


캡스톤 디자인 I

종합설계 프로젝트

프로젝트 명	MASK(Malware Analysis System in Kookmin)
팀 명	NCNP(No Commit No Pay)
문서 제목	계획서

Version	1.7
Date	2018-MAR-09

팀원	한 채연(조장)
	김 영재
	명 준우
	이 유정
	허 준녕

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09


CONFIDENTIALITY/SECURITY WARNING

이 문서에 포함되어 있는 정보는 국민대학교 전자정보통신대학 컴퓨터공학부 및 컴퓨터공학부 개설 교과목 캡스톤 디자인I 수강 학생 중 프로젝트 **"MASK(Malware Analysis System in Kookmin)"**를 수행하는 팀 **NCNP**의 팀원들의 자산입니다. 국민대학교 컴퓨터공학부 및 팀 **NCNP**의 팀원들의 서면 허락없이 사용되거나, 재가공 될 수 없습니다.

문서 정보 / 수정 내역

Filename	수행계획서-NCNP.docx
원안작성자	한채연
수정작업자	한채연, 허준녕, 김영재, 명준우, 이유정

수정날짜	대표수정자	Revision	추가/수정 항목	내 용
2018-03-04	한채연	1.0	최초 작성	개요 초안 작성, 개발 목표 초안 작성
2018-03-05	허준녕	1.1	내용 수정	개발 목표 수정
2018-03-05	한채연	1.2	내용 추가	개발 일정 초안 작성, 개발 목표 및 내용 초안 작성, 역할 분담 초안 작성, 시스템 기능 요구사항 초안 작성
2018-03-05	한채연	1.2.1	내용 추가	배경 기술 초안 작성 시스템 기능 요구사항 추가
2018-03-06	이유정	1.3	내용 추가	세부 개발 내용 추가
2018-03-06	명준우	1.4	내용 추가	개발 추진 배경 내용 추가 세부 개발 내용 추가
2018-03-06	한채연	1.5	내용 추가	개요 추가, 그림 수정
2018-03-06	김영재	1.6	내용 추가	시스템 구조도 그림 및 내용 추가
2018-03-08	한채연	1.7	최종	최종 수정 및 작성

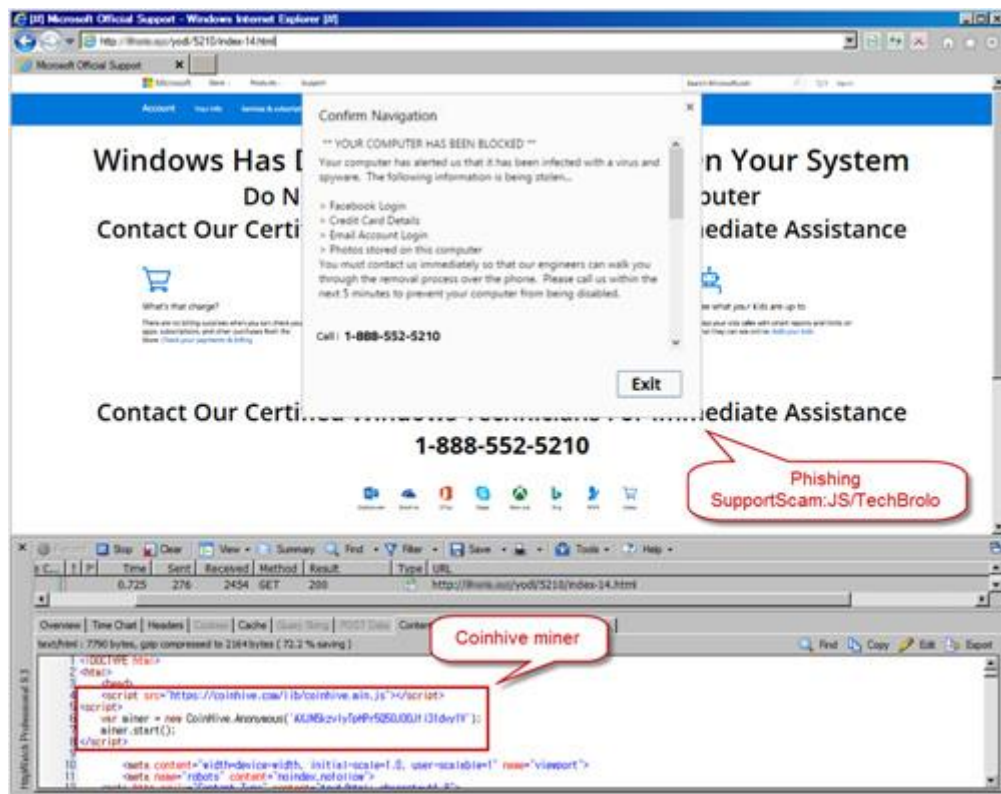
 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

목 차

1	개요	4
1.1	프로젝트 개요	4
1.2	추진 배경 및 필요성	7
1.2.1	바이러스 분석 기술의 시장 현황	7
1.2.2	현재 바이러스 분석 시스템의 한계점과 개선 방향	10
2	개발 목표 및 내용	11
2.1	목표	11
2.2	연구/개발 내용	12
2.2.1	Scenario1 – 데이터베이스에 분석 결과가 존재하는 경우	12
2.2.2	Scenario2 – 데이터베이스에 분석 결과가 존재하지 않는 경우	13
2.2.3	세부 연구/개발 내용	13
2.3	개발 결과	18
2.3.1	시스템 기능 요구사항	18
2.3.2	시스템 비기능(품질) 요구사항	19
2.3.3	시스템 구조	20
2.3.4	결과물 목록 및 상세 사양	20
2.4	기대효과 및 활용방안	21
3	배경 기술	22
3.1	기술적 요구사항	22
3.2	현실적 제한 요소 및 그 해결 방안	23
3.2.1	하드웨어	23
3.2.2	소프트웨어	23
4	프로젝트 팀 구성 및 역할 분담	24
5	프로젝트 비용	25
6	개발 일정 및 자원 관리	25
6.1	개발 일정	25
6.2	일정별 주요 산출물	26
6.3	인력자원 투입계획	27
6.4	비 인적자원 투입계획	28
7	참고 문헌	29

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

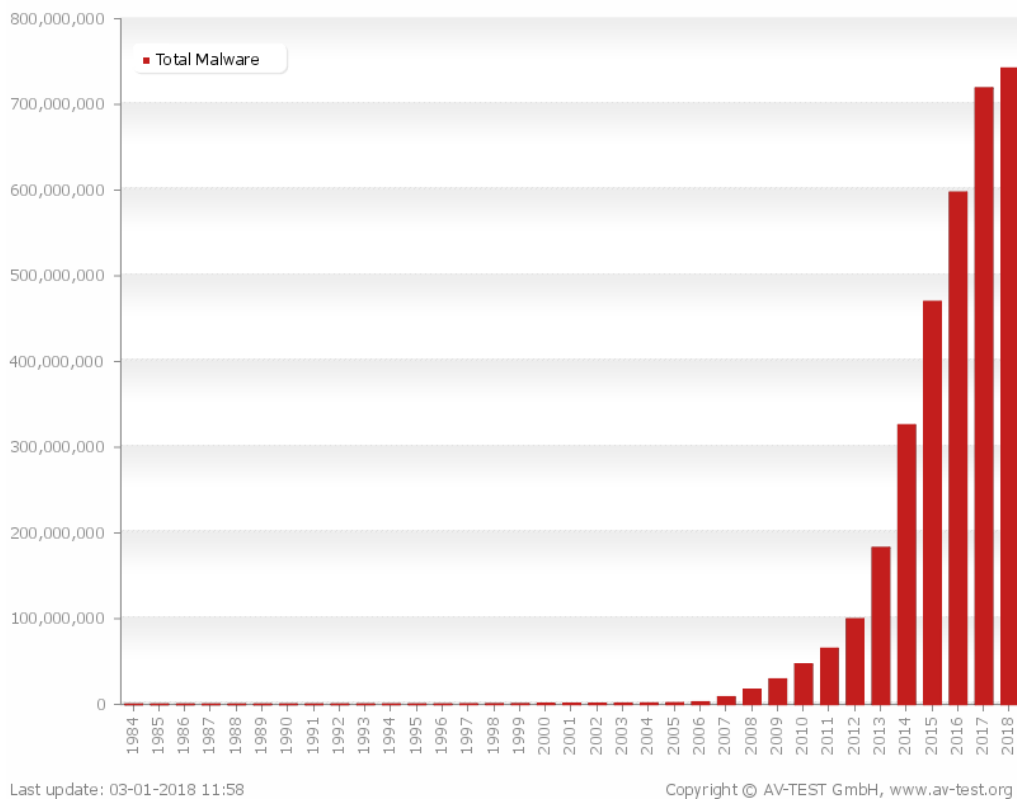
악성코드 “마이너”가 기승을 부리고 있다. 이렇듯 능동적이고 지능화된 악성코드의 공격이 지속적으로 범람하고 있다.



[그림 3] 피싱 사이트에서도 이용되는 비트코인 마이너[출처=제로서트]

안티바이러스 테스트 업체 AV-TEST에 따르면 해마다 발견되는 악성코드는 늘어나는 추세이며, 현재 하루에 대략 100만개 정도의 악성코드가 발견되고 있다. 이에 비해 악성코드 전문가는 현격히 부족하다. 게다가 안티 바이러스 제품이나 기존 탐지 솔루션은 복잡해지고 정교해지는 악성코드에 대해 효과적으로 대응하지 못한다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09



[그림 4] 연도 별 발견되는 총 악성코드 개수[출처=AV-TEST]

본 프로젝트는 악성코드로 의심되는 파일을 분석하여 그 결과들을 악성코드 분석가에게 제공함으로써 보다 정교하게 악성코드를 분석할 수 있도록 한다. 본 프로젝트에서 제공하는 기능은 아래와 같다.

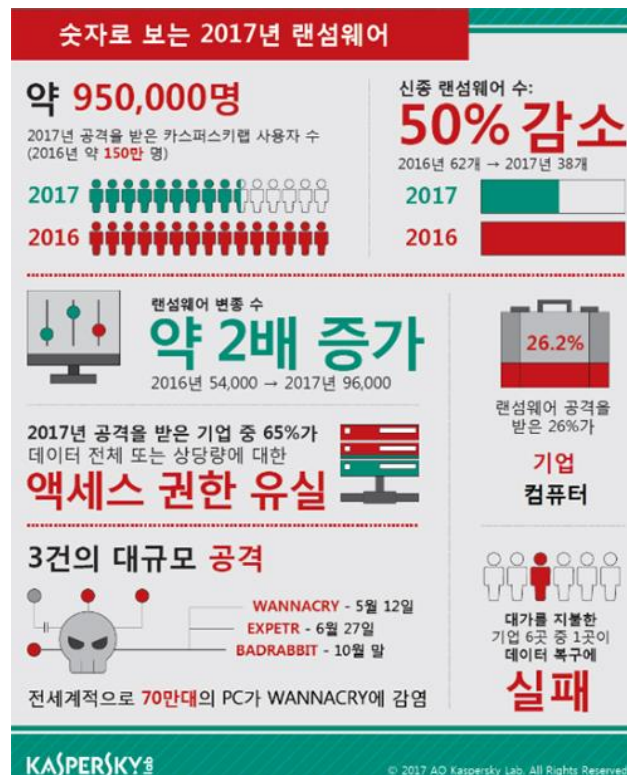
첫째, 파일에 대한 다양한 분석이다. 사용자가 악성으로 의심되는 파일을 웹에 업로드하면, 그 파일에 대한 정적 분석 결과와 동적 분석 결과를 모두 제공해 준다.

둘째, 딥러닝 모델에 의한 결과이다. 사용자가 악성으로 의심되는 파일을 웹에 업로드하면, 그 파일에 대한 다양한 결과값들을 제공한다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

1.2 추진 배경 및 필요성

악성코드의 개수가 나날이 증가함에 따라 경제적, 사회적 문제를 발생시키고 있다. 현저히 늘어나는 악성코드 개수에 비해 악성코드 분석 전문가의 수는 부족하다. 또한 매일 수십만 개의 신종/변종 악성코드들이 생성되고 있다. 수많은 악성코드를 비롯한 각종 사이버 침해 사고로 인한 경제적 피해규모가 상당하며, 국가 사회적인 혼란을 유발하여 국민생명과 국가안보에 심각한 위협이 된다.



[그림 5] 숫자로 보는 2017년 랜섬웨어[출처=카스퍼스키랩]

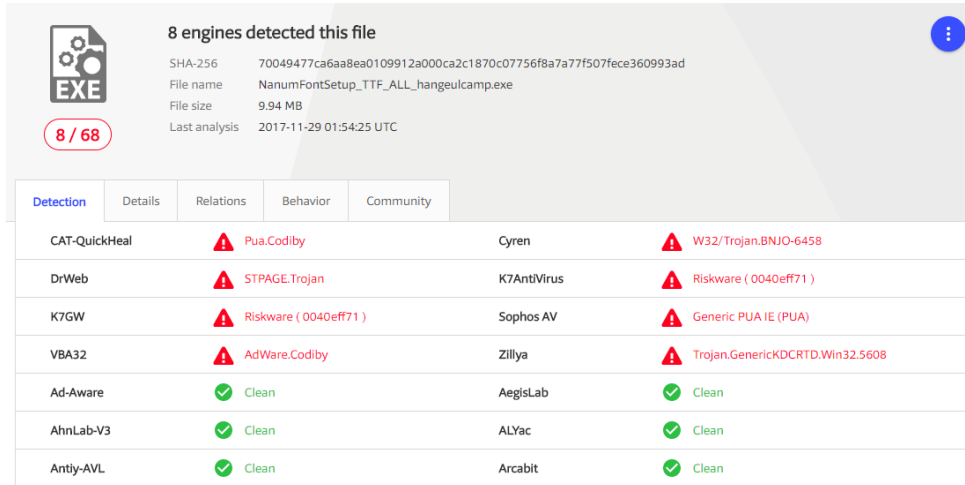
기존의 악성코드 분석 기술과 전문 분석가에 의한 대응으로는 신종/변종 악성코드의 생성 속도를 따라갈 수가 없다. 따라서 대량의 악성코드를 효율적으로 식별하기 위한 자동화된 기술이 반드시 필요하다.

1.2.1 바이러스 분석 기술의 시장 현황

1) VirusTotal

‘VirusTotal’은 구글(Google)의 자회사로, 파일이나 URL을 안티바이러스 엔진과 웹사이트 스캐너를 이용하여 악성여부를 식별해주는 무료 온라인 서비스이다. V3, 알약, 비트디펜더 등 약 60여가지의 바이러스 검사 소프트웨어 제품을 사용한다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09



Detection	Details	Relations	Behavior	Community
CAT-QuickHeal	⚠ Pua.Codliby			⚠ W32/Trojan.BNJO-6458
DrWeb	⚠ STPAGE.Trojan			⚠ Riskware (0040eff71)
K7GW	⚠ Riskware (0040eff71)			⚠ Generic PUA.IE (PUA)
VBA32	⚠ AdWare.Codliby			⚠ Trojan.GenericKDCRTD.Win32.5608
Ad-Aware	✅ Clean			✅ Clean
AhnLab-V3	✅ Clean			✅ Clean
Antiy-AVL	✅ Clean			✅ Clean

[그림 6] Virustotal 분석 결과 화면 중 일부

2) malwares.com

‘malwares.com’은 세인트시큐리티(Saint Security)에서 개발한 한국 최초의 빅데이터 기반 악성코드 자동 분석 플랫폼으로, 다양한 수집 채널로부터 유입된 악성코드를 자동으로 분석하고, 분석된 결과를 공유하여 악성코드 유포 정황을 빠르게 인지하고 크게 확산되는 것을 방지하여 각종 악성코드로부터의 공격에 능동적으로 대처하는 것을 목적으로 만들어진 인텔리전스 서비스다.



70049477CA6AA8EA0109912A000CA2C1870C07756F8A7A77F507FECE360993AD

MD5 : AD7587A2CAE0ED77E8774447F2F28726
SHA-1 : 84A9BC095D31A2995CC62FECA7CDB933D2EDA971
SHA-256 : 70049477CA6AA8EA0109912A000CA2C1870C07756F8A7A77F507FECE360993AD
파일 크기 : 10,425,776 bytes
파일 유형 : exe_32bit
알려진 날짜 : 2014-03-29 14:59:05 (3년 11개월 전)

AI safe 66/100 malware

태그 #peexe #overlay #user-directory #signed #nsis #packing #exe_32bit

[그림 7] malware.com 분석 결과 화면 중 일부

3) Hybrid Analysis

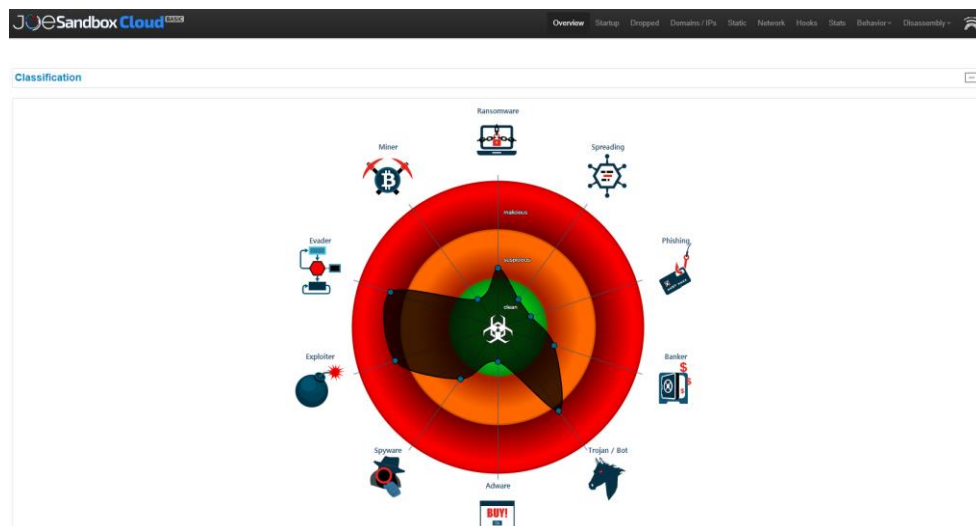
‘Hybrid Analysis’는 독일의 페이로드시큐리티(Payload Security)에서 개발한 악성 소프트웨어 분석 서비스로, 하이브리드 분석 기술을 사용하여 알려지지 않은 위협을 탐지 및 분석한다. Hybrid Analysis는 분석 정보 생성 단계에서 심층적인 정적 분석을 수행할 수 있도록 심볼 정보와 모니터

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

링 된 런타임 프로세스의 메모리 덤프 스냅샷을 미세한 단위로 저장하고, 정적 분석과 동적 데이터를 결합하여 실행에 관계없이 악성 행위를 탐지한다.

4) Joe Sandbox

‘Joe Sandbox’는 스위스의 ‘Joe Security’에서 개발한 악성코드 정밀 분석 샌드박스로, 샌드박스 내에서 얻은 분석정보를 시각화하여 분석 리포트에서 보여준다. 윈도우 운영체제 기반 데스크톱 PC부터 맥OS, 안드로이드와 iOS 등 다양한 모바일 OS까지 다양한 지원 제품군으로 구성되어 있는 것이 가장 큰 특징이다.




[그림 8] Joe Sandbox 분석 결과 화면 중 일부

5) Clam AntiVirus

‘Clam AntiVirus’는 CISCO Systems에서 지원하는 오픈소스 소프트웨어로 자유 크로스플랫폼 형식의 바이러스 검사 소프트웨어 툴킷이며, 바이러스를 비롯한 수많은 종류의 악성 소프트웨어를 탐지해낸다. 주로 메일 게이트웨이에서의 이메일 스캐닝을 위하여 설계되었다.

6) Kicom AntiVirus


‘Kicom AntiVirus’는 오픈소스 소프트웨어로 악성코드를 탐지하고 치료하기 위하여 설계된 안티바이러스 엔진이다. 1995년 C/C++로 작성되었다가 1998년 HAURI의 ViRobot 엔진과 통합된 후에 파이썬 언어로 다시 개발되었다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

1.2.2 현재 바이러스 분석 시스템의 한계점과 개선 방향

위에 언급한 대부분의 제품들은 Public API를 제공하지만 한정적인 서비스만을 이용할 수 있으며, 더 나은 서비스를 제공받기 위해서는 유료로 이용해야 한다. 또한 오픈소스 소프트웨어가 아니므로 추가적인 개발을 할 수 없다. 따라서 본 프로젝트는 오픈소스 소프트웨어를 지원하여 사용자가 원하는 기능을 추가할 수 있도록 한다.

ClamAV와 KicomAV는 신종/변종 악성코드에 대한 자동 업데이트가 이루어지지 않으므로 신종/변종 악성코드 탐지에 취약하다. 따라서 본 프로젝트는 정적 분석 뿐 만 아니라 동적 분석도 진행하고 더 나아가 딥러닝을 적용하여 라벨링까지 하므로 신종/변종 악성코드에 실시간으로 대응하기 유리하다. 그리고 악성코드 분석가에게 자세하고 보다 더 다양한 분석 결과를 제공하는 방향으로 나아간다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

2 개발 목표 및 내용

2.1 목표

본 프로젝트는 악성코드 분석 전문가에게 악성코드로 의심되는 파일의 정적, 동적 분석에 대한 결과와, 그와 유사한 파일에 대한 분석 결과, 자체적으로 학습한 모델에 의해 결정된 악성코드의 라벨을 제공하는 것을 목표로 한다. 또한 오픈소스 소프트웨어로 개발하는 것을 목표로 한다.

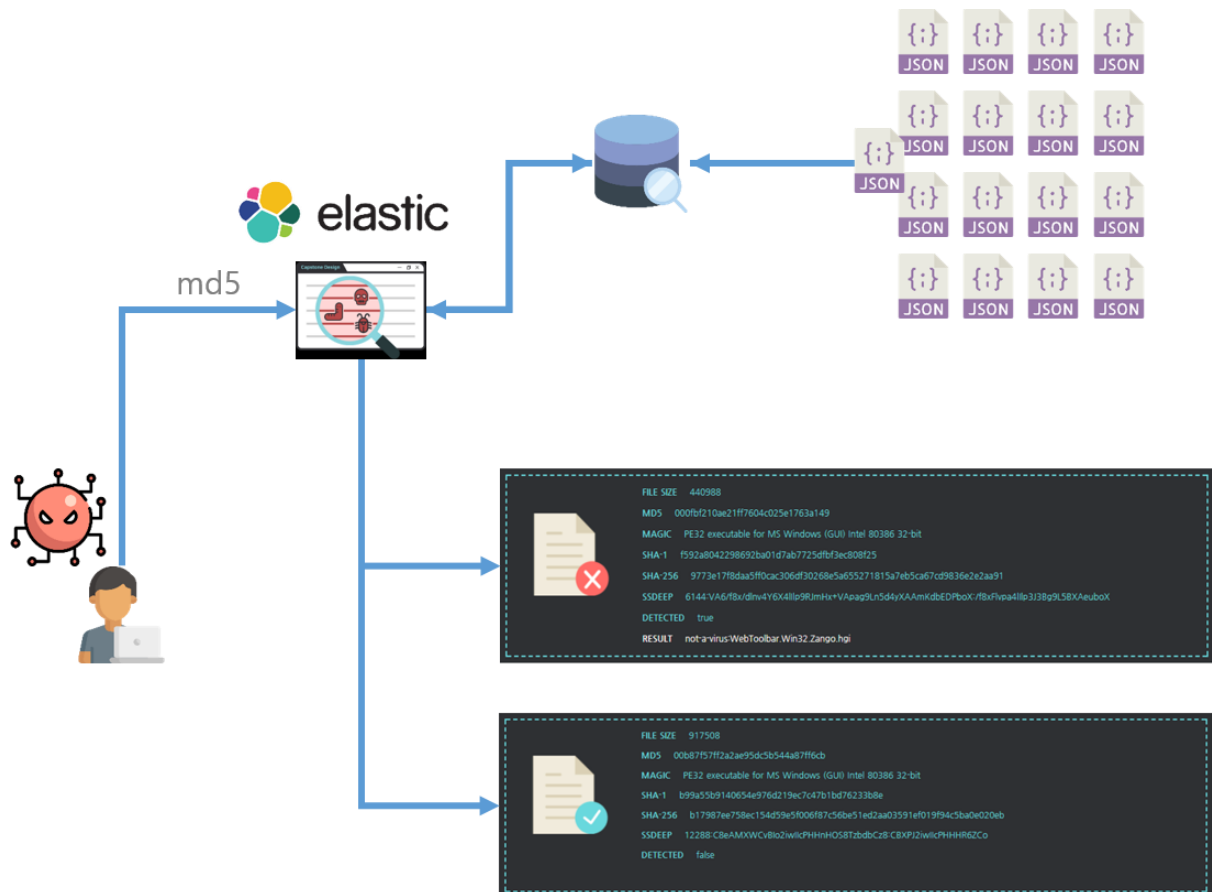
■ 세부 목표

- 업로드 한 파일과 유사한 파일을 데이터베이스에서 찾아 그 파일의 정보를 제공한다.
- 분석된 악성코드의 세부정보를 사용자가 한 눈에 볼 수 있도록 웹에 시각화 한다.
- 업로드한 파일에 대한 정적 분석, 동적 분석 결과 리포트를 자동으로 생성한다.
- 자동으로 생성된 리포트로부터 피처를 자동으로 추출한다.
- 주기적으로 재학습을 시켜서 새로운 악성코드에 대해 대응을 한다.

 <div> <p>국민대학교</p> <p>컴퓨터공학부</p> <p>캡스톤 디자인 I</p> </div>	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

2.2연구/개발 내용

2.2.1 Scenario1 – 데이터베이스에 분석 결과가 존재하는 경우

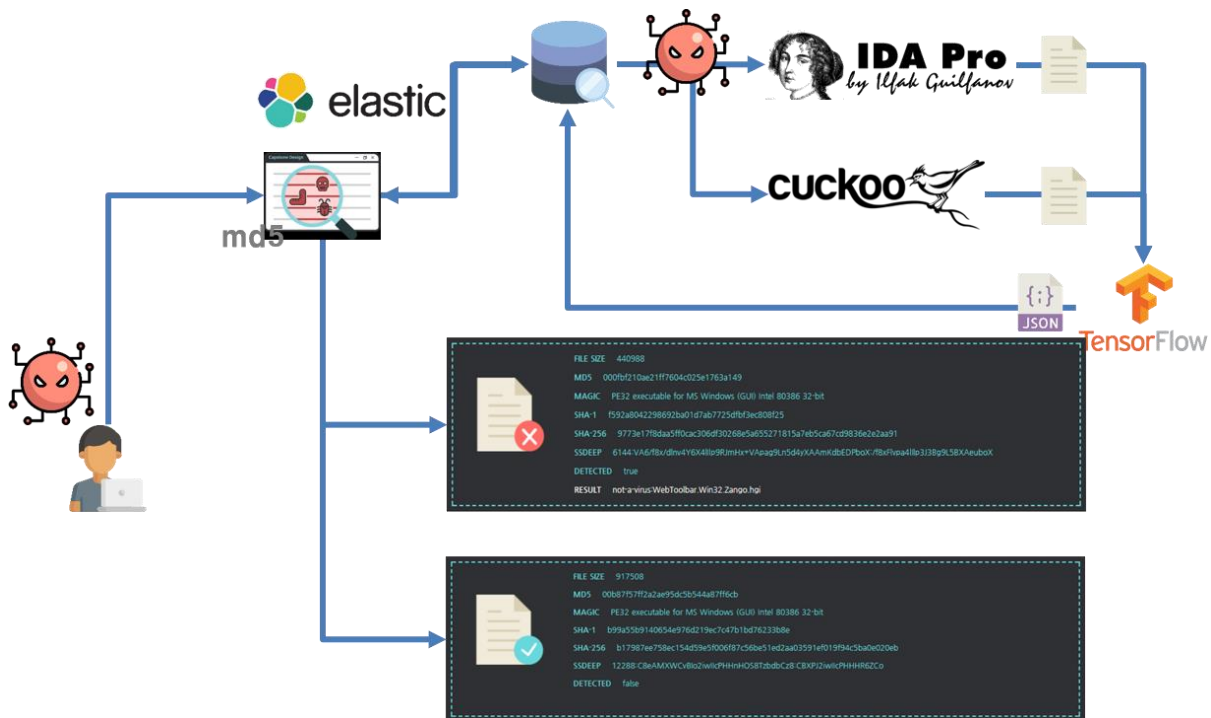


[그림 9] 데이터베이스에 분석 결과가 존재하는 경우

본 프로젝트에서는 검색엔진을 활용한다. 사용자가 악성으로 의심되는 파일을 웹에 업로드 하면, 그 파일의 md5 값을 구하여 데이터베이스와 연동된 엘라스틱서치를 통해 악성코드 분석 결과를 찾는다. 리포트가 데이터베이스에 존재할 경우 리포트 파일과 그 연관된 정보들을 웹을 통해서 유저에게 보여준다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

2.2.2 Scenario2 – 데이터베이스에 분석 결과가 존재하지 않는 경우



[그림 10] 데이터베이스에 분석 결과가 존재하지 않는 경우

- 리포트가 데이터베이스에 없을 경우, 파일에 대해 각각 정적 분석(Static Analysis)과 동적 분석(Dynamic Analysis)을 진행하고 리포트를 생성한다. 분석 후 생성되는 각각의 리포트로부터 피쳐들을 추출한다. 추출한 피쳐를 이용하여 학습된 모델로부터 탐지 결과를 구한다. 이 결과들을 데이터베이스에 저장한 후 웹을 통해 유저에게 분석 결과를 보여준다.

2.2.3 세부 연구/개발 내용

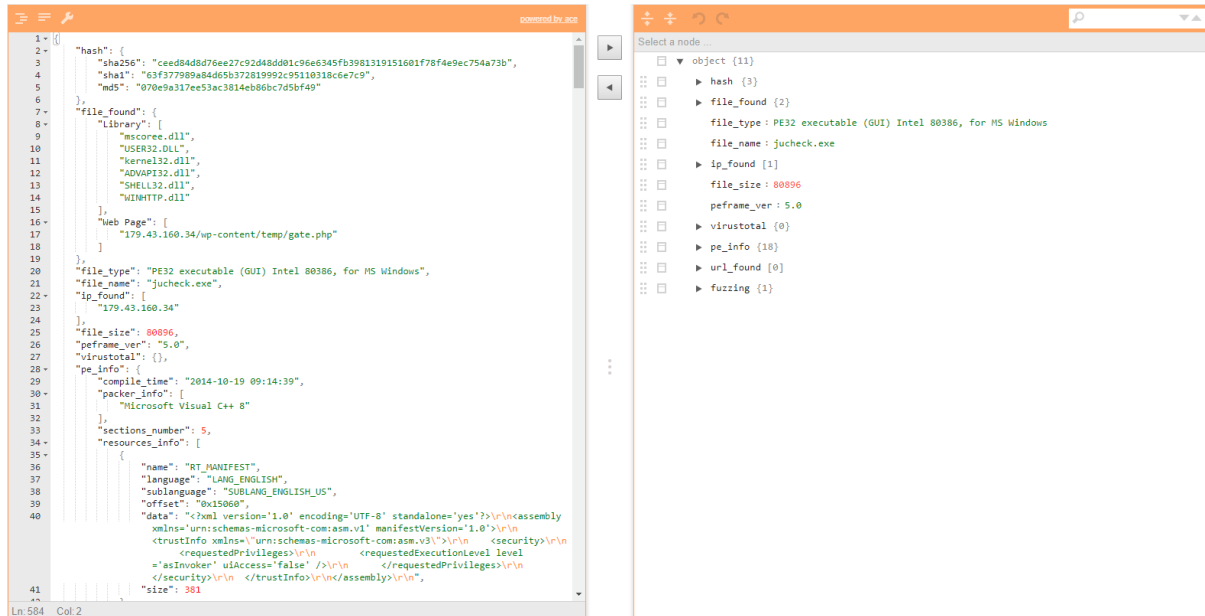
본 프로젝트는 각 연구/개발 내용을 다음과 같이 분야 별로 나누어 세부 목표를 정하고 연구 및 개발을 진행한다.

1) 정적 분석

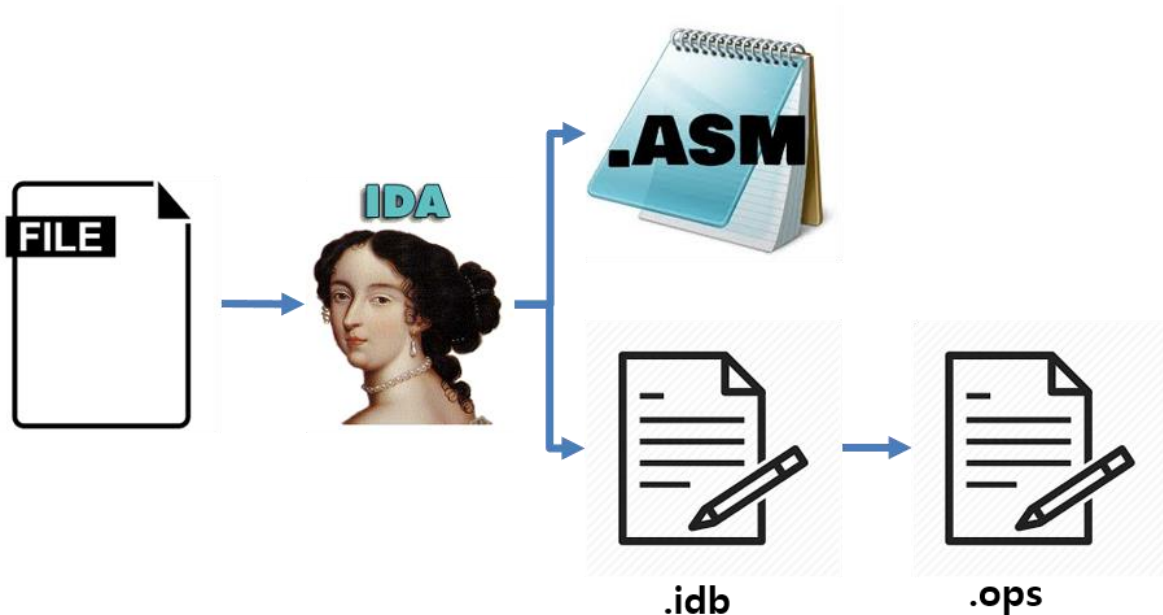
악성코드 정적 분석을 위해서는 IDA Pro 와 파이썬 오픈 소스인 peframe을 이용할 것이다. peframe을 이용하여 파일의 정적 분석 정보를 뽑아서 json 형태로 저장을 하고, IDA Pro를 이용하여 파일의 어셈블리 코드와 idb파일을 생성해 데이터베이스에 저장을 한다. 저장된 데이터는 웹

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

서비스와 딥 러닝의 피쳐로 사용된다.



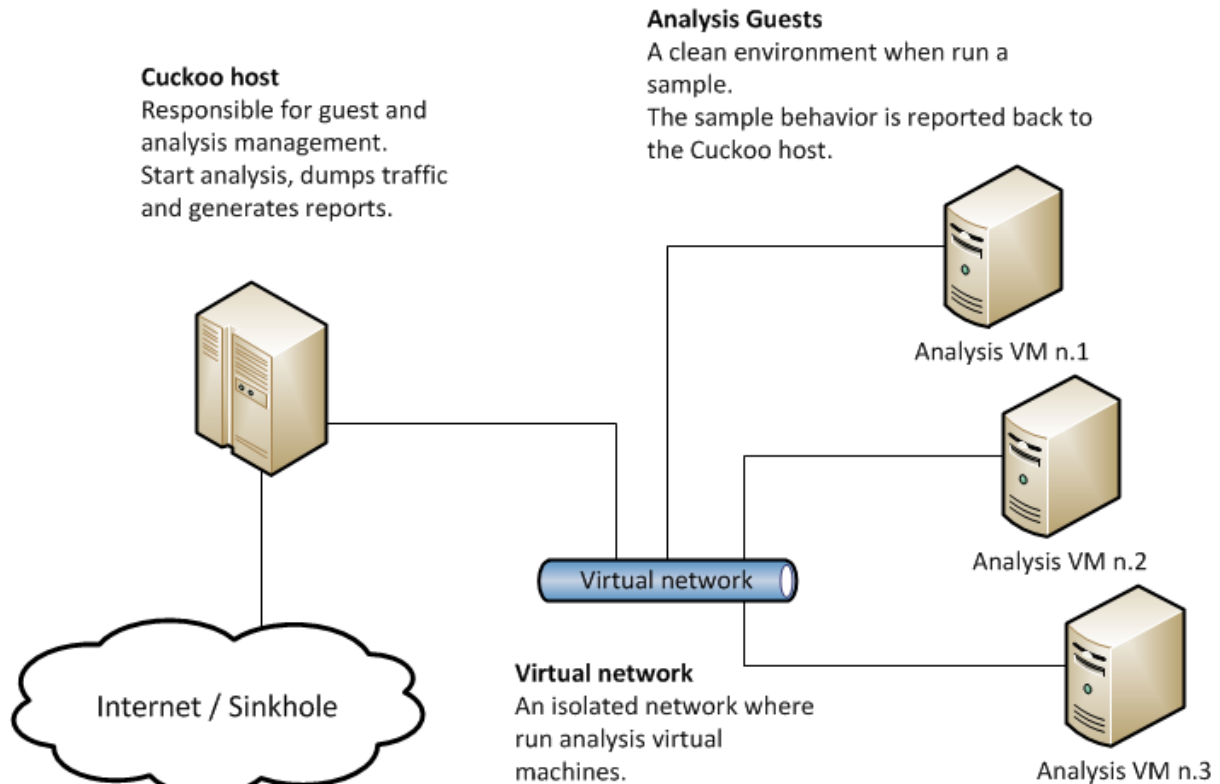
[그림 11] peframe 을 이용해 만든 정적 정보 json파일



[그림 12] ida pro를 이용한 정적 분석 계획

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

2) 동적 분석



[그림 13] 쿠쿠샌드박스의 메인 아키텍처 [출처=쿠쿠샌드박스 문서]

악성코드 동적 분석을 위한 오픈소스 소프트웨어인 쿠쿠샌드박스를 이용한다. 먼저 쿠쿠샌드박스에 대해 분석 및 조사를 한다. 쿠쿠샌드박스의 Host와 Guest instance들의 설정을 프로젝트 환경에 맞게 변경한다. 대략적인 기본 세팅을 마친 후 시스템 최적화 작업을 진행한다. 효율적인 분석과 리포트 생성을 위하여 최대한 안정성을 높이고 처리량을 늘리는 방향으로 인스턴스들을 구성하는 방법에 대해 연구한다. 이와 더불어 효율적인 시스템 이용을 위하여 쿠쿠샌드박스에서 제공하는 API를 분석한다. 또한 추출된 리포트를 분석하여 어느 피처가 사이버 킬체인 관점에서 유용한 정보인지 분석 후 피처를 추출하는 작업을 한다. 마지막으로 이 모든 것을 자동화하는 작업을 진행한다.

3) 라벨링

본 프로젝트에서는 악성코드에 대한 라벨링은 정적 분석과 동적 분석에 대한 정보를 이용하여 특징에 따라 분류를 진행한다. 라벨은 AV-TEST에서 각 부분 최고점을 받은 안티 바이러스인 카스퍼스키 라벨을 기반으로 독자적인 라벨을 제작할 예정이다. 먼저 안티 바이러스의 명명법을 매기는 방법을 연구하고 이를 기반으로 독자적인 라벨을 제작 한다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

4) 딥러닝

본 프로젝트는 정적, 동적 분석 결과로부터 피처를 생성하여 딥러닝 모델을 설계할 예정이다. 먼저 정적 분석 결과를 이용한 피처 생성 방법은 다음과 같다. IDA Pro를 이용하여 입력 파일을 idb 파일을 생성 한다. 생성된 idb파일을 ida python을 이용하여 피처를 뽑은 다음, 문자열 추출 기법을 이용하여 추출한 문자열을 이용하여 피처를 생성 한다. 쿠쿠샌드박스를 이용해 생성한 리포트로부터 문자열 추출 기법을 적용하여 피처를 생성한다. 생성된 피처를 이용하여 딥러닝 모델을 학습을 하며 모델 및 피처를 보안할 예정이다.

5) 데이터 처리

본 프로젝트는 대량의 데이터가 저장되고 그 데이터는 검색을 위해 사용될 예정이다. 정적 분석과 동적 분석의 리포트를 처리해야하는 프로젝트 특성 상 데이터 필드의 유연성이 요구되며 빅 데이터 처리의 속도 이슈가 예상되므로 기존의 RDBMS 보다 빠르고 유연한 NoSQL 기반 데이터 베이스 도입을 목표로 할 것이다. 데이터 검색의 경우 대용량 검색 처리를 해야 하므로 역 인덱싱 기법을 이용하는 엘라스틱서치 검색엔진 도입을 목표로 하고있다. 파일 유사도 검사를 하기 위해 어떠한 알고리즘을 사용하는 것이 정확하고 빠른 결과를 얻을 수 있는지 연구한다.

6) 악성 코드 크롤러

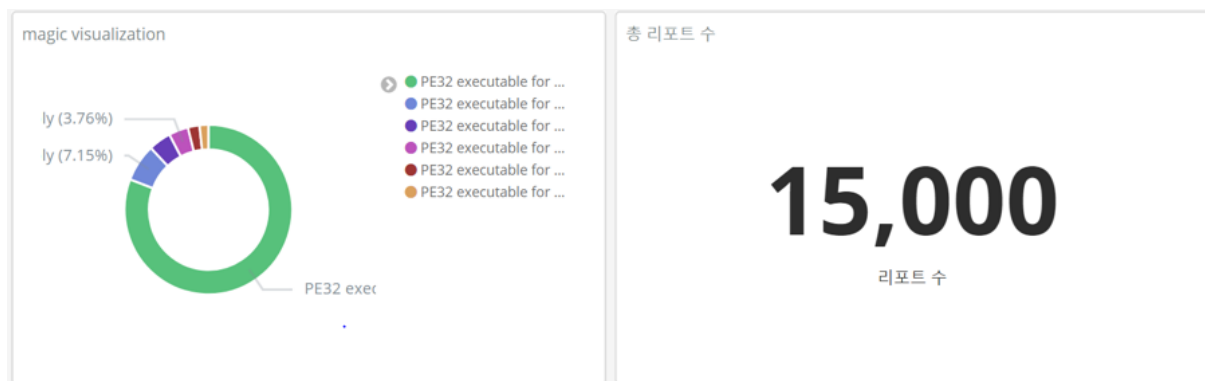
사이트 이름	사이트 주소	다운로드 조건
contagiodump	http://contagiodump.blogspot.kr/	무료(가입필요)
kernelmode	http://www.kernelmode.info/forum/	무료(가입필요)
malshare	http://malshare.com/	무료
AVCesar	https://avcaesar.malware.lu/	무료(가입필요)
malwareblacklist	http://www.malwareblacklist.com/	무료(가입필요)
malwr	https://malwr.com/	무료(가입필요)
minotaur	http://minotauranalysis.com/	무료(링크만 공유)
openmalware	http://openmalware.org/	무료
secuoxlabs	http://secuoxlabs.fr/	무료
virusign	http://www.virusign.com/	무료
virusshare	http://virusshare.com/	무료(회원 초대 필요)

[표 1] 악성코드 크롤러 사이트 리스트[출처=서준석 보안프로젝트 부대표]

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

악성 코드를 수집을 자동화 하기 위한 크롤러를 제작할 것이다. 먼저 무료 악성 코드 공유 사이트를 조사를 한 다음, 자동으로 악성코드를 크롤링하는 코드를 작성해 악성코드를 수집할 생각이다. 수집된 악성코드는 서비스를 운영하면서 새로 수집된 악성코드와 함께 정적 및 동적 분석을 한 다음 딥러닝 모델로 주기적으로 학습을 할 것이다.

7) 웹



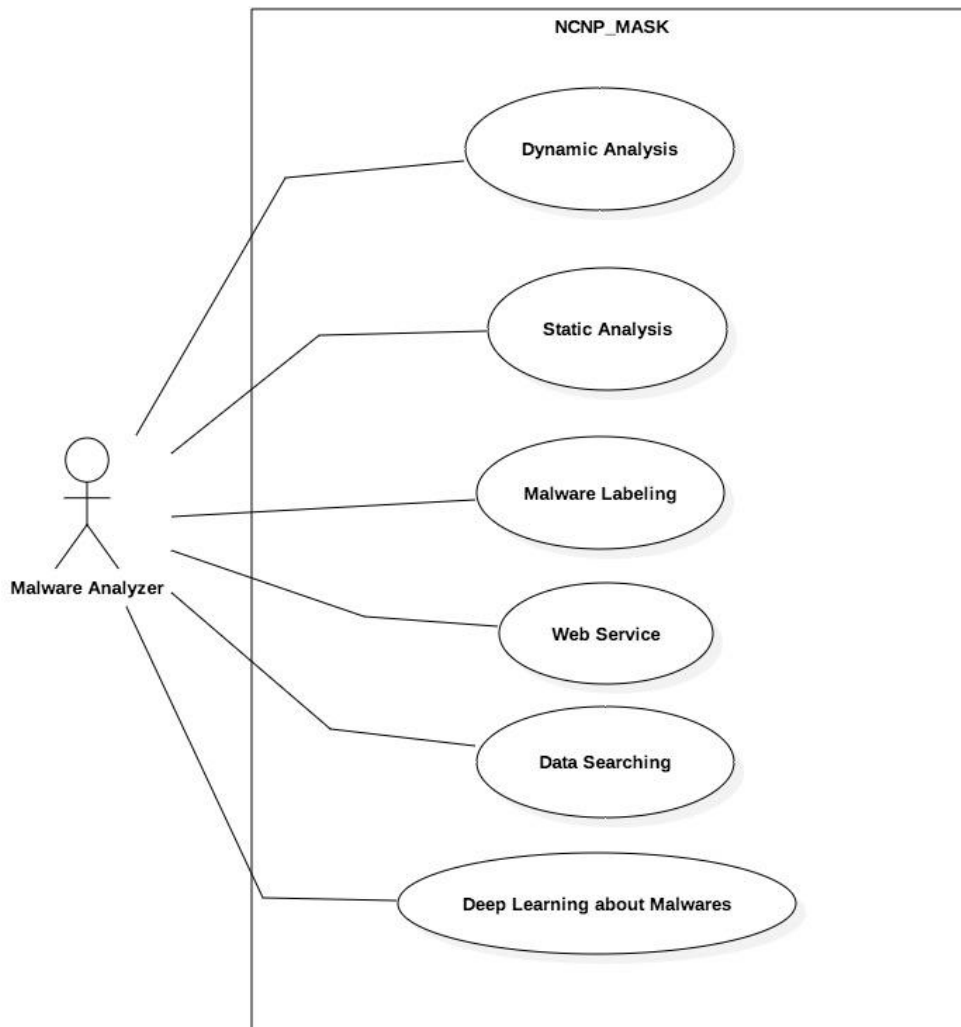
[그림 14] 예상 시각화 결과

분석된 악성코드의 세부정보를 사용자가 쉽게 알 수 있도록 웹에 시각화 하기 위해 Django를 사용하여 웹 서버를 구축하고 데이터베이스와 연동할 수 있도록 한다. 유저가 악성으로 의심되는 파일을 웹에 업로드 하게 되면, md5 값을 구하여 database로 넘겨진다. 분석된 결과는 기존 데이터들에 대한 요약 정보를 텍스트와 이미지로 시각화 하여 보여 주고, 업로드한 악성코드의 행동 정보를 그래프로 연결하여 나타내도록 한다. 그래프는 Nwagon 오픈소스를 사용할 예정이며, javascript와 jquery가 프론트엔드에 기본적으로 사용된다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

2.3 개발 결과

2.3.1 시스템 기능 요구사항



[그림 15] 시스템 기능 요구사항 use case diagram

1) Dynamic Analysis

- 악성코드를 분석환경에서 실행시켜 행동 변화를 확인하는 분석 방법이다.
- 악성코드에 대해 동적 분석을 진행한다.
- 악성코드 분석을 위한 시스템으로는 오픈소스 소프트웨어인 Cuckoo Sandbox 를 이용한다.

2) Static Analysis

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

- 실제 실행 없이 컴퓨터 소프트웨어를 분석하는 방법이다.
- 악성코드에 대해 정적 분석을 진행한다.
- 악성코드 분석을 위한 시스템으로는 컴퓨터 소프트웨어용 디스어셈블러인 IDA 와 파이썬 기반의 오픈소스 도구인 peframe 을 사용한다.

3) Malware Labeling

- 분석한 악성코드에 대해 라벨을 부여한다.

4) Data Searching

- 많은 양의 샘플 데이터를 효과적으로 검색 및 저장 할 수 있어야 한다.
- 요청한 파일과 유사한 파일에 대한 검색을 지원한다.

5) Deep Learning about Malwares

- 라벨을 붙이기 위해 데이터를 학습시켜 모델을 만든다.

2.3.2 시스템 비기능(품질) 요구사항

1) 보안성

본 서비스는 악의적인 대용량 파일의 업로드로 인한 분석 시간 지연으로 시스템이 마비될 경우를 고려하여 일정 크기 이상의 대용량 데이터 업로드를 제한한다.

2) 사용성

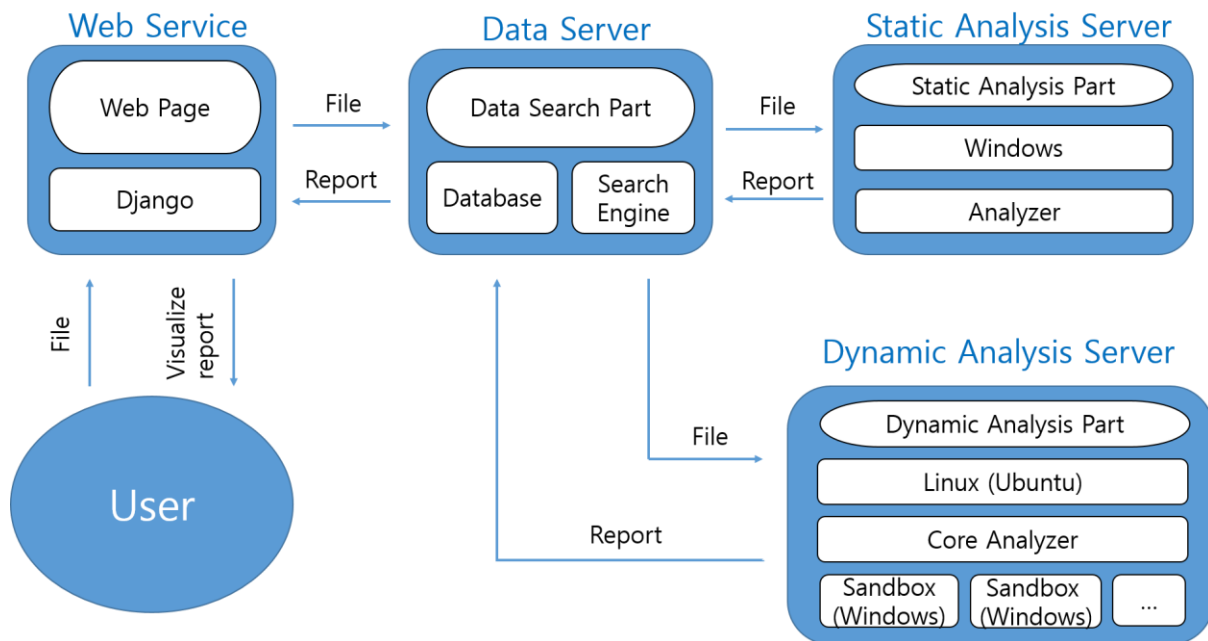
본 서비스의 목적은 악성코드 분석 전문가를 대상으로 악성코드 분석결과와 유사한 악성코드에 대한 정보를 보여 주어 전문가가 분석 할 수 있게 도와주는 것이다. 따라서 분석에 용이 하도록 기존 데이터들에 대한 요약 정보를 시각화 하여 보여 주고, 업로드한 악성코드의 행동정보를 그 래프로 연결하여 나타내도록 한다.

3) 성능

본 서비스는 악성코드 분석으로 정적 분석과 동적 분석을 사용하게 되는데 동적 분석의 경우 악성코드의 행동 분석을 하는데 정적 분석에 비해 오랜 시간이 걸리게 된다. 따라서 여러 개의 인스턴스들을 구성하여 병렬적으로 분석할 수 있도록 가이드라인을 제시해야 한다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

2.3.3 시스템 구조




[그림 16] 시스템 구조도

- 웹 서비스 : 사용자가 파일을 업로드 할 수 있도록 웹 서비스를 구성한다. 웹 서비스는 파이썬 기반의 웹 프레임워크인 Django를 이용하여 구성한다.
- 데이터 서버 : 사용자가 업로드한 파일의 분석결과가 데이터베이스에 존재하는지 검색한다. 만약 존재하지 않는다면 정적, 동적 분석 서버에 각각 업로드한 파일을 전송한다. 분석이 완료되어 분석 결과가 도착하면 데이터베이스에 저장하고 웹으로 전송한다.
- 정적 분석 서버 : 정적 분석을 완료한 후 분석 리포트를 생성한다.
- 동적 분석 서버 : 동적 분석 코어와 분석이 이루어질 1개 이상의 샌드박스로 구성한다. 각각의 샌드박스가 분석을 완료하면 분석 포트를 생성한다.

2.3.4 결과물 목록 및 상세 사양

대분류	소분류	기능	형식	비고
웹 서비스	레이아웃	사용자에게 편리한 User Interface 를 제공한다.		
	파일 업로드	사용자가 악성으로 의심되는 파일을 업로드한다.	모듈	

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

	데이터베이스 연동 및 검색	업로드한 파일에 대해 데이터베이스에서 검색한다.	모듈	
	대시보드	데이터베이스에 있는 데이터를 시각화하여 사용자에게 보기 쉽게 보여준다.		
악성코드 분석 서비스	정적 분석	Peframe 과 IDA pro 를 사용하여 파일에 대해 정적 분석한다.	모듈	
	동적 분석	Cuckoo Sandbox 를 사용하여 파일에 대해 동적 분석한다.		
딥러닝	학습 모델	정적 분석과 동적 분석으로부터 피처를 추출하여 학습하고 테스트를 지원한다.	모듈	

2.4기대효과 및 활용방안

(1) 기대효과

악성코드 분석 전문가가 악성코드로 의심되는 파일을 효율적으로 분석할 수 있다. 최근 생성되는 악성코드는 신종 뿐만 아니라 기존 악성코드의 변종이 많기 때문에, 유사한 악성코드가 포함된 분류를 알아내기만 하면 이후의 행위들을 예측하여 대응하기 용이하다. 더 나아가 악성코드 분류 기술에 딥러닝을 적용하여 분석 효율성을 높인다. 우리 서비스는 시그니처 기반 탐지가 아닌 행위 기반 탐지 시스템이기 때문에 악성코드 행동 예측을 함으로써 사이버공격에 대한 선제적 대응 능력을 확보할 수 있고, 신종/변종 악성코드에 대한 피해의 최소화를 기대한다.

(2) 활용 방안

악성코드 분석 전문가가 자신의 환경에 맞게 설치 및 서비스 이용을 할 수 있다. 우리와 유사한 타 서비스들 중 대부분은 오픈소스 소프트웨어가 아닌 것에 비해, 우리 프로젝트는 오픈소스 소프트웨어이므로 전문가가 자신만의 서비스를 구축함으로써 악성코드 분석 시장이 활성화되는 것을 기대한다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

3 배경 기술

3.1 기술적 요구사항

1) 개발 환경

- Xeon® CPU E5-2620 v3, 2T HDD * 5, 64G RAM, Ubuntu 16.04
- Xeon® CPU E5-2630 v2, 4T HDD * 5, 1T HDD * 1, 56G RAM, Ubuntu 16.04
- Intel® Core™ i7-7700, GTX 1080 Ti 11G, 1T SSD + 512G SSD + 4T HDD, 64G RAM

2) 프로젝트 결과물 확인 환경

- Xeon® CPU E3-1231 v3, 32G RAM

3) 악성코드

- Virussign
- KISA R&D challenge

4) 오픈소스 소프트웨어

- Cuckoo Sandbox 2.0.4
- peframe
- Tensorflow 1.6

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

3.2현실적 제한 요소 및 그 해결 방안

3.2.1 하드웨어

- AV-TEST 에 따르면 2017 년 기준으로 하루에 대략 30 만 개의 신규 악성코드가 유입된다. 이를 우리 서비스를 이용해서 동적 분석하기 위해서는 프로세서 자원이 많이 필요하다. 현실적인 방안으로 좋은 서버를 구축하는 거보다 비교적 저 사양의 서버들을 많이 확보하여 분산 환경 구축을 통해 가격 효율성을 높이는 것을 목표로 한다.


3.2.2 소프트웨어

- 이진 파일로부터 어셈블리어를 추출할 수 있는 디컴파일러가 필요하다.
- PE 파일 외에는 지원하지 않는다.
- 악성코드가 실행 코드 암호화나 패킹(Packing) 등 은닉기술이 적용될 경우, 혹은 신규 악성코드가 발견되었을 경우 정적 분석에 불리하다. 또한, 악성코드가 분석 환경을 인지하여 회피가 가능할 경우 동적 분석에 불리하다. 따라서 본 프로젝트는 정적 분석과 동적 분석을 동시에 사용하여 두 분석 방법의 단점을 상호보완한다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

4 프로젝트 팀 구성 및 역할 분담

이름	역할
김영재	<ul style="list-style-type: none"> - 정적분석 환경 구축 - 딥러닝 모델 설계 및 구축 - 악성코드 크롤러 제작
명준우	<ul style="list-style-type: none"> - VirusTotal 분석 리포트 수집 - 동적, 정적 분석 정보를 이용하여 파일 간 유사도 추출 - 악성코드 라벨링
이유정	<ul style="list-style-type: none"> - 웹 프론트엔드 제작 - 포스터, 팀 로고 등 디자인
한채연	<ul style="list-style-type: none"> - Software Project Leader, 문서작업 - 동적 분석 시스템 최적화, 자동화 - Report로부터 유용한 피쳐들 추출 - 논문 분석 및 연구
허준녕	<ul style="list-style-type: none"> - 프로젝트 인프라 관리 - 데이터베이스 설계 및 SQL 작성, 관리, 검색엔진 구축 - ssdeep 을 이용한 파일 유사도 분석 - 웹서버 구축 및 관리

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09


5 프로젝트 비용

항목	예산치 (MD)
분석 환경 구축	20
Database 및 검색 엔진 환경 구축	20
서비스 알고리즘 구성	40
웹 서비스 구현	40
동적 분석 시스템 최적화	25
학습 모델 구성	40
프로젝트 테스트 및 유지보수	20
프로젝트 평가 및 보고서 작성	20
합	185

6 개발 일정 및 자원 관리

6.1 개발 일정

항목	세부내용	1월	2월	3월	4월	5월	6월	비고
요구사항분석	요구 분석							
	정보 수집							
관련분야연구	주요 기술 연구							
	관련 시스템 분석							
설계	시스템 설계							
구현	코딩 및 모듈 테스트							
테스트	시스템 테스트							

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

6.2 일정별 주요 산출물

마일스톤	개요	시작일	종료일
계획서 발표	개발 환경 구축 산출물 : 계획서 1. 프로젝트 수행 계획서 2. 계획서 발표 ppt	2018-03-02	2018-03-09
설계 완료	시스템 설계 완료(IDA 및 Cuckoo 환경 설치, 기본 응용 작성 및 테스트 완료) 산출물 : 1. 프로젝트 1 차 중간 보고서 2. 프로젝트 진도 점검표 3. 1 차분 구현 소스 코드	2018-03-09	2012-03-20
중간 보고	개발환경 및 프로그램 기본기능구현 완료 산출물 : 1. 프로젝트 중간 보고서 2. 프로젝트 진도 점검표 3. 1 차분 구현 소스 코드	2012-03-21	2018-04-12
구현 완료	시스템 구현 완료 산출물: 악성코드 분석 프로그램	2018-04-01	2018-04-15
테스트	시스템 통합 테스트 산출물: 악성코드 분석 프로그램 최종본	2018-04-16	2018-05-13
최종 보고서	최종 보고 산출물: 최종 보고서	2018-05-14	2018-05-29

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

6.3인력자원 투입계획

이름	개발항목	시작일	종료일	총개발일(MD)
전원	<i>Project Study</i>	2018-01-02	2018-03-02	20
전원	<i>프로젝트 절차 구성</i>	2018-02-24	2018-02-28	5
한채연	<i>쿠쿠샌드박스 설치 및 환경 구축 시스템 자동화 및 최적화 동적분석 피쳐 추출</i>	2018-03-02	2018-05-20	70
김영재	<i>정적분석 환경 구축 딥러닝 모델 설계 및 구축 악성코드 크롤러 제작</i>	2018-03-02	2018-05-20	70
명준우	<i>VirusTotal 분석 리포트 수집 동적, 정적 분석 정보를 이용하여 파일 간 유사도 추출 악성코드 라벨링</i>	2018-03-02	2018-05-20	70
이유정	<i>웹 프론트엔드 제작 포스터, 팀 로고 등 디자인</i>	2018-03-02	2018-05-20	70
허준녕	<i>데이터베이스 구축 및 설계 데이터베이스 검색엔진 연동 파일 유사도 검사 기능 구현 웹 서버 구축</i>	2018-03-02	2018-05-20	70
전체	<i>성능 테스트</i>	2018-05-21	2018-05-29	14

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

6.4 비 인적자원 투입계획

항목	Provider	시작일	종료일	Required Options
개발용 PC 2 대	조립 PC	2018-01-02	2018-05-29	
개발용 노트북 2 대	Lenovo	2018-03-02	2018-05-29	

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	계획서		
	프로젝트 명	MASK(Malware Analysis System in Kookmin)	
	팀 명	NCNP	
	Confidential Restricted	Version 1.7	2018-MAR-09

7 참고 문헌

번호	종류	제목	출처	발행년 도	저자
1	웹 페이지	virustotal	https://www.virustotal.com/ko/about/		
2	기사	그림 1	http://www.econovill.com/news/articleView.html?idxno=314861		
3	기사	그림 2	http://www.ahnlab.com/kr/site/securityinfo/secunews/secuNewsView.do?seq=26509		
4	기사	그림 3	http://www.boannews.com/media/view.asp?idx=58181		
5	웹 페이지	그림 4	https://www.av-test.org/en/statistics/malware/		
6	기사	그림 5	http://www.boannews.com/media/view.asp?idx=58349		
7	기사	2017 년을 강타한 주요 보안이슈 5 가지는?	http://www.boannews.com/media/view.asp?idx=58065		
8	기사	안랩(AhnLab)	http://www.ahnlab.com/kr/site/securityinfo/secunews/secuNewsView.do?seq=26045		
9	웹 페이지	Malwares.com	http://story.malwares.com/		
10	웹 페이지	Hybrid-analysis	https://www.payload-security.com/technology/hybrid-analysis		
11	웹 페이지	그림 11	http://docs.cuckoosandbox.org/en/latest/introduction/what/		
12	논문	지능형 악성코드 분석을 위한 리얼머신 기반의 바이너리 자동실행 환경	KIISE Transactions on Computing Practices, Vol. 22, No. 3, pp.139-144	2016.3	