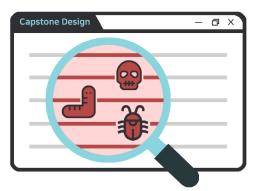
# Malware Analysis System in Kookmin



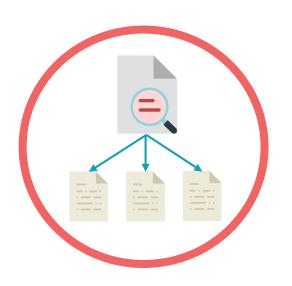
16조 - 한채연 김영재 명준우 이유정 허준녕

# NCNP

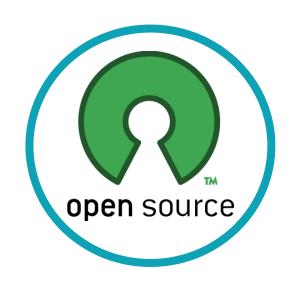
# "기하급수적으로 증가하는 악성코드"



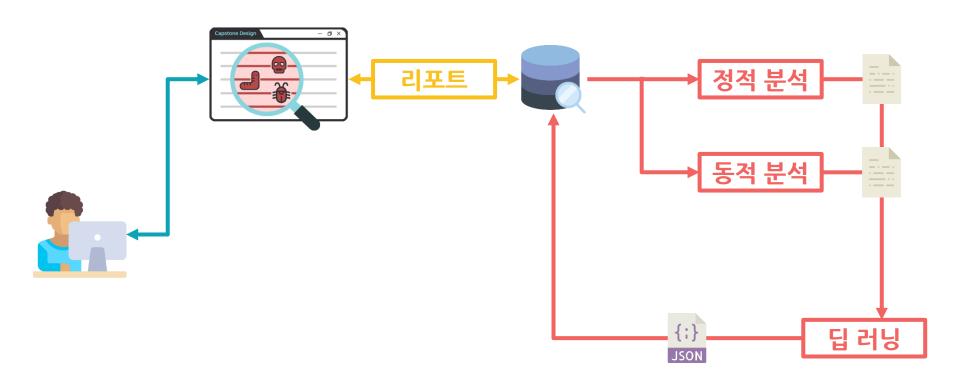
# 01.프로젝트목표







# 01.프로젝트목표



# 02.수행내용





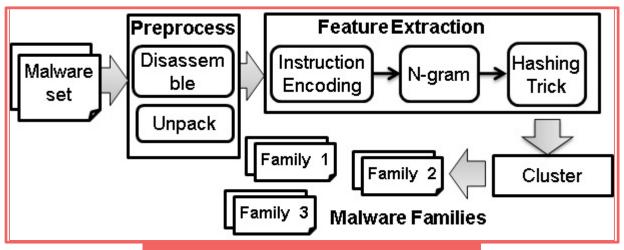






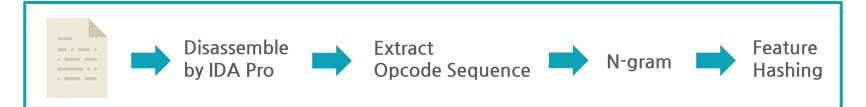


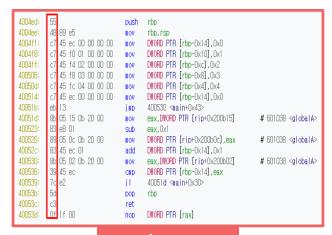




**A system overview of MutantX-S**(출처: MutantX-S: Scalable Malware Clustering Based on Static Features)







opcode sequence

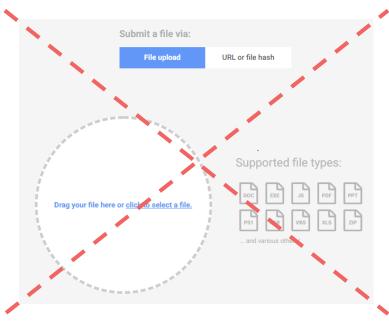




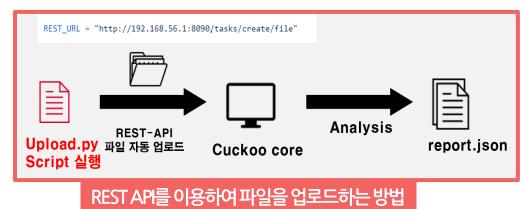
# 비정상적인 접근을 탐지하기 위해 의도적으로 설치해 둔 시스템



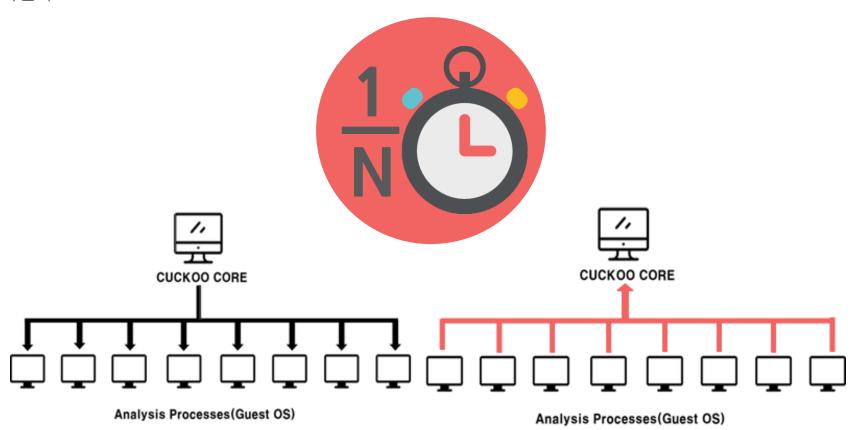




웹 인터페이스를 이용하여 파일을 업로드하는 방법









# Process memory

프로세스에 대한 메모리 덤프 분석

# 02 Target

yara rule에 의해 탐지되었을 경우 나타나는 정보이다.

```
☐ { } target

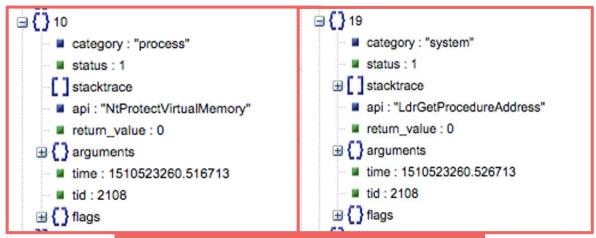
     ■ category : "file"
  ∃ { } file
     ⊟{}0
                description : "Possibly employs anti-virtualization techniques"
                author: "nex"
              name: "vmdetect"

⊕ { } offsets

           sha1: "37ea8c0215242241950aa4a803c33c7cd9e2c1c6"
        name: "71c76877401eda920f377199a9afe99e.vir"
        ■ type: "PE32 executable (GUI) Intel 80386, for MS Windows, UPX compressed"
        sha256: "59056fe589edf664ac8f730477235d53e8e1d750b2713661d5b928c67789e444"
                       · vmdetect - Possibly employs anti-virtualization techniques
            Yara
```



- 네트워크 프로토콜, 악성코드를 실행한 host 정보
- 04 정적 분석 결과(Strings..)
- 05 Behavior(API 통계, API call sequence..)



한프로세스의 10번째, 19번째 API호출 기록(리포트)



# 05 Behavior(API 통계, API call sequence..)

쿠쿠샌드박스는 아래와 같이 323개 함수의 API 호출을 기록하고, 자체적으로 17개의 카테고리로 분류한다.

class	description	example	# of APIs
Α	file/directory	CopyFile, CreateDirectory, GetFileType,	47
В	registry	RegCreateKeyEx, NtCreateKey, RegDeleteValue,	38
C	internet explorer	CDocument_write, CScriptElement_put_src,	7
D	user interface	DrawText, FindWindow, LoadString,	11
Е	net API	NetGetJoinInformation, NetShareEnum,	6
F	network	DnsQuery_A, GetAdaptersInfo, HttpOpenRequestA,	62
G	OLE	CoCreateInstance, CoInitialize, ···	3
Н	process	CreateProcess, CreateThread, Module32First,	41
I	synchronization	GetLocalTime, GetSystemTime,	8
J	resource	FindResource, LoadResource,	6
K	services	ControlService, CreateService,	12
L	system	GetNativeSystemInfo, LdrLoadDll, NtClose,	26
M	certificate	CertControlStore, CertOpenStore,	5
N	encryption	CryptCreateHash, CryptGenKey,	19
0	exception	SetUnhandledExceptionFilter, RtlDispatchException,	6
P	misc	GetUserName, GetDiskFreeSpace, WriteConsole,	20
Q	notification	_anomaly_, _exception_,	4

API Table (출처:고동우, 김휘강(2017) "API콜시퀀스와 Locality Sensitive Hashing을 이용한 악성코드 클러스터링 기법에 관한 연구", 정보보호학회논문지)





# 06 Signatures

위의 악성코드 정보들을 바탕으로 나타난 악성코드 특징(description)

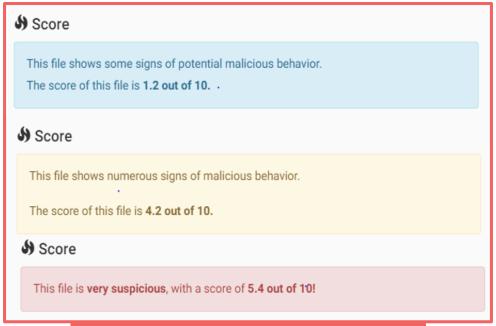
⊞ Signatures				
Queries for the computername (1 event)				
• Checks amount of memory in system, this can be used to detect virtual machines that have a low amount of memory available (1 event)				
A process attempted to delay the analysis task. (1 event)				
Drops a binary and executes it (1 event)				
Checks adapter addresses which can be used to detect virtual network interfaces (1 event)				
Potentially malicious URLs were found in the process memory dump (50 out of 124 events)				
Attempts to identify installed AV products by installation directory (3 events)				
O Deletes its original binary from disk (1 event)				
A process performed obfuscation on information about the computer or sent it to a remote location indicative of CnC Traffic/Preperations. (4 e				

웹인터페이스에서확인가능한파일에대한signatures



# 07 Score

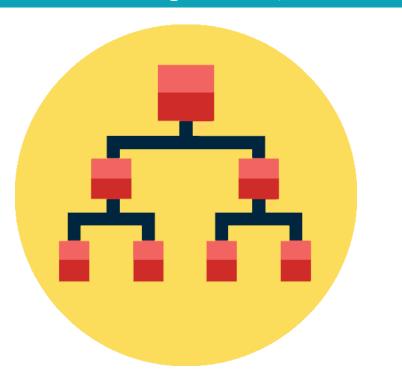
signatures로 식별한 패턴을 통해 의심스러운 평균 수준을 수치화한 정도



웹인터페이스에서확인가능한파일에대한score

# **02. 수행 내용** 라벨링

# 악성코드를 어떻게 분류할 것인가



# **02. 수행 내용** 라벨링



### The best antivirus software for Windows Home User







VB100 results from 2018-02 (latest) on Windows 7 Professional, Windows 10 Professional

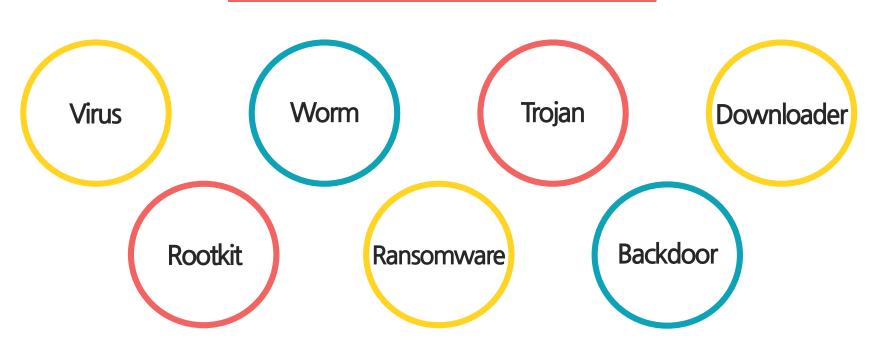
Read the full review, or download it.

SELECT INDICATORS EXPORT TO CSV										
Tested product	Result +	RAP Overview	WildList (%)	WildList (%)	False positives	False positives				
Kaspersky Lab K Kaspersky Endpoint Security 10 for Windows	Passed virus 100		100.00	100.00	0	0				





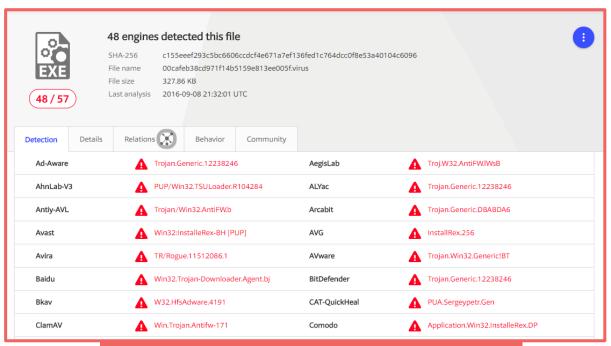
# "악성코드는 일반적으로 7가지로 분류할 수 있다."







# 01 카스퍼스키 라벨 가져오기



바이러스토탈에서 약60여개의 안티바이러스의 분석 결과





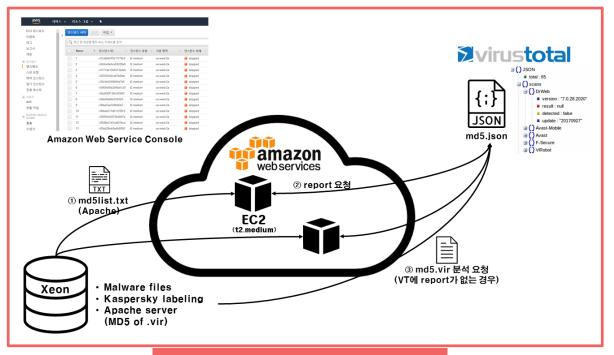
# 카스퍼스키 라벨 가져오기



바이러스토탈에서 제공되는 API를 통하여 받을 수 있는 json 형식의 분석 결과

# **02. 수행 내용** 라벨링

# 대량의 악성코드의 바이러스토탈 리포트를 가져오기



AWS EC2 인스턴스를 활용한 분산 분석

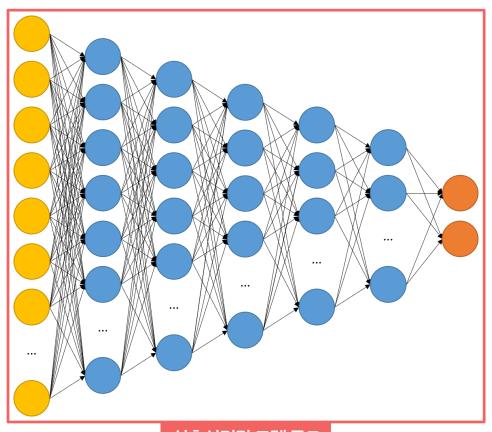
# **02.수행내용** 답러닝





# **02. 수행 내용** <sub>딥러닝</sub>





심층신경망모델구조

# **02.수행내용** 데이타처리

대량의 악성코드 데이타를 어떻게 운용하여 데이타 사용에 대한 시간 비용을 최소화 할 것 인가?



# **02.수행내용** 데이타처리-DBMS구축

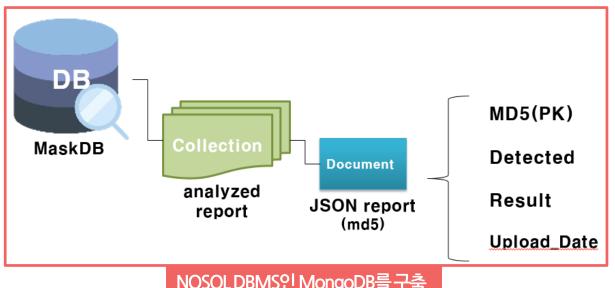


# DB의 유연성

실험적인 Data 사용 및 저장을 위한 DB

# Read 연산 위주의 DB

서비스 중 DB 접근은 주로 Data 검색을 위해 사용



NOSQL DBMS인 MongoDB를 구축

# **02.수행내용** 데이타처리

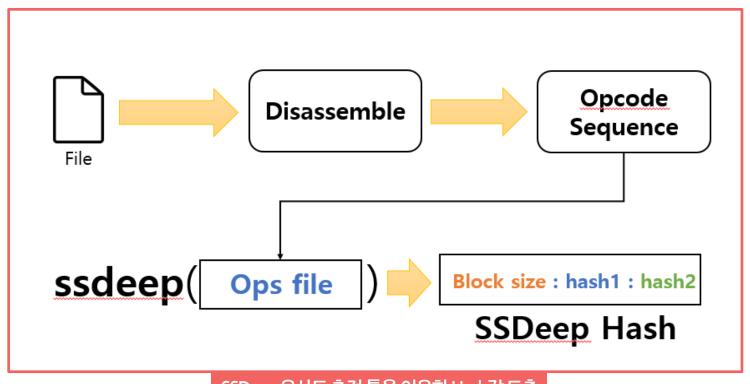
# 업로드 한 파일과 기존 DB에 있는 파일에서 유사한 파일을 어떻게 찾아서 보여줄 것인가?



**02. 수행 내용** 데이타처리-유사한파일 검색

# **SSDeep**

# **이 어떤 방법으로 유사도 측정을 할 것인가?**

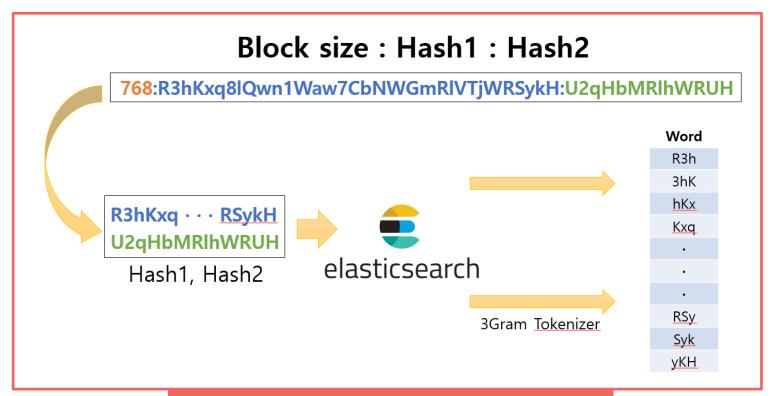


SSDeep 유사도 측정 툴을 이용한 Hash 값 도출

# **02. 수행 내용** 데이타처리-유사한파일 검색



# ◉◢ 도출된 Hash값 간의 유사도 비교를 위한 전 처리

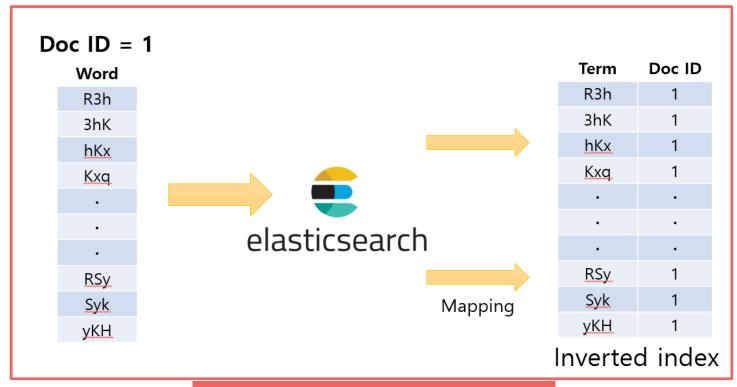


역인덱싱의 전처리 위해 N-Gram Tokenize 중 3-gram을 이용

# **02.수행 내용** 데이타처리-유사한파일 검색



# 05 빠른 검색위한 방법

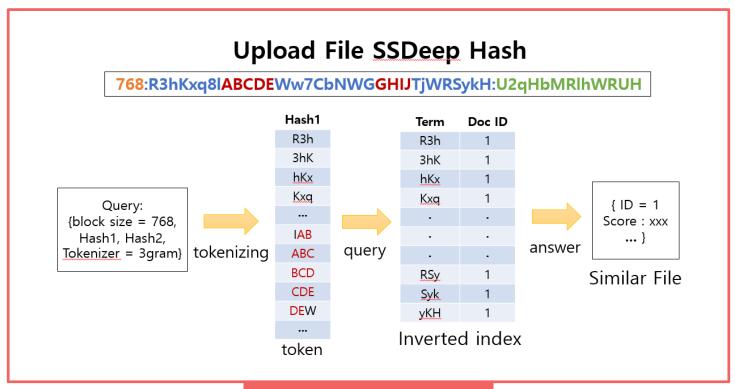


Tokenize된 word의 빠른 검색을 위해 역인덱싱

# **02. 수행 내용** 데이타처리-유사한파일 검색

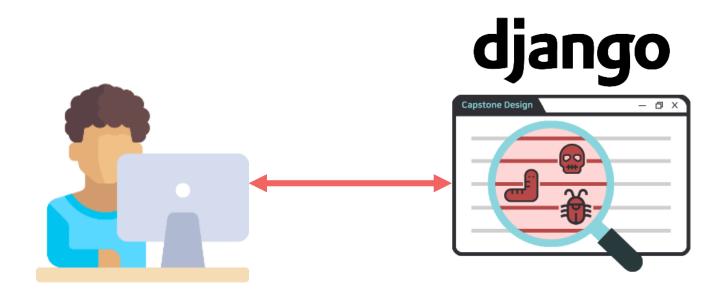


# 업로드 한 파일과 유사한 파일 검색



유사한 파일을 찾는 질의의 과정

# **02.수행내용** 웹



# 02.수행내용 <sub>웹</sub>

# O Bootstrap을 이용한 반응형 웹 제작



브라우저의 폭 768px 이상일 때웹 초기화면



# **02.수행내용** 웹

# 02 파일 업로드 방법



# **02.수행내용** 웹

# 03 분석 결과 제공

NCNP: MASK 분석	통계	Sign in
파일 정보		
	파일 MDS virussign.com_00cee1918c8925a2a7782f1a0d5759c7 바이러스 유무 True 바이러스 분석 결과 None 리포트 수집 날짜 None	
AI SAFE	100%	
시그니저 (Primary)		
유사 파일		

사용자에게 보여지는 분석 결과 화면

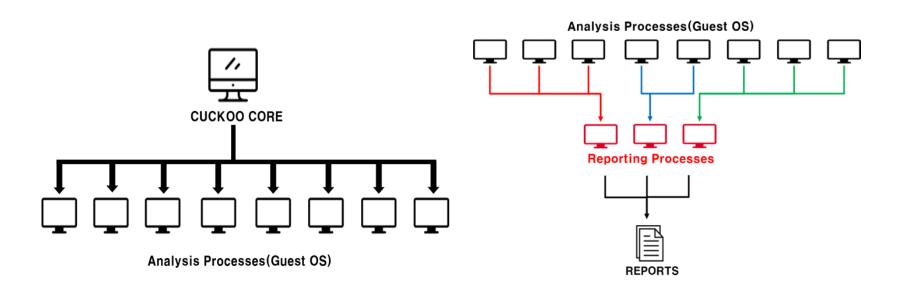
# 03.수정사항

# 크롤링자동화〉〉수동화

크롤러를 개발하여 악성코드를 수집하려 하였으나 유료서비스나 토렌트 등을 이용해야 하기 때문에 자동화가 어려움 수집 채널을 수동으로 변경함

# **04. 향후 추진계획** 동적 분석

# 분석 프로세스와 리포팅 프로세스를 분리함으로써 시스템의 안정성을 높임



# **04. 향후 추진계획**

## 더 정확한 악성코드의 특징에 따른 분류를 위해 세분화된 라벨을 제작

현재 사용중인 7가지의 라벨을 기준으로 잡고 세분화 악성코드의 정적/동적 분석 정보에서 뽑은 특징점을 이용한 실험

### AWS 인스턴스를 이용한 악성코드 분산 분석 자동화

인스턴스를 프로그램을 통해 조작 (python boto3 라이브러리…)

인스턴스에 악성코드를 분산 업로드 (신규 분석 요청시 인스턴스에서 직접 파일 보냄)

# **04. 향후 추진계획**

### 피처 해싱을 다양하게 해볼 예정

다양한 기법으로 피처 해싱을 시도하여 모델의 성능 향상을 시도함

### 하이브리드 모델 개발

동적 분석 결과로부터 추출된 피처도 적용되는 동적 분석 모델과 정적 분석 결과와 동적 분석 결과를 복합적으로 사용하는 하이브리드 모델 개발을 목표로 함

# **04. 향후 추진계획** 웹

## 파일의 정적, 동적, 혼합 분석 기능 제공

제공중인 정적 분석을 포함하여 동적, 혼합 분석 기능을 모두 제공할 예정

### 분석 결과 화면에 정보 추가

동적 분석 시에 나타나는 파일의 C&C 서버를 시각화 하여 제공 매일 분석되는 악성코드의 수와 발견되는 종류를 도식화 한 표 등 각종 자료 제공

# THANKYOU CAA