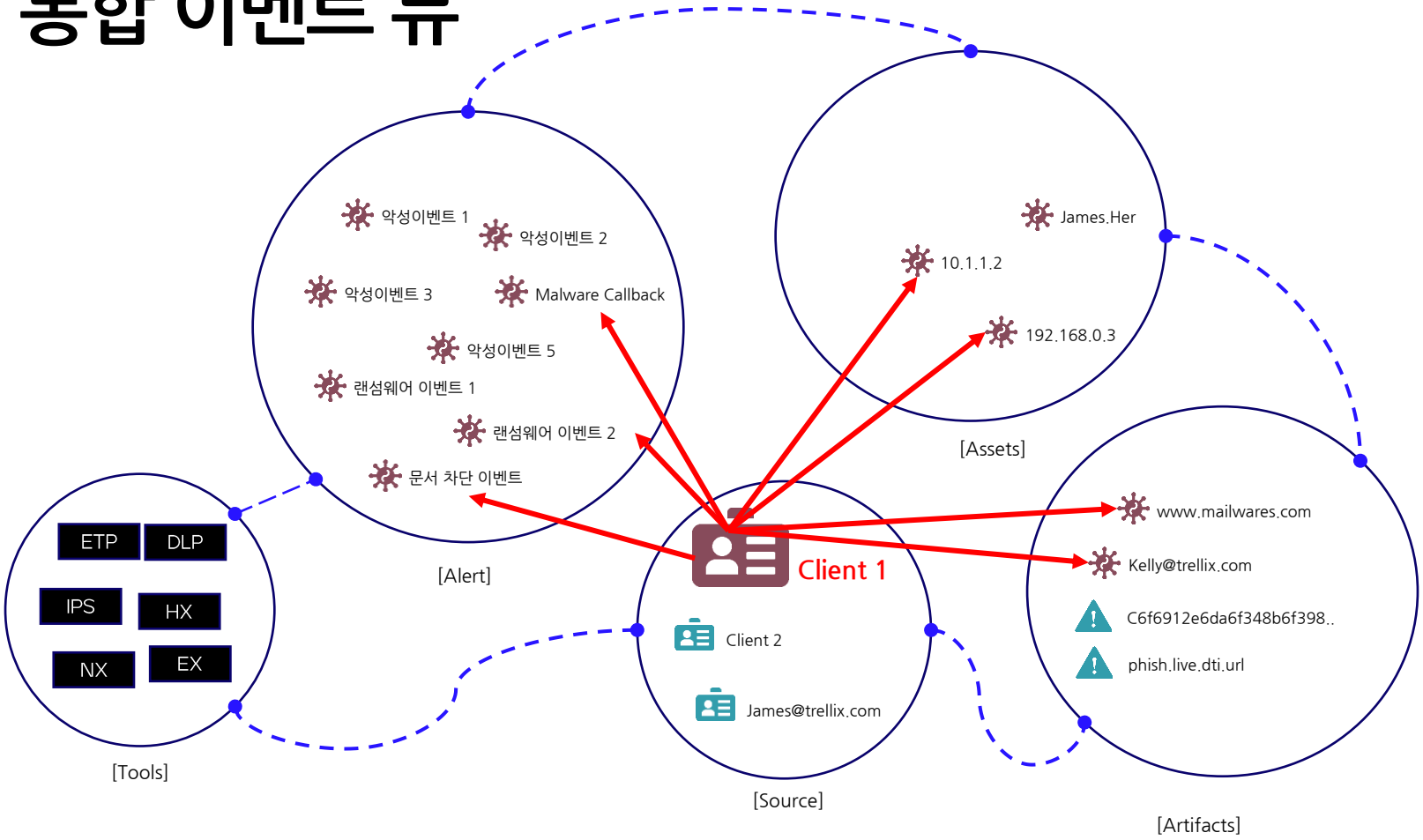
Trellix


Trellix
Helix
INDEX SEARCH
Start hunting for evil...
PAST 24 HOURS
History
Favorites
Syntax Help
THREAT LIST
ID: 397169
Correlations Details
Assignee: Filippo
Status: Open
Export
Actions
Fix Now
854
Collection(+5) tactic(s) using Archive Collected Data(+14) technique(s) detected, but not blocked
6/14 TACTICS
--- Open
Filippo
Overview
Intel
Events 62
Related Alerts 17
Related Assets 3
Response 6
Total Timeframe: 41d 17h 20m 15s |
Expand All Nodes
Collapse All Nodes
TRELLIX INTEL HIT [Hash - A...
Web Gateway - Uncategorized...
EDR Modify Registry
EDR Modify Registry
EDR Archive via Utility
EDR Credentials in Registry
EDR [sysmon.exe] High Risk
EPP Threat Event [fw.intrus...
TRELLIX INTEL HIT [URL - At...
EDR Local Data Staging
EDR File Deletion
EDR Credentials in Registry
TRELLIX INTEL HIT [Hash - A...
EPP Threat Event [av.detect]
EDR Malicious File
EDR Local Data Staging
EDR OS Credential Dumping
Multiple Alerts (17)
Involving
10.0.0.6
http://bimafu.com/sysmon.zip
client1
john.butter
activdir
10.0.0.7
Multiple Assets (6)
From interacting with
cloud
proxyweb
mcafee_insights
03df884985aef3790306771aa50...
a77e5deeb382adb108ee42d9b1c...
voltron_edr_detections
proxyweb
mcafee_epo
3b375dcda1f6019d986de1f7ae3...
277dd68f829604502b3d08f115b...
mcafee_epo
786c1fee8477d77c349c5377b8a...
Multiple Artifacts (12)
Which triggered
client1
activdir
10.0.0.1
Multiple Sources (4)
Did not block
Tools

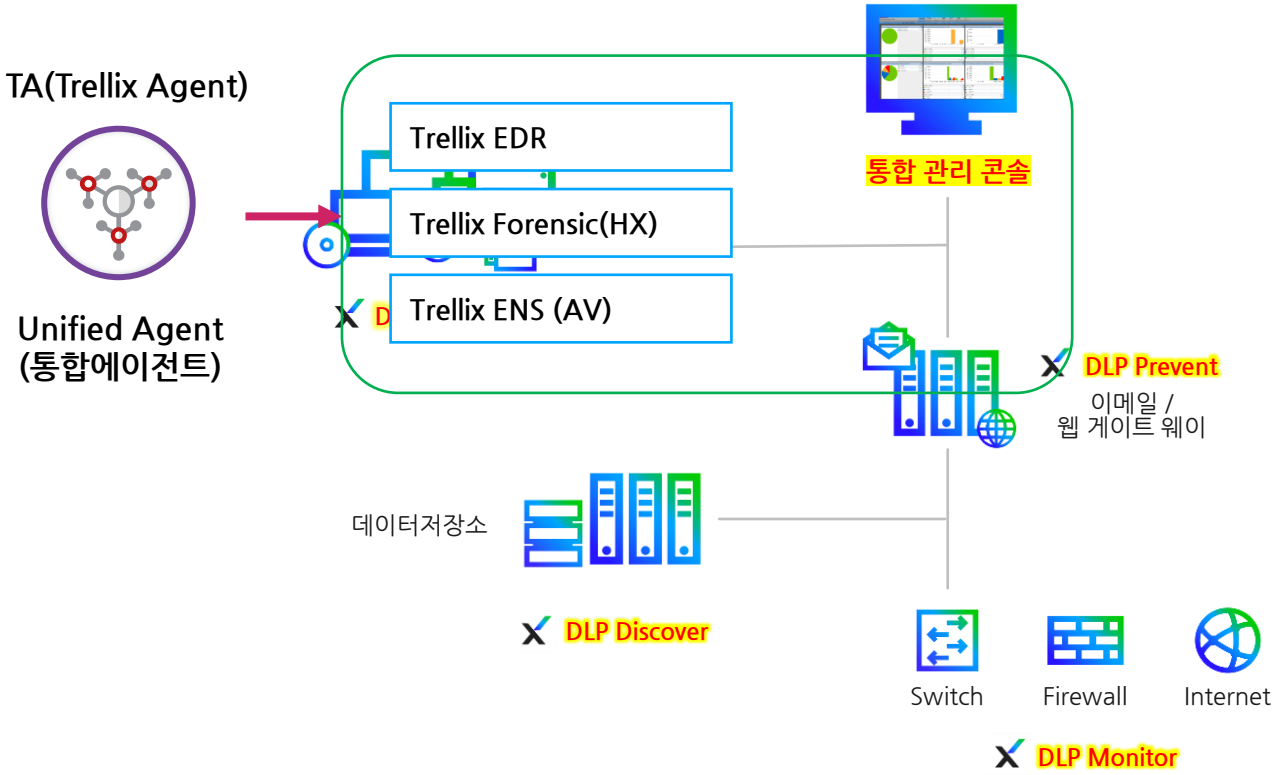
통합 이벤트 뷰


악성이벤트 1
악성이벤트 2
악성이벤트 3
Malware Callback
악성이벤트 5
랜섬웨어 이벤트 1
랜섬웨어 이벤트 2
문서 차단 이벤트
[Alert]
James.Her
10.1.1.2
192.168.0.3
[Assets]
ETP
DLP
IPS
HX
NX
EX
[Tools]
Client 1
Client 2
James@trellix.com
[Source]
www.mailwares.com
Kelly@trellix.com
C6f6912e6da6f348b6f398..
phish.live.dti.url
[Artifacts]

통합 이벤트 뷰


악성이벤트 1
악성이벤트 2
악성이벤트 3
Malware Callback
악성이벤트 5
랜섬웨어 이벤트 1
랜섬웨어 이벤트 2
문서 차단 이벤트
[Alert]
James.Her
10.1.1.2
192.168.0.3
[Assets]
ETP
DLP
IPS
HX
NX
EX
[Tools]
Client 1
Client 2
James@trellix.com
[Source]
www.mailwares.com
Kelly@trellix.com
C6f6912e6da6f348b6f398..
phish.live.dti.url
[Artifacts]

Trellix Data Protection 아키텍쳐



Trellix

Trellix

Thank you

트렐릭스 코리아 – Trellix Korea Ltd.

Office:+82-2-2092-6580
Fax: +82-2-2092-6585
E-mail: korea.info@trellix.com