


Code Engn 11

세인트시큐리티
김 기 홍
kog@stsc.com



빅 데이터 기반 악성코드 자동 분석 & 시스템

1 Day



250,000

신규 악성코드



2,500,000

경유지



50,000

악성코드
유포지/배포지 주소

1 Day



250,000

신규 악성코드



증가 추세!

2,500,000

경유지



50,000

악성코드
유포지/배포지 주소

그냥 단순하게 산술적으로 계산 해보면...



= 25

1 Super Geek

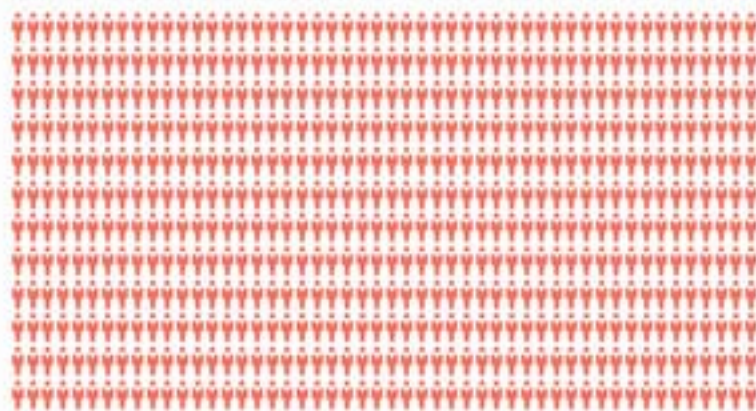
바이너리 분석



= 50

1 Super Genius

바이너리 분석



1,000 Super Genius

=  = 50,000

바이너리 분석

국내 1위 보안 업체
전체 직원수

그냥 단순히 산술적으로 계산 해보면...



15개 회사

= 1 Day

사실상,
모든 파일을 일일이 분석한다는 것은
현실적으로

불가능.

그럼 어떻게 할 건데?!



빅 데이터(Big Data) & 집단 지성



빅 데이터(Big Data) & 집단 지성

얼마나 커야 **빅 데이터**라고 할 수 있을까?

malwares.com 1일 기준



약 130,000,000
총 누적 샘플 수집 량



500,000
신규 파일 수집



2,000,000
URL 크롤링



malwares.com 1일 기준



약 130,000,000

총 누적 샘플 수집 량



10Tbyte

개당 약 15~20Mbyte 파일
50만개



2,000,000

URL 크롤링

10Tbyte?



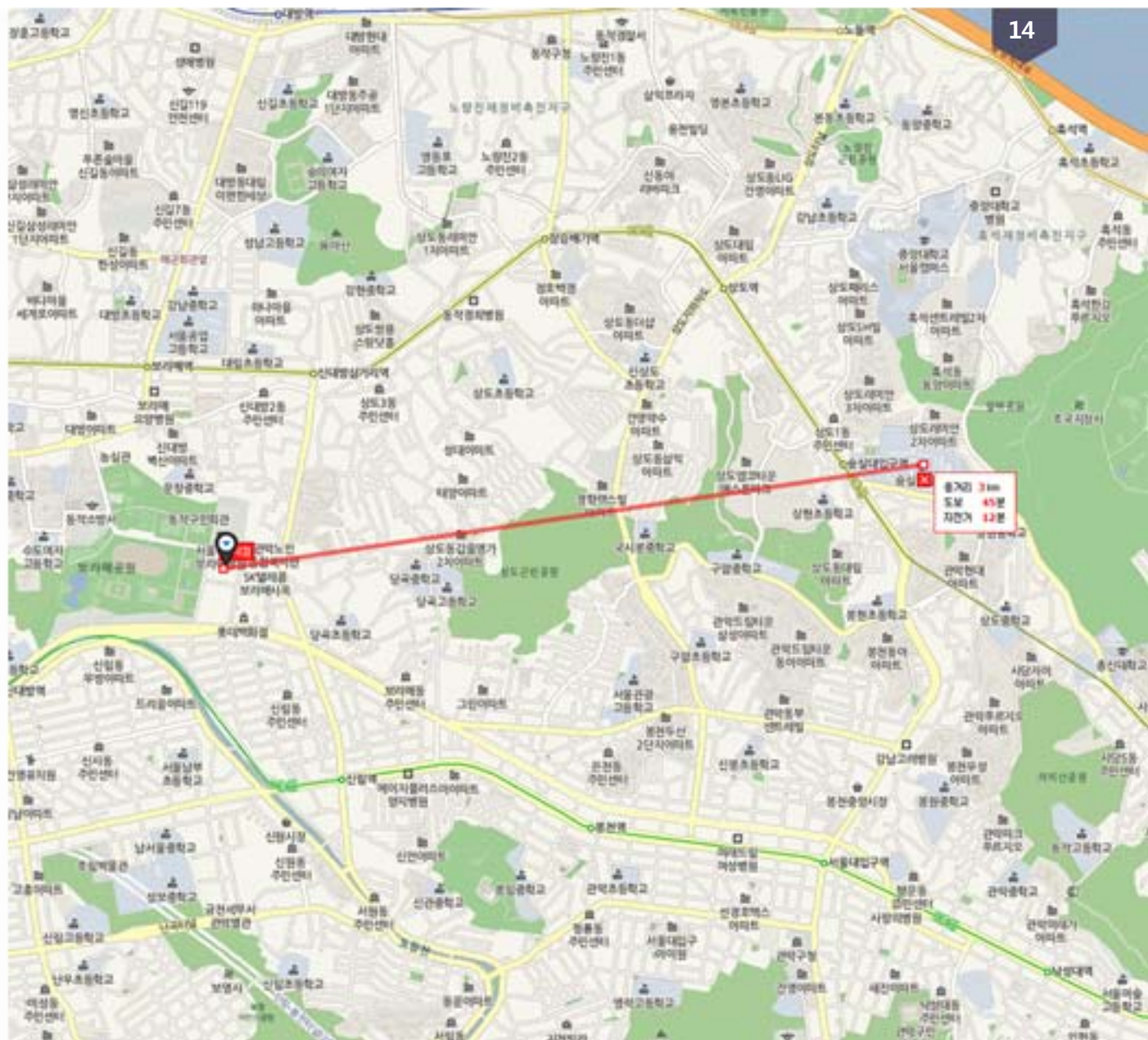
iPhone 6+ **128Gbyte** X **80개** = **10Tbyte**

매일 iPhone 6+ **128Gbyte** **80개** 분량의 데이터가
생성되고 있는 중

2014년 4월 1일 OPEN
 오늘까지 243일.
 7개월 28일.

iPhone 6+ 128Gbyte
 19,440개 소비중.

iPhone 6+
 $19,440 = 3\text{Km}$



```
select * from mws_total_sample_hash_list;
```

필요한 내용을 전부 쪼개서 저장할 수 있어야 함.
RDB 사용은 거의 불가.

NoSQL 사용.
Amazon Elastic Search

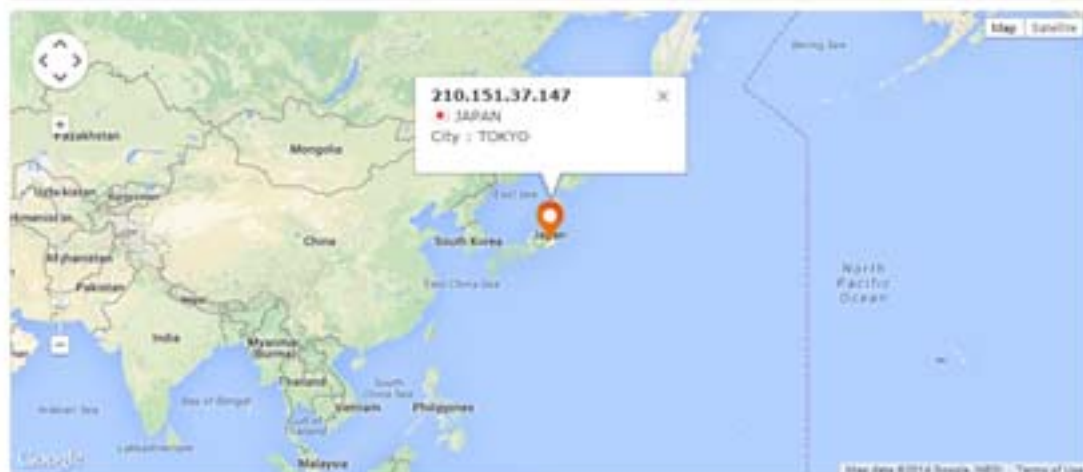
무엇을 저장하고
관리할지 정해보자!

- ▶ 파일명
- ▶ A/V 탐지 결과(히스토리 관리)
- ▶ MD5, SHA1, SHA256
- ▶ 정적 분석 결과
 - Strings
 - PE 구조 등 파일 구조 특화된 정보 추출
 - YARA 룰 매칭 정보
 - 바이너리 자체에 패턴 매칭 정보
- ▶ 동적 분석 결과
 - 가상머신 Windows XP / 7 행위 분석 결과
 - 리얼머신 Windows XP / 7 행위 분석 결과
 - 행위 분석 시 발생한 pcap / Screenshot
- ▶ 네트워크 수집 정보 연동
 - 배포지
 - C&C 통신 대상

이렇게 분석해 보니 어떤 결과가 나왔지?

Case A. 그냥 봐줄 만한 녀석들

210.151.37.147



Malicious URL history of this IP

No.	URL	Anti-virus	Scan Date
36	http://210.151.37.147/82/1125/jauned.exe	3 / 41	2014-11-25 10:36:47
35	http://210.151.37.147/A2/1125/jauned.exe	2 / 41	2014-11-24 20:20:30
34	http://210.151.37.147/92/1125/jauned.exe	2 / 41	2014-11-24 19:36:18
33	http://210.151.37.147/882/1125/jauned.exe	2 / 41	2014-11-24 18:48:43
32	http://210.151.37.147/1x2/1125/jauned.exe	2 / 41	2014-11-24 18:48:42
31	http://210.151.37.147/52/1125/jauned.exe	2 / 41	2014-11-24 18:45:25
30	http://210.151.37.147/72/1125/jauned.exe	2 / 41	2014-11-24 18:40:57
29	http://210.151.37.147/712/1125/jauned.exe	2 / 41	2014-11-24 18:39:56
28	http://210.151.37.147/A2/11231/jauned.exe	2 / 41	2014-11-23 23:29:02
27	http://210.151.37.147/92/11231/jauned.exe	6 / 41	2014-11-23 20:49:00
26	http://210.151.37.147/712/11231/jauned.exe	6 / 41	2014-11-23 20:43:16
25	http://210.151.37.147/52/11231/jauned.exe	2 / 41	2014-11-23 19:42:57

http://smsvn.org



Malicious URL history of this Hostname

No.	URL	Anti-virus	Scan Date
5	http://smsvn.org/astore/anh-1x.apk	10 / 10	2014-09-19 23:29:13
4	http://smsvn.org/astore/1ho-anh.apk	10 / 10	2014-09-19 23:28:51
3	http://smsvn.org/astore/anh-18.apk	10 / 10	2014-09-19 19:29:56
2	http://smsvn.org/xen/android/QUAY-LEN-9X.apk	6 / 10	2014-09-10 17:10:10
1	http://smsvn.org/xen/android/XEN-CLIP-NONG.apk	7 / 10	2014-08-09 23:22:42

IP usage history of the Hostname

No.	IP	Country	City	Last Check Date
2	123.30.50.125	VIET NAM	HANOI	2014-11-26 06:09:27
1	123.30.50.188	VIET NAM	HANOI	2013-09-07 13:00:00

Case B. 너와 나의 집념 싸움

같은 HASH 파일의 배포 주소
10월경 서버 변경(이때쯤 ISP를 통한 차단이 이루어짐)
최근까지 배포 중.

98	http://205.164.5.236:8914/ex/142001.exe	2004-11-25 11:26:22
97	http://205.164.5.236:8914/ex/141762.exe	2004-11-25 11:27:30
96	http://205.164.5.236:8914/dl/775076.exe	2004-11-25 11:22:29
95	http://205.164.5.236:8914/ex/141758.exe	2004-11-25 11:04:41
94	http://205.164.5.236:8914/ex/141754.exe	2004-11-25 10:43:05
93	http://205.164.5.236:8914/ex/141752.exe	2004-11-25 10:32:36
92	http://205.164.5.236:8914/ex/141993.exe	2004-11-25 10:23:28
91	http://205.164.5.236:8914/dl/775075.exe	2004-11-25 10:12:09
...		
66	http://205.164.5.237:8914/dl/774856.exe	2004-11-05 18:59:20
65	http://205.164.5.237:8914/dl/774844.exe	2004-11-04 18:47:57
64	http://205.164.5.237:8914/dl/774832.exe	2004-11-03 20:01:23
63	http://205.164.5.237:8914/dl/774820.exe	2004-11-02 18:06:09
62	http://99.46.84.80:8914/dl/774796.exe	2004-10-31 18:35:21
61	http://99.46.84.59:8914/dl/774783.exe	2004-10-30 18:07:26
60	http://99.46.84.59:8914/dl/774796.exe	2004-10-29 18:42:31
59	http://99.46.84.59:8914/ex/136186.exe	2004-10-29 18:03:33
58	http://99.46.84.59:8914/dl/774784.exe	2004-10-24 19:14:53
...		
5	http://99.46.84.59:8914/dl/773839.exe	2004-08-07 19:41:23
4	http://99.46.84.59:8914/dl/773838.exe	2004-08-07 18:18:57
3	http://99.46.84.59:8914/ex/115919.exe	2004-08-07 18:08:52
2	http://99.46.84.59:8914/dl/773828.exe	2004-08-06 22:11:35
1	http://99.46.84.59:8914/dl/773827.exe	2004-08-06 20:58:04
0	http://99.46.84.59:8914/dl/774084.exe	2004-08-06 20:52:07

Case C. 또 다른 집념의 소유자

2014년 7월에
갑자기 바이너리 배포

하루에 2-3번씩
바이너리 바꾸면서
계속 배포

주소는 같지만
HASH값을 계속 다르게해서 배포함.

142.0.131.57/mbc.exe

20

URL Detection

Additional Info

Malicious samples downloaded from this URL

No.	SHA-256	Anti-virus	Scan Date
34	627F73589406ED8E1AD1E330EC943A84888CCE1E38D0207866A8B04E7830A3D	36 / 11	2014-07-11 12:06:29
33	848D3C05C62D3C8F8F8F400E8C6354D191E1FF33279432D64080606E4363D	36 / 11	2014-04-30 10:00:18
32	81CC3AF4E9E1390E78E2C148497888953351091E40481E131E18981D898F90	25 / 11	2014-04-15 09:17:16
31	20F46AAE04D0450290492A790F98F3C4E06F1185612185D40087354102C45AA	37 / 11	2014-04-30 10:12:11
30	8406ED45C3E3580E21279F20991A42044ED13CE5AA8128C119130832A3D96C	24 / 11	2014-04-15 09:44:35
29	F325CBFF8718CE3331E88AD78240F43013NE15A1E27E81F70606E1F6C8F8927D	22 / 11	2014-04-15 10:13:22
28	81540483CE4E1841883CF10A75CED3C864887E8E21FF30058F3C6225FF188C91D	37 / 11	2014-04-30 10:44:12
27	26FA713CCE8DE9DA4112218D4003A059E821C02F7E7327E79831E3F98130ED7	37 / 11	2014-04-30 10:12:12
26	83F830A083F985554E19AF3FCE3A0B093E13AC4881E807897CC344E1882F190	23 / 10	2014-04-15 09:19:19
25	E38841DA48F5D3AF4F8D990CF3905D6E58C18A4D03ED24F0E8298240817E7E	22 / 11	2014-04-15 10:12:16
24	82C0793F8DCF04E84C1482E8A78ADD31848D887136A20F84DF021323F7A00E1	37 / 11	2014-04-30 10:10:18
23	471E9856E0564E1D7EF96352518AE7F94AE814408901D37C81818ACAA05927B	37 / 11	2014-04-30 10:17:44
22	08A70A5096C8C954104022753291E817B18F5BC88E6449A2C18386705F06DD	35 / 11	2014-04-30 10:15:18
21	8164FFDCEAAED2773A9FCC5C32F298DAF8D81980F81D111F88F3FC648843E	30 / 11	2014-04-14 09:23:31
20	813505D75240D41888F2819A811F488BC0FF815F11B458F4E74E3C1196088D54	10 / 11	2014-04-14 09:15:19
19	348CA43D347C2713C033989674FAA388D8F3FC9FC483F087C0561882181D8ED	36 / 11	2014-04-30 10:14:41
18	8D8831080A6684681908CE838738815E181AC42D881904E3A60F148D437E43CD	10 / 11	2014-04-14 08:41:22
17	232F30F840CF838973AAC02C1716065283889673109FA1F545710F0480E342F	34 / 11	2014-04-30 10:11:41
16	79E7773F3F323072E116A380C1F4CE79CB871943C8C60E692F6A084C27D48C55	35 / 11	2014-04-30 10:46:47
15	FF0C88FE198C6D5F8A131C8AD13D64193C0F8A3190DE45A987F8C89571746CB	10 / 11	2014-04-14 08:25:10
14	6321890570EE40F4800131896C3268973AC9E131A4217AD41D8617A81F43CA	10 / 11	2014-04-14 08:11:17
13	1624F58C7F88D70883C9C03DA1758F785307190AA2A318A70882913C083A1	35 / 11	2014-04-30 10:12:16
12	A5CC02325F8184992940000080C3DF9FED68947C6AC1883343006ABC1CC418	35 / 11	2014-04-30 10:15:12
11	AA7A18E188D405724380E008FE8F6AC40CEFA0013295DA1079C263EAB8E1E8	35 / 11	2014-04-30 10:18:05
10	8A8268139871318CF72D143181A88903190C81LAC021E1D0ED68181AF871298E	35 / 11	2014-04-30 10:10:08
9	D4F4FA31FDF73DF1202785A33207EAD677C80DF09E904993475C68582FAAC939	10 / 11	2014-04-14 07:18:15
8	6CE73F0C3C88D0C8DE2EFC1AF43CB328F87838A84E3A8C2F395D67C2E88011	18 / 11	2014-04-30 10:43:12
7	88E9177AA684268874438C381317A745FC2542538BA1E14FD84180064F8E3E40	10 / 10	2014-04-14 07:06:42
6	814362939E6E8D010138AC18DE4A6C1FD6193CB189F9545737679C40E149828	36 / 11	2014-04-30 10:43:18

Case D. 애쓰다...

**F2499A1F54E3C6AC088FBAECCC0610D5A43196D1083E3BB051CB74
405DA32E27**

MD5 : F2479008F3284EEF0A51E774EE026040
SHA-1 : 45421048690EC1863FFD0015F974E1D9E0AD7022
SHA-256 : F2499A1F54E3C6AC088FBAECCC0610D5A43196D1083E3BB051CB74405DA32E27
File Size : 132,408 bytes
Extension : exe_32bit
Packer : UPX
Scan Date : 2014-11-24 09:50:14 (1 day 21 hours ago)

Tag exe_32bit packing service executable autorun unpacking process bug msg

Analysis Environment

Behavior analysis history 2014-11-25 (Windows XP KOR) ▾

OS : Microsoft Windows XP Professional Service Pack 3 (KOR)
Installed Programs : Microsoft Office Professional 2010 / Windows Internet Explorer 8 / Nango! 2010 / Adobe Reader 10.1.0 / Adobe Flash Player 11 Active X / Java(TM) 7

Detection Summary

PID	Process path	Detected suspicious actions
208	\\MS_SAMPLE_DISK\jshwncz.exe	UAC function off
1388	c:\documents and settings\mwise\application data\msallat.exe	Detection of window host file alteration behavior
1388	c:\documents and settings\mwise\application data\msallat.exe	UAC function off
1388	c:\documents and settings\mwise\application data\msallat.exe	Program auto-run registration
1388	c:\documents and settings\mwise\application data\msallat.exe	Window service registration

System Behavior

[208] <MS_SAMPLE_DISK\jshwncz.exe>

Modified Host File

Host

IP	Domains
127.0.0.1	threatsense.net
127.0.0.1	www.threatsense.net
127.0.0.1	www.zna.com.ar
127.0.0.1	zna.com.ar
127.0.0.1	store.ca.com
127.0.0.1	avira.com
127.0.0.1	www.antiVir.com
127.0.0.1	antiVir.com
127.0.0.1	www.antiVir.com.tr
127.0.0.1	www.avg.com
127.0.0.1	avg.com
127.0.0.1	www.scanwith.com
127.0.0.1	scanwith.com
127.0.0.1	www.avast.gen.tr
127.0.0.1	avast.gen.tr
127.0.0.1	www.avast.com
127.0.0.1	avast.com
127.0.0.1	forum.avast.com
127.0.0.1	www.nod32.com
127.0.0.1	nod32.com
127.0.0.1	novirusthanks.org
127.0.0.1	novirusthanks.org
127.0.0.1	vscan.novirusthanks.org
127.0.0.1	virustotal-uploader.en.softonic.com
127.0.0.1	virscan.org
127.0.0.1	pandasecurity.com
127.0.0.1	www.arcaBit.com
127.0.0.1	arcabIt.com
127.0.0.1	www.arcabIt.pl

Case E. 호스팅 서버 탈취

호스팅 서버를 탈취해서
서버 IP는 동일하지만
여러 호스트명과 서버 주소를 갖고
악성코드를 배포함

101.79.5.66

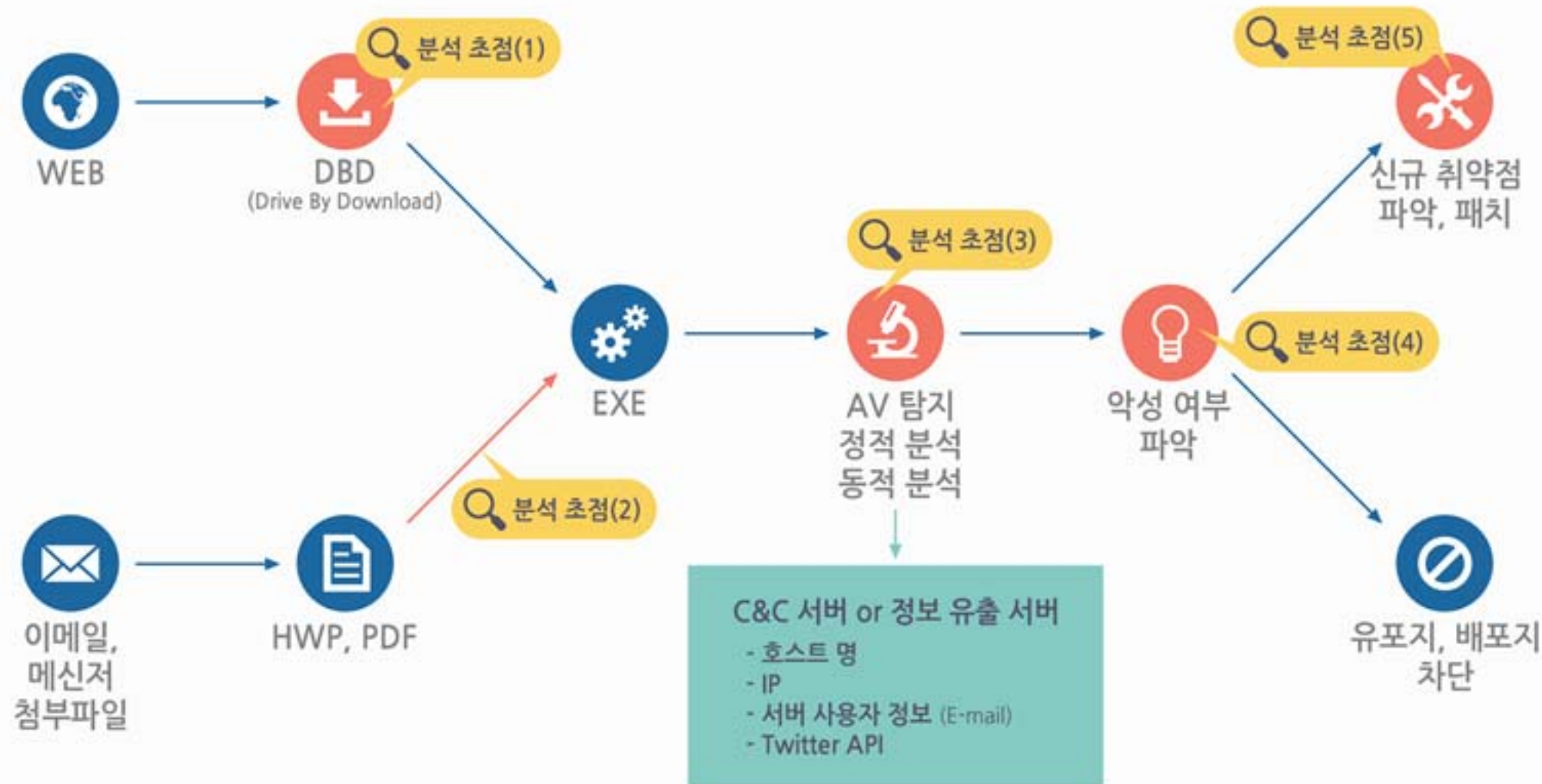
22



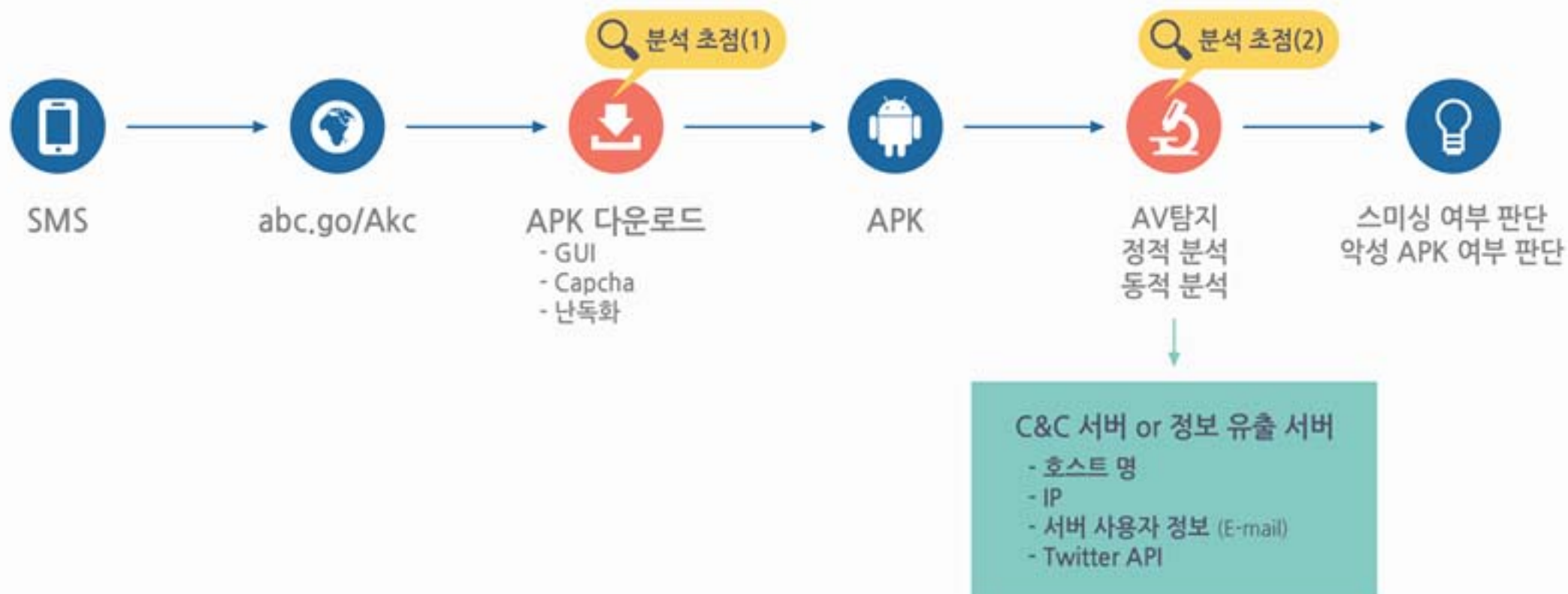
Malicious URL history of this IP

No.	URL	Anti-virus	Scan Date
61	http://m1ja.kr/ver.exe	10 / 10	2014-10-02 17:40:17
60	http://hoseu-sanjang.co.kr/admin/?_auth=001	1 / 10	2014-08-04 00:30:12
59	http://hoseu-sanjang.co.kr/admin/?_auth=001	4 / 10	2014-08-03 21:48:44
58	http://hoseu-sanjang.co.kr/admin/?_auth=	3 / 10	2014-07-28 09:24:03
57	http://befrend.pe.kr/20132.exe03Cn	1 / 10	2014-06-22 00:35:42
56	http://linkzone.com/css/index.rar	1 / 10	2014-06-18 19:47:42
55	http://rljinlaser.com/	4 / 10	2014-06-05 20:06:17
54	http://rljinlaser.com/iei/3pg.js	1 / 10	2014-06-05 05:19:45
53	http://ssfire0k.net/da/index.html	2 / 10	2014-04-08 01:35:02
52	http://eunjangdo7.net/	4 / 10	2014-04-01 21:44:28
51	http://andongseran.net/	1 / 10	2014-03-12 06:47:21
50	http://wkwg.co.kr/pop/index.html	1 / 10	2014-03-11 00:12:04
49	http://inow.co.kr/admin/a.scr	7 / 10	2014-03-04 08:42:15
48	http://widehove.com/	7 / 10	2014-02-17 13:03:49
47	http://andong114.co.kr/pop/xx.html	1 / 10	2014-02-13 05:28:15
46	http://andong114.co.kr/pop/zz.html	1 / 10	2014-02-12 08:45:04
45	http://301,79.5.66/	1 / 10	2014-02-11 05:13:13
44	http://smoc.co.kr/	6 / 10	2014-02-09 21:52:49

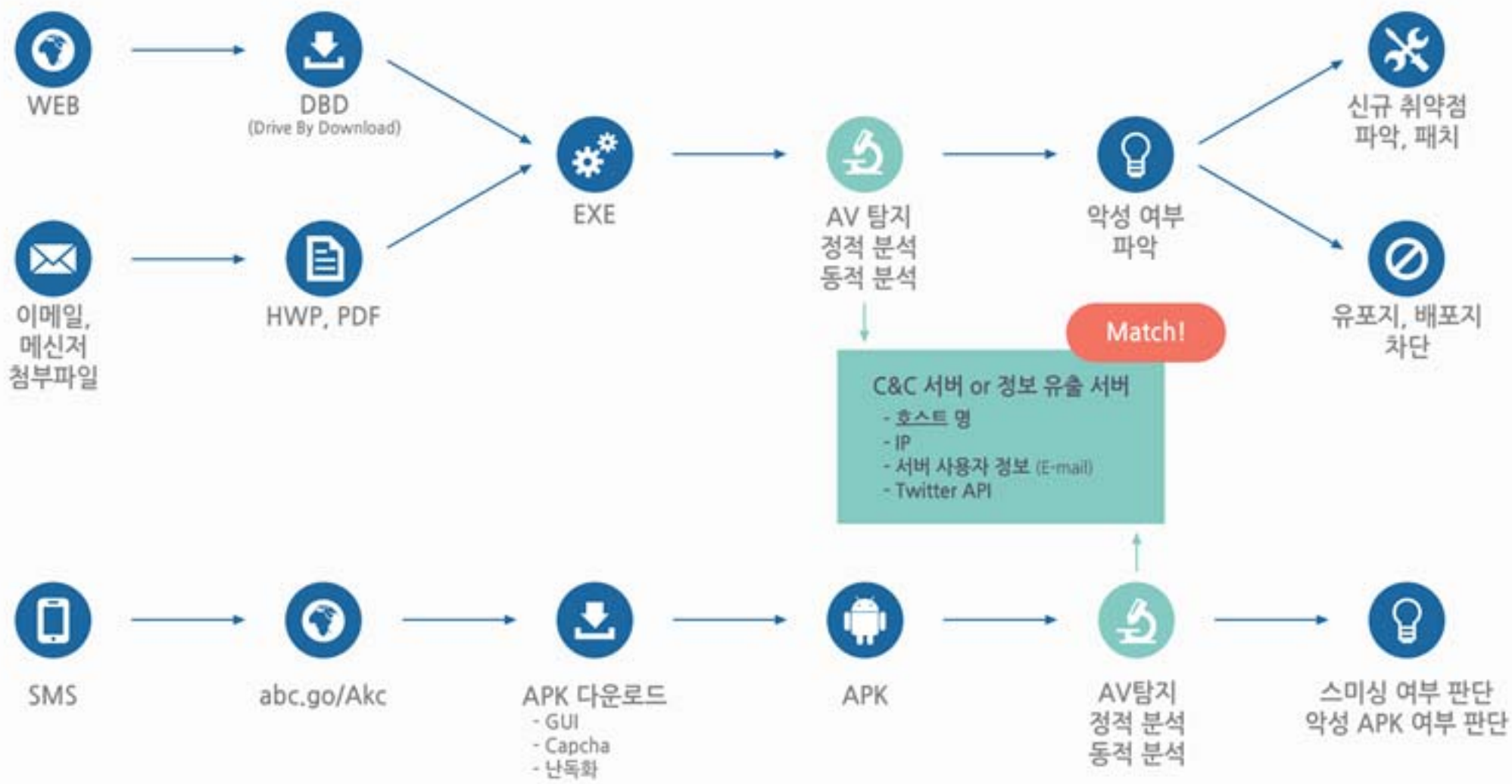
Windows OS 기반 악성코드



Android OS 기반 악성코드



연관 관계 추적





하나의 파일을 분석하면?

- ▶ 24개의 Strings 정보
- ▶ 1.4개의 IP 정보
- ▶ 38가지의 행위 패턴
- ▶ 0.8개의 배포지 정보

$$\infty \times \infty \times \infty \times \infty \times \infty = \infty$$

무한대의 데이터를 가공하면
무한의 데이터를 다시 생산할 수 있다.



빅 데이터(Big Data) & 집단 지성

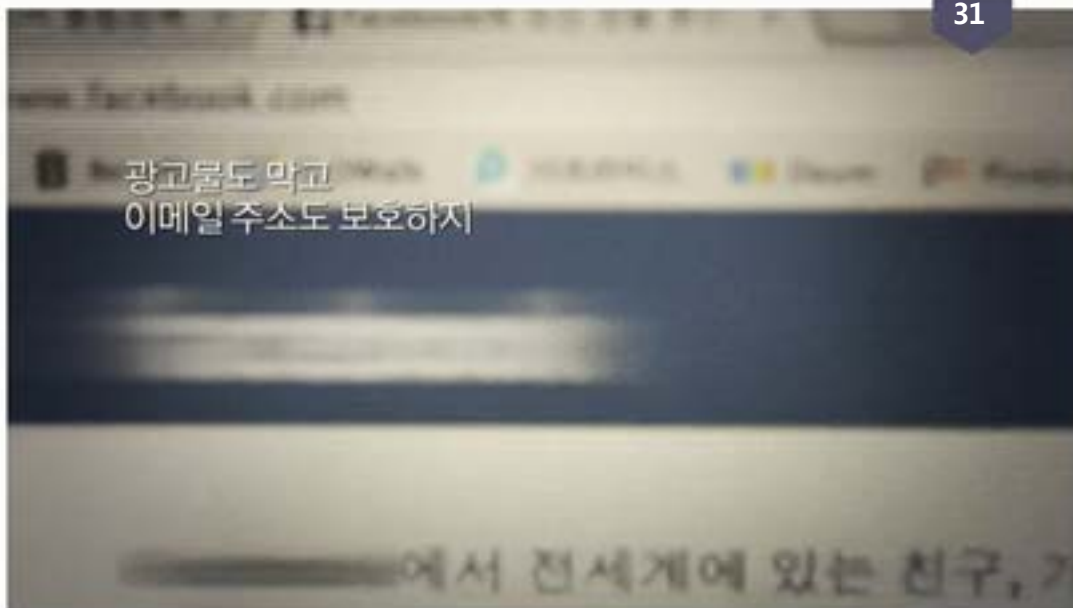
ReCaptcha

10초

너도 해봤을 거야

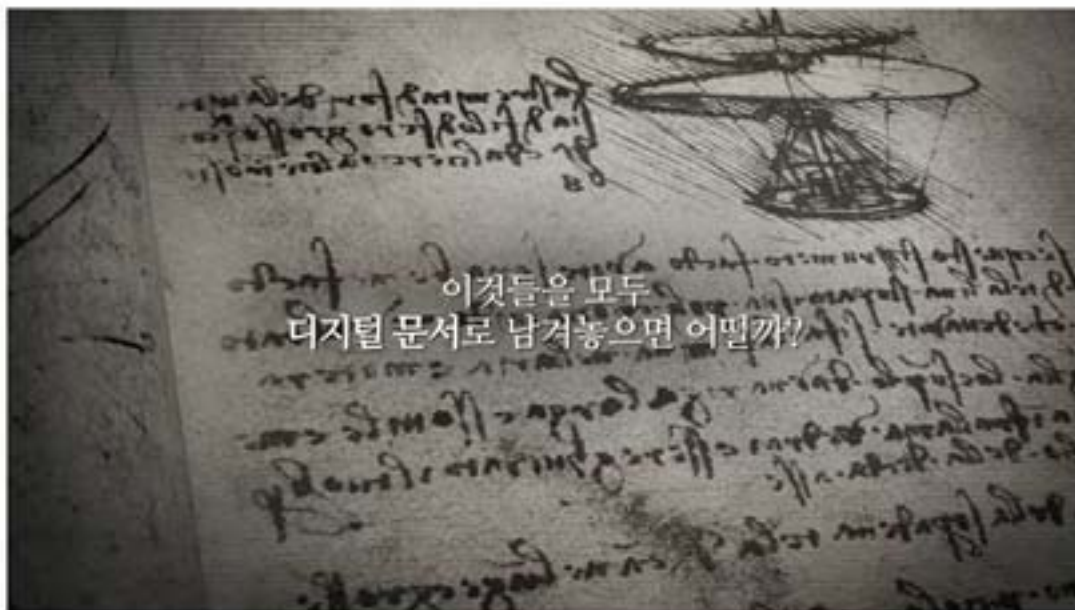
이런 문자
구불구불하고 왜곡된 문자열

캡차 CAPTCHA
컴퓨터 자동 가입 방지 프로그램





이 시간을
인류를 위해 쓸 수는 없을까?





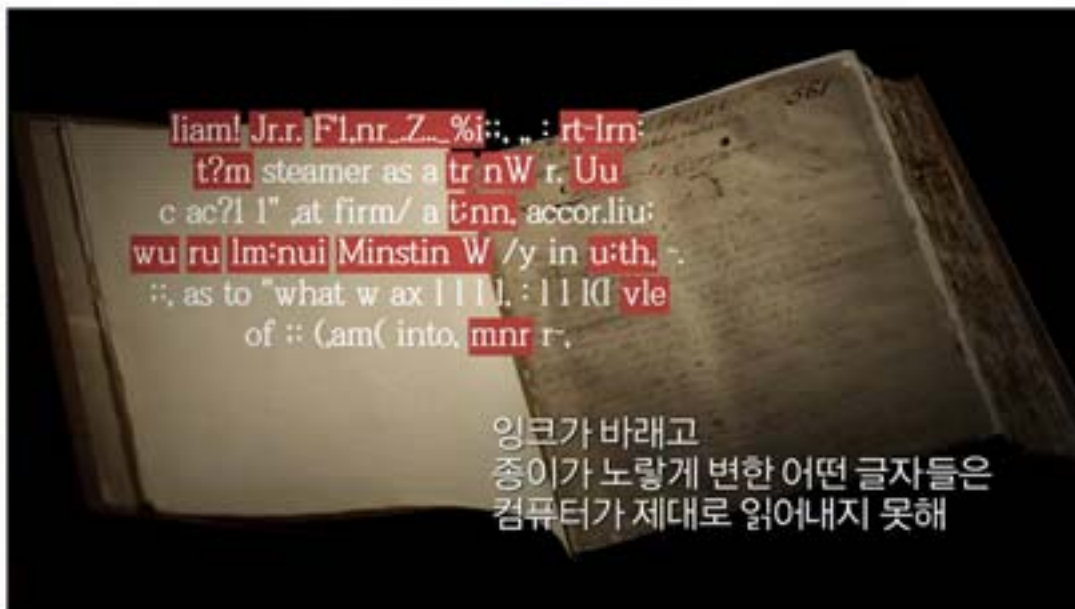
인류에게 너무도 소중한 유산들



디지털화하려면
먼저 책을 스캔해야 해



스캔한 글자 이미지를
컴퓨터가 다시 읽어내는 거지



liam! Jr.r. Fl.nr. Z. %i::, .: rt-lm:
t?m steamer as a tr nW r. Uu
c ac?l l" ,at firm/ a t'nn. accor.liu:
wu ru lm:nui Minstin W /y in u:th, -
::, as to "what w ax l l l l. : l l l d vle
of :: (am(into, mn r.

잉크가 바래고
종이가 노랗게 변한 어떤 글자들은
컴퓨터가 제대로 읽어내지 못해

일일이 사람이 구별해야 하지

그 인력, 돈, 시간...
불가능하지

그래서  생각해 낸 거야

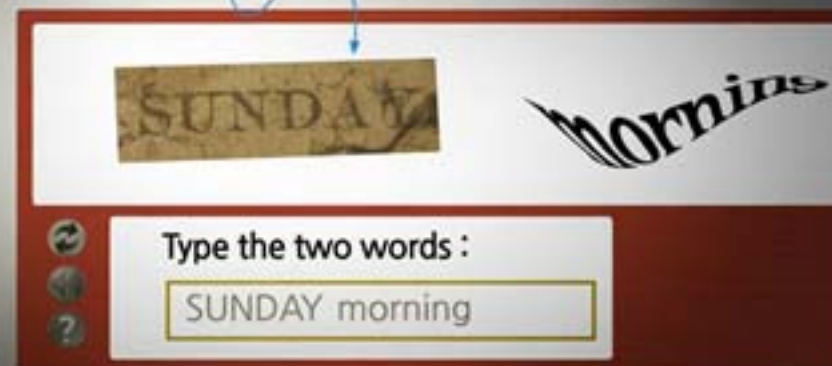
전 세계 2억 명이 함께 하면
가능하지 않을까?



입력하는 글자 둘 중 하나는
기존의 자동 가입 방지 기능

read books

또 다른 하나는
컴퓨터가 읽지 못한 글자



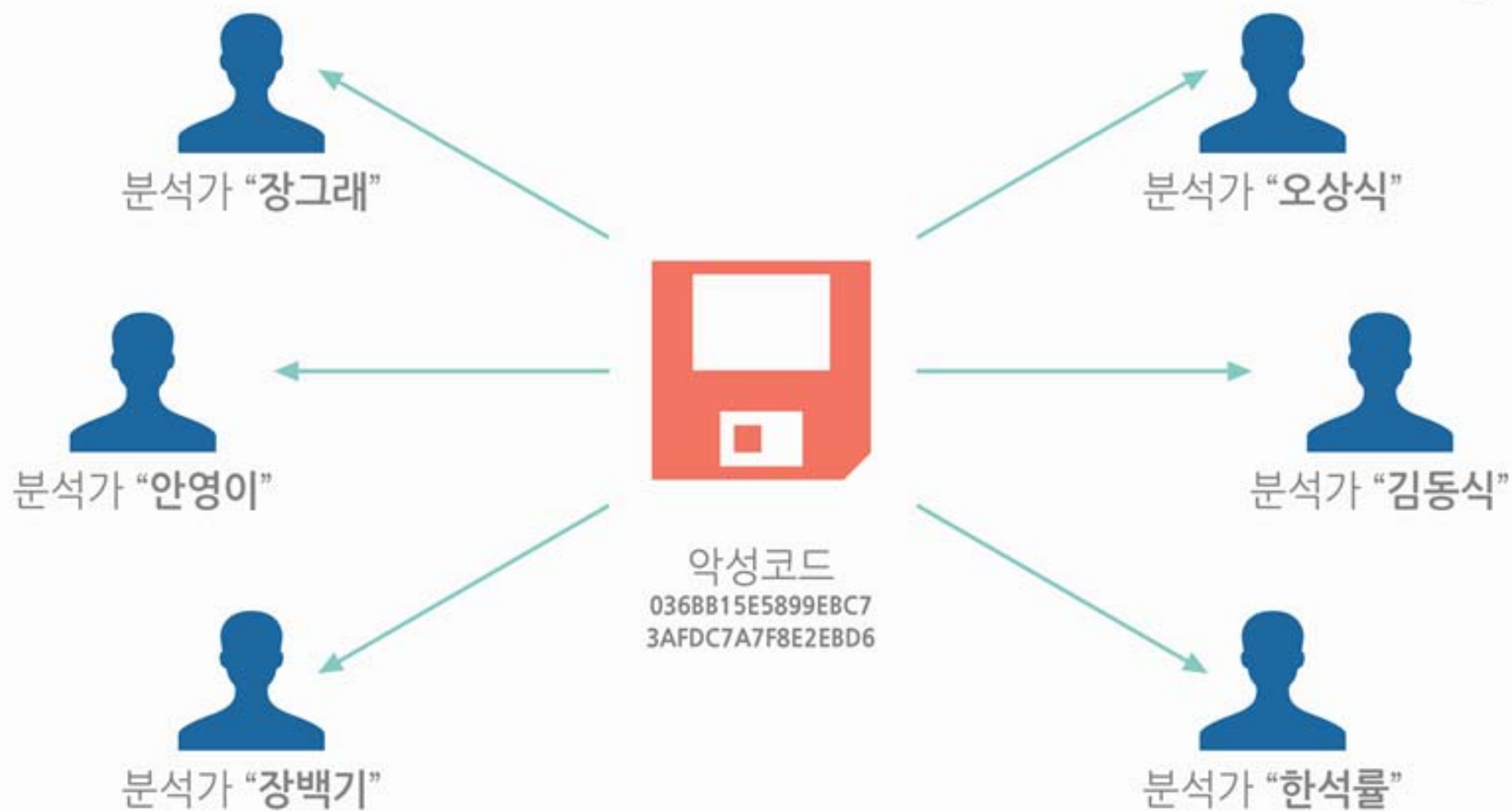




일년에 책 250만 권이야



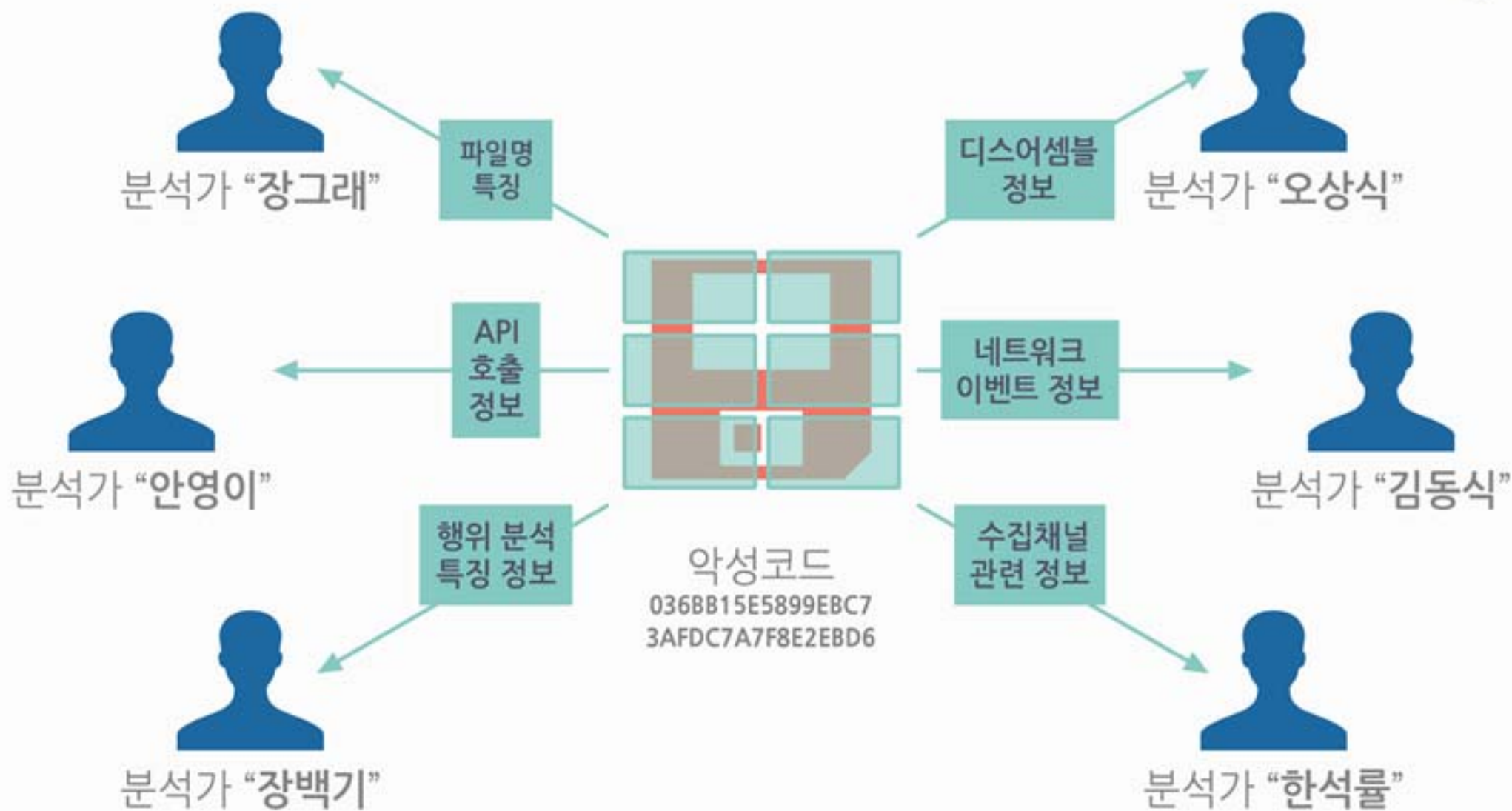
인류에게 중요한 자료를 모아서
전자정보로 저장하고
배포하는 프로젝트지

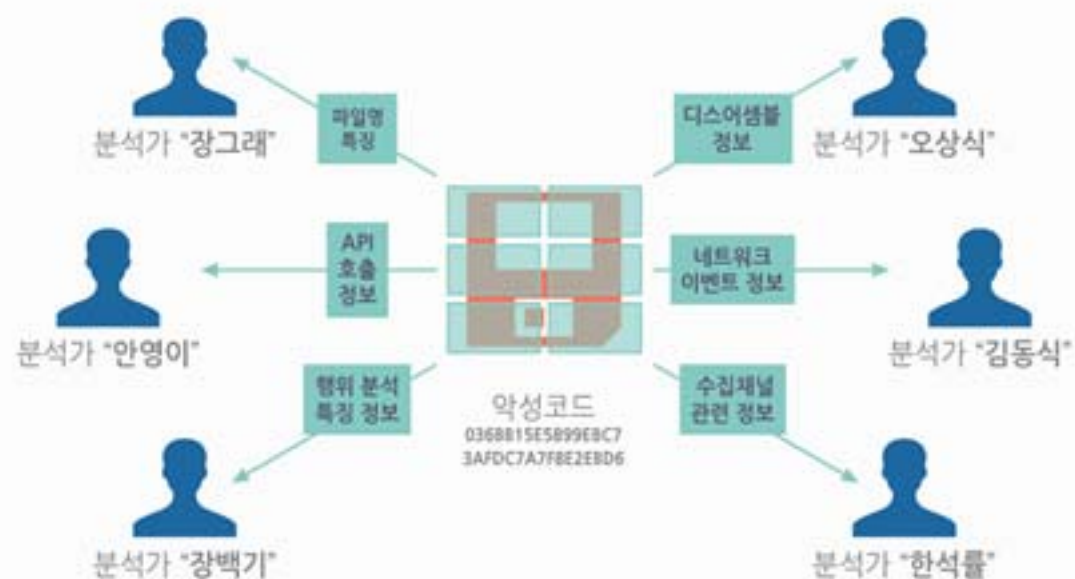


다 비슷비슷한
분석 보고서.
인재들이 비슷한 일을 하고 있다.

이 능력 있는 사람들이
한곳에 모여 **정보를 공유**한다면?

그리고 공유된 정보로 더 뛰어난 분석내용이 나온다면!?





+

비전문가들의 간단한 투표



다수의 Anti-Virus 탐지 정보

win32/Trojan/abc
abc.Trojan.win32

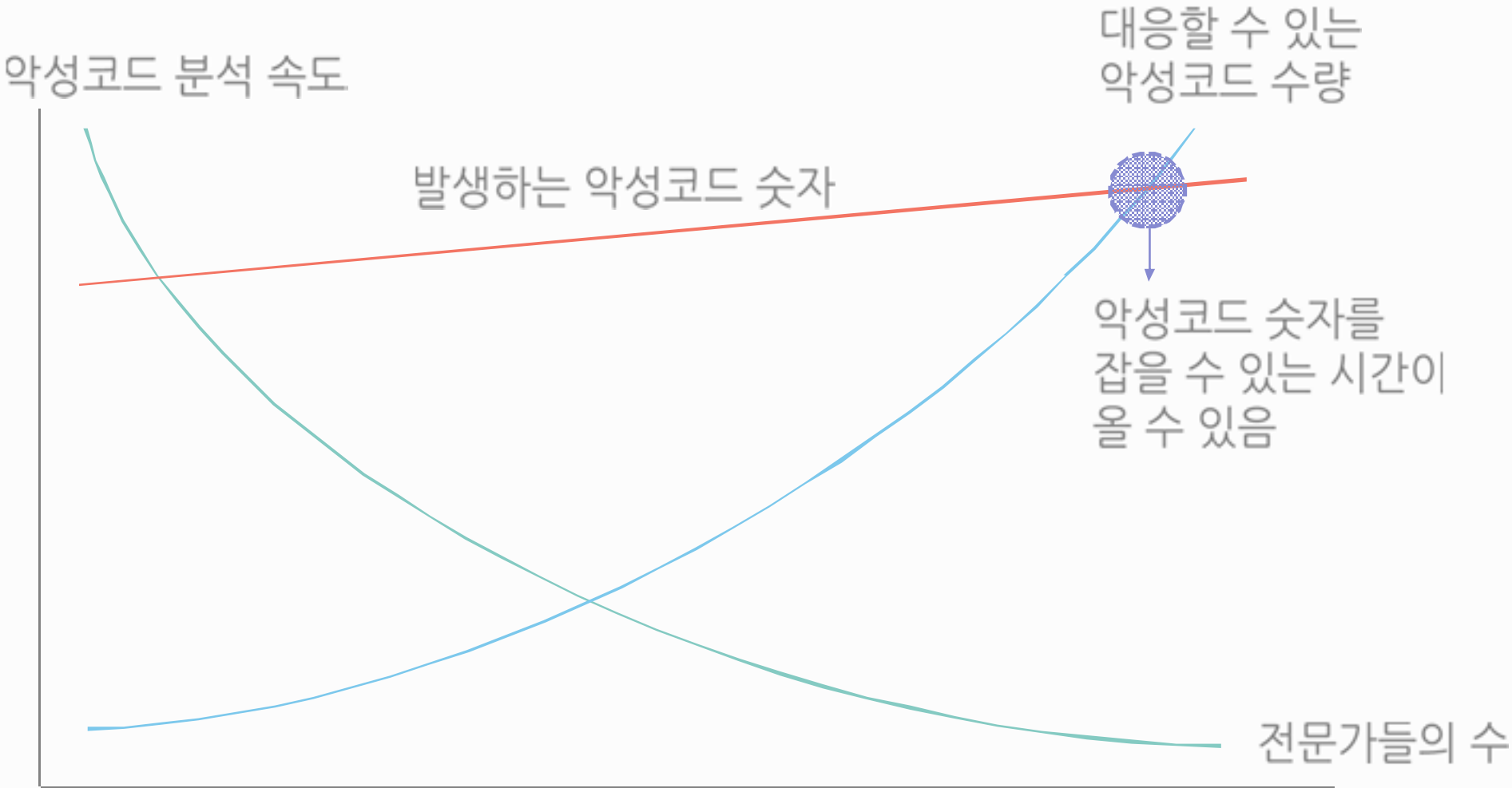
⋮

전세계 사용자들이 모여서
하나씩 관련된 **정보**를 모으고 **프로파일링**하게 된다면...

중복된 시간 소비를 줄여서
효율을 **극대화**시킬 수 있음.

그러기 위해서는?

그것을 모을 수 있는 **시스템**과 **체계**가 필요함!



Just for **FUN** :)

Q. 해외 A/V 평균 대응 시간?

Q. 국내 A/V 평균 대응 시간?



Q. 오탐이 가장 많은 A/V?

Q. 정탐(악성을 악성으로, 정상을 정상으로 탐지하는)을
가장 잘하는 A/V?



감사합니다
Thank you