a security insight

캡스톤 디자인I 5조 어시스트

01

배경과 필요성

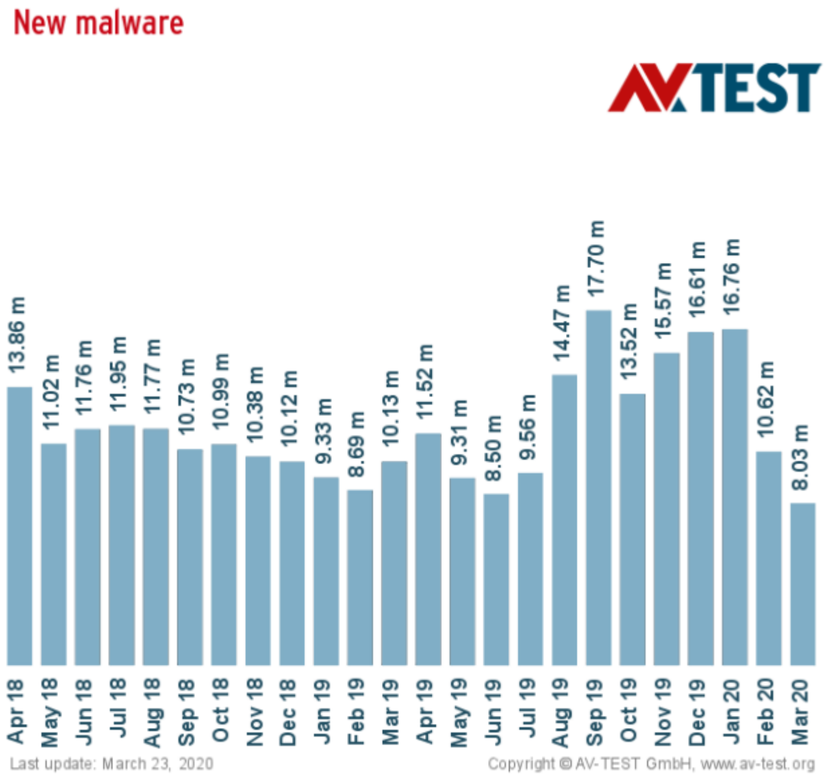
02

프로젝트 소개

03

개발 계획

늘어나는 신종 악성코드

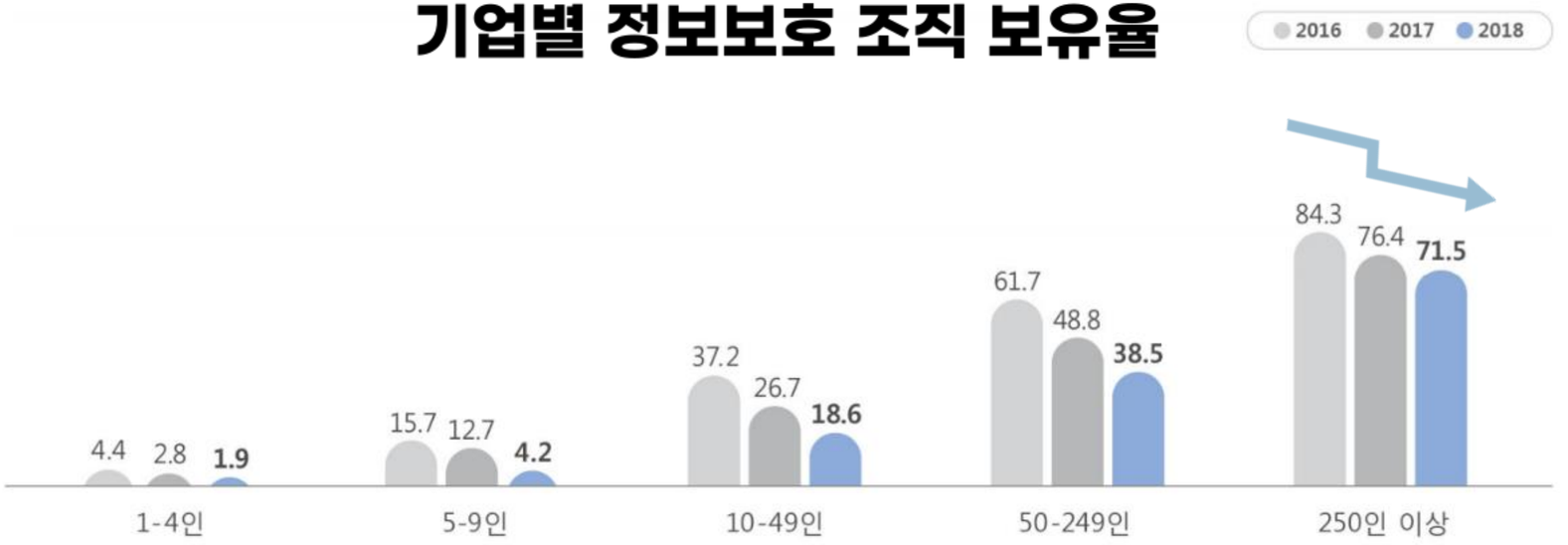


[그림 1] AV-Test의 신종 악성코드 발생량 통계조사 결과 [1]

01 매일 생성되는 350,000 개의 악성코드

02 한정된 전문가 인력 ∿→ 자동화 분석 도구

전문인력 부족



[그림 2]과학기술 정보통신부, 2018년 정보보호 실태조사 [2]

약 94%
정보보호 조직을 보유하지 않은 기업

약 63%
정보보호/개인정보보호에 투자하지 않는 기업

약 89%
IT예산 중 1%미만을 투자하는 기업

악성코드 분석 소요시간

최소

시간 단위

최대

주 단위

[기사 1]드라마 유령 ' 속 악성코드 , 실제로는? [3]

01

02

03

배경과 필요성

프로젝트 소개

개발 계획

전문가 분석 보조 도구

자동 분석 도구

연구/투자 활발

전문가 분석 보조 도구

연구/투자 부족

asi - a security insight

악성 코드 분석 보조 도구

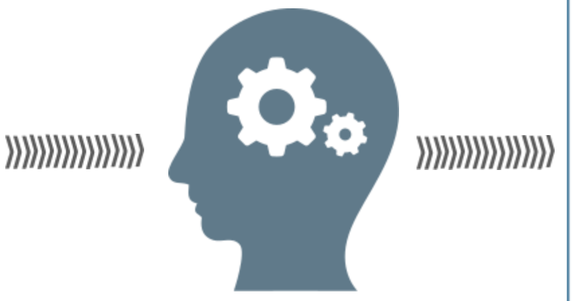
asi - 핵심 아이디어

01 파일 입력

02 disassemble

```
10001180:    55                   push   %ebp
10001181:    8b ec                mov    %esp,%ebp
10001183:    6a ff                push   $0xffffffff
10001185:    68 21 80 00 10       push   $0x10008021
1000118a:    64 a1 00 00 00 00    mov    %fs:0x0,%eax
10001190:    50                   push   %eax
10001191:    83 ec 08             sub    $0x8,%esp
10001194:    a1 00 40 02 10       mov    0x10024000,%eax
10001199:    33 c5                xor    %ebp,%eax
1000119b:    89 45 f0             mov    %eax,-0x10(%ebp)
1000119e:    53                   push   %ebx
1000119f:    56                   push   %esi
100011a0:    57                   push   %edi
100011a1:    50                   push   %eax
100011a2:    8d 45 f4             lea    -0xc(%ebp),%eax
100011a5:    64 a3 00 00 00 00    mov    %eax,%fs:0x0
100011ab:    6a 14                push   $0x14
100011ad:    6a 0c                push   $0xc
100011af:    ff 15 44 b2 01 10    call   *0x1001b244
100011b5:    83 c4 08             add    $0x8,%esp
100011b8:    89 45 ec             mov    %eax,-0x14(%ebp)
100011bb:    8b c8                mov    %eax,%ecx
100011bd:    c7 45 fc 00 00 00 00 movl   $0x0,-0x4(%ebp)
100011c4:    e8 c7 04 00 00       call   0x10001690
```

03 RNN



04 악성 행위 예측

```
10001180:    55                   push   %ebp
10001181:    8b ec                mov    %esp,%ebp
10001183:    6a ff                push   $0xffffffff
10001185:    68 21 80 00 10       push   $0x10008021
1000118a:    64 a1 00 00 00 00    mov    %fs:0x0,%eax
10001190:    50                   push   %eax
10001191:    83 ec 08             sub    $0x8,%esp
10001194:    a1 00 40 02 10       mov    0x10024000,%eax
10001199:    33 c5                xor    %ebp,%eax
1000119b:    89 45 f0             mov    %eax,-0x10(%ebp)
1000119e:    53                   push   %ebx
1000119f:    56                   push   %esi
100011a0:    57                   push   %edi
100011a1:    50                   push   %eax
100011a2:    8d 45 f4             lea    -0xc(%ebp),%eax
100011a5:    64 a3 00 00 00 00    mov    %eax,%fs:0x0
100011ab:    6a 14                push   $0x14
100011ad:    6a 0c                push   $0xc
100011af:    ff 15 44 b2 01 10    call   *0x1001b244
100011b5:    83 c4 08             add    $0x8,%esp
100011b8:    89 45 ec             mov    %eax,-0x14(%ebp)
100011bb:    8b c8                mov    %eax,%ecx
100011bd:    c7 45 fc 00 00 00 00 movl   $0x0,-0x4(%ebp)
100011c4:    e8 c7 04 00 00       call   0x10001690
```

악성행위 의심 영역 하이라이팅

asi - 웹 서비스

<파일 업로드 시각화 안>

asi - 웹 서비스

<분석 결과 시각화 안>

01 배경과 필요성

02 프로젝트 소개

03 개발 계획

데이터 수집

데이터 수 : 정상 15만 + 악성 2만 = 17만

전처리 - op code 추출

disassemble

IDA

```
10001180:    55                    push   %ebp
10001181:    8b ec                 mov    %esp,%ebp
10001183:    6a ff                 push   $0xffffffff
10001185:    68 21 80 00 10        push   $0x10008021
1000118a:    64 a1 00 00 00 00     mov    %fs:0x0,%eax
10001190:    50                    push   %eax
10001191:    83 ec 08              sub    $0x8,%esp
10001194:    a1 00 40 02 10        mov    0x10024000,%eax
10001199:    33 c5                 xor    %ebp,%eax
1000119b:    89 45 f0              mov    %eax,-0x10(%ebp)
1000119e:    53                    push   %ebx
1000119f:    56                    push   %esi
100011a0:    57                    push   %edi
100011a1:    50                    push   %eax
100011a2:    8d 45 f4              lea    -0xc(%ebp),%eax
100011a5:    64 a3 00 00 00 00     mov    %eax,%fs:0x0
100011ab:    6a 14                 push   $0x14
100011ad:    6a 0c                 push   $0xc
100011af:    ff 15 44 b2 01 10     call   *0x1001b244
100011b5:    83 c4 08              add    $0x8,%esp
100011b8:    89 45 ec              mov    %eax,-0x14(%ebp)
100011bb:    8b c8                 mov    %eax,%ecx
100011bd:    c7 45 fc 00 00 00 00  movl   $0x0,-0x4(%ebp)
100011c4:    e8 c7 04 00 00        call   0x10001690
```

parser

자체 개발

```
push
mov
push
push
mov
push
sub
mov
xor
mov
push
push
push
push
lea
mov
push
push
call
add
mov
mov
movl
call
```

file

assembly code

op code

전처리 - 특징벡터 축소

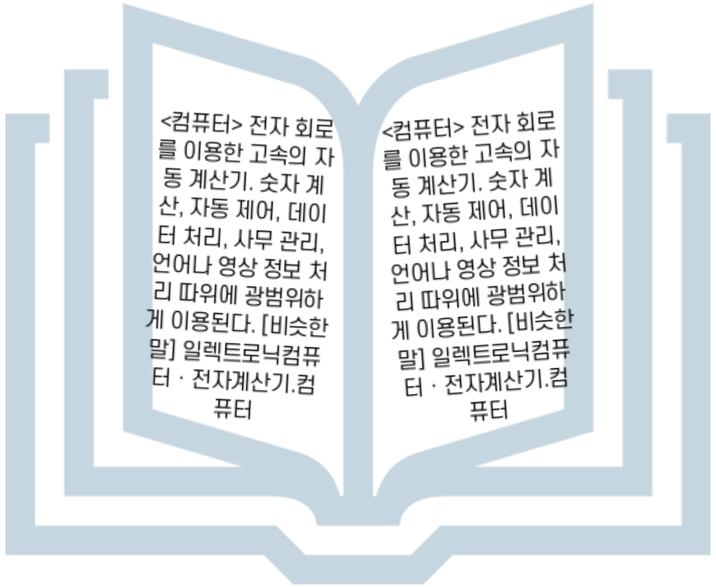
push
mov
push
push
mov
push
sub
mov
xor
mov
push
push
push
push
lea
mov
push
push
call
add
mov
mov
movl
call

sub
mov
xor
mov
push
push
push
push
lea
mov
push
push

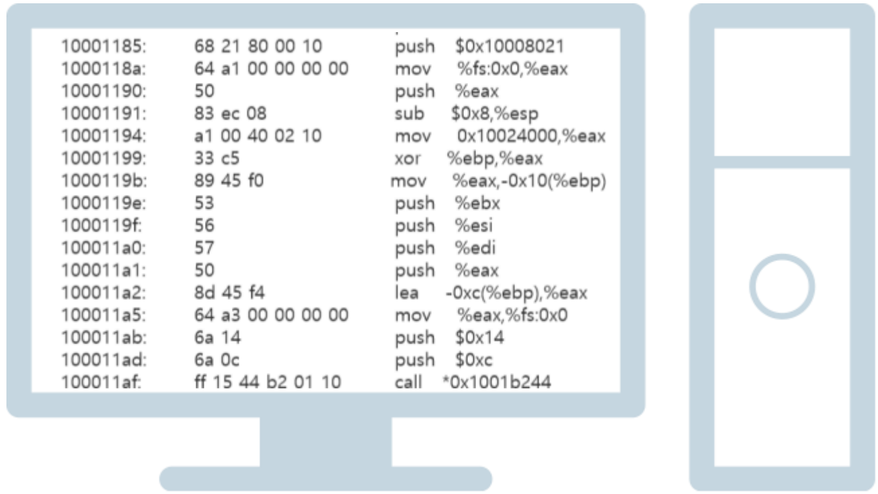
01 one-hot 인코딩

02 단어 임베딩

모델 구현 - RNN 근거



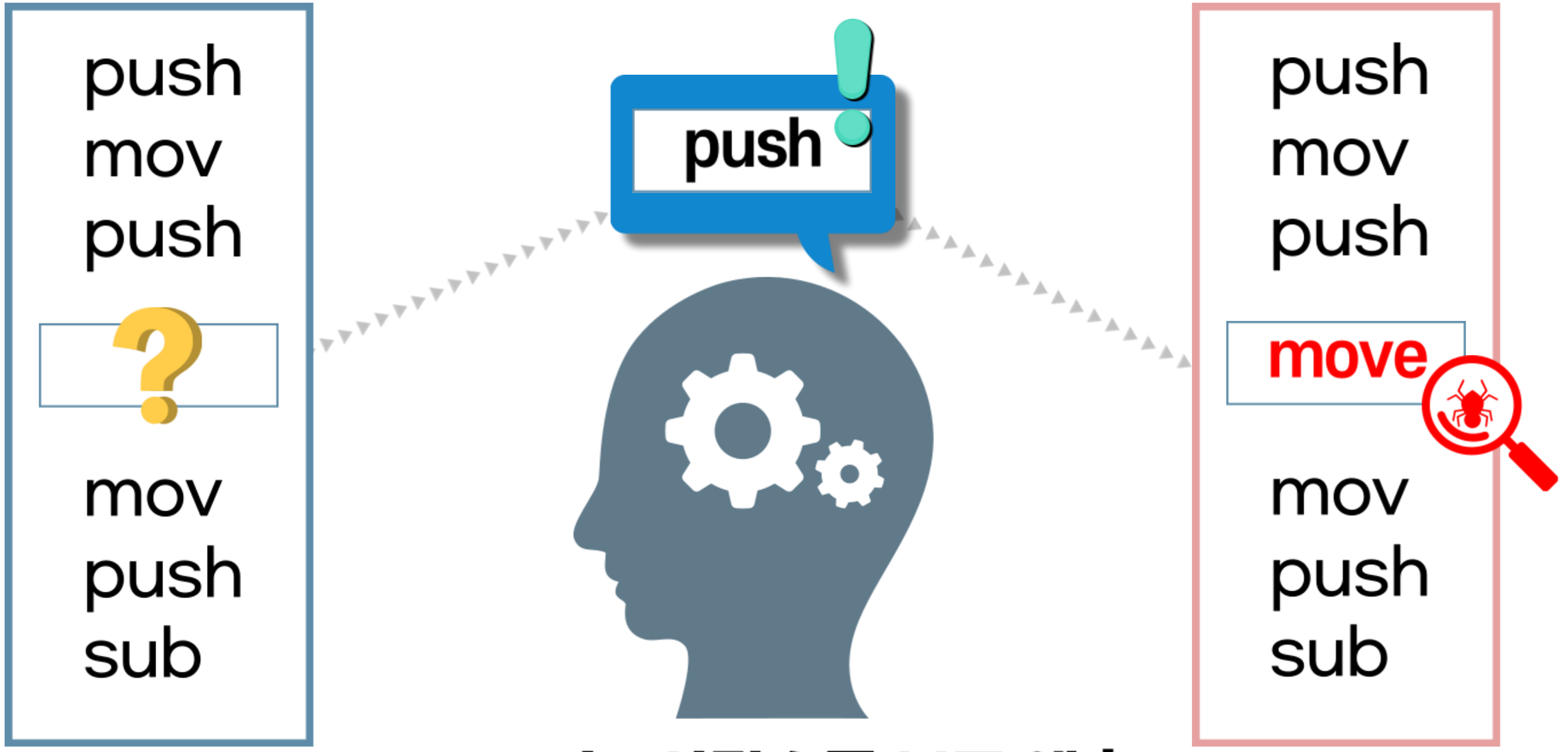
시퀀스 구조



자연어

assembly code

모델 구현 - 방법 2



opcode 시퀀스를 보고 예측

모델 구현 - 고려사항

하이퍼 파라미터 ~~> 실험적 도출

01 윈도우 사이즈

02 중심 명령어 위치 선정

03 신경망 종류
- vanilla RNN
- LSTM
- GRU
- ...

웹 구현

프론트엔드

백엔드

월별 구현 계획

항목	세부내용	1월	2월	3월	4월	5월	6월
요구사항분석	요구 분석	✓					
	SRS 작성	✓					
관련분야연구	딥러닝 기술 연구		✓	✓			
	관련 논문 동향조사		✓	✓			
설계	시스템 설계				✓	✓	
구현	코딩 및 모듈 테스트				✓	✓	
테스트	시스템 테스트						✓

팀원 별 역할 분담

손현기

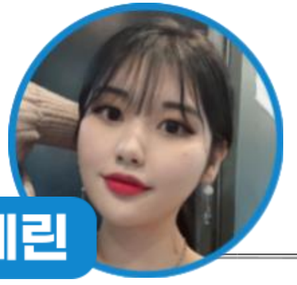
정상파일 크롤러 개발
opcode 파서 개발
신경망 구현 및 튜닝
서버 구축

김주환

논문 동향조사
제안서 및 보고서 작성
신경망 구현 및 튜닝

김호준

자료 조사
문서작업 보조
웹 프론트 개발



오예린

디자인(발표자료, 로고 등)
웹 UI/UX 기획
검색엔진(추가 시도)

이동윤

정상파일 크롤러 개발
신경망 구현 및 튜닝

ruslan

opcode 파서 개발
웹 프론트 개발

참고 문헌

번호	종류	제목	출처	발행년도	저자	기타
[1]	기사	Malware Statistics	https://www.avtest.org/en/statistics/malware/	2020	AV Test	
[2]	보고서	2018년 정보보호 실태조사	https://www.msit.go.kr/web/msipContents/contentsView.do?cateId=_status&artId=1513388	2018	과학기술정보통신부, 한국정보보호산업협회	
[3]	기사	드라마 '유령 ' 속 악성코드 , 실제로는	https://www.ahnlab.com/kr/site/securityinfo/secunews/secuNewsView.do?cmd=print&seq=19768&menu_dist=3	2012	AhnLab	

감사합니다.