연구논문/작품 제안서

2020 년도 제 2 학기

논문/작품	○논문(O) ○작품( ) ※ 해당란에 체크
제목	클라우드 기반 머신러닝 서비스 보안 프레임워크
GitHub URL	https://github.com/Seo-han-gyeol/Graduate-Report.git
팀원명단	서한결 (인) (학번: 2015311152) 강동윤 (인) (학번: 2015312912)

2020 년 9 월 25 일

지도교수 : 이 호 준 서명

1. 과제의 필요성

1. 1 Abstract

AI 기술이 적용된 서비스 제공에 강제되는 높은 메모리 사용량을 해결하기 위해 일반적으로 클라우드 컴퓨팅 기술을 이용합니다. 클라우드 기반 서비스는 애플리케이션 개발자로 하여금 메모리 사용량에 대한 걱정을 덜어주어 성능적인 부분을 좀 더 신경 쓸 수 있게 하며 이용자는 편리하게 양질의 서비스를 제공받을 수 있게 합니다. 하지만 보안 대책이 미흡한 클라우드 서비스는 서비스를 제공받아 얻는 이익만을 생각하기에는 보안사고로 인한 피해가 막대할 수 있습니다. AI 기술이 인간의 삶에 깊이 파고든 현 상황에서 우리가 사용하는 AI 기술이 적용된 애플리케이션 그 중에서 많은 부분을 차지하고 있는 클라우드 기반 애플리케이션의 보안은 그 중요도가 높다고 할 수 있습니다. 이를 위해 본 논문에서는 클라우드 기반 AI 서비스를 분석하여 어떤 공격이 이루어질 수 있는지 분석하고 그에 대한 방어법으로 기존에 개발된 하드웨어 기반 보안을 활용하여 하나의 AI 서비스 보안 프레임워크를 만들어 보려고 합니다.

1. 2 키워드

인공지능(AI), 클라우드 컴퓨팅, 클라우드 서비스, 신뢰 실행 환경(TEE), 머신러닝(ML)

1. 3 서론

머신 러닝(Machine Learning)에 대한 활발한 연구를 통해 비약적인 기술 발전이 이루어졌습니다. 사람이 구별하기 힘든 이미지를 인식하고, 방대한 양의 데이터에서 쓸모 있는 정보를 추출하는 등 인간의 능력을 넘어서는 기술들이 등장하며 기업에서는 모바일 또는 IoT 애플리케이션에 이런 머신 러닝 기술을 접목시켜 다양한 서비스를 제공하려는 노력을 하고 있습니다. 당연히 기업에서는 빠른 계산능력과 높은 정확도를 가진 모델을 사용하여 양질의 서비스를 제공하고자 하지만, 이는 많은 메모리 사용이 불가피합니다.

보통 클라우드 컴퓨팅 기술로 이를 해결하려 하지만 보안 대책이 미흡한 클라우드 서비스는 여러 보안 문제가 발생할 수 있습니다. 앞으로 머신 러닝 기술은 더욱

발전하여 우리 생활의 모든 방면에서 사용될 것이 분명한 이 시점에서 보안사고로 인해 제공한 개인 정보가 보호되지 못할 수도 있다는 것은 굉장히 심각한 문제입니다.

본 논문에서는 머신 러닝 기술을 제공하는 클라우드 기반 서비스에 어떤 공격을 시도할 수 있는지 다양한 관점에서 분석하고 이를 신뢰 실행 환경(TEE)과 암호화된 ONNX 포맷을 활용하여 서비스의 보안성을 높이는 시도를 하려합니다.

신뢰 실행 환경(TEE)이란 보안성을 높이기 위해 메인 프로세서 내 별도로 독립된 보안 영역이 제공하는 안전한 실행 환경으로서 하드웨어 기술의 도움으로 구현될 수 있습니다.[1] TEE 를 구현하여 그 안에서 작업을 할 수 있다면, 해당 이용자가 아닌 다른 외부 공격자 심지어는 서버 관리자 또한 input, output 데이터에 접근할 수 없습니다. 따라서 데이터 유출을 막을 수 있을 것이고 보안 영역 내부에 모델 외의 다른 추가적인 시스템 구축을 통해 보안 영역 외부에서의 데이터 손상에 대한 대책도 마련할 수 있을 것으로 기대됩니다.

머신러닝 서비스를 제공하는 애플리케이션의 경우 모델이 공개되어 있는 경우도 있지만 그렇지 않은 경우도 있습니다. 모델 그 자체가 기업의 중요한 자산일 수도 있고, 또한 모델에 대한 정보가 유출된다면 이를 이용한 공격이 가능할 수 있다는 점에서 모델에 대한 정보 역시 보호되어야 합니다. 이를 위해 암호화된 ONNX 를 사용해 볼 수 있습니다. ONNX 란 머신러닝 프레임워크 간의 변환을 가능하게 해주는 포맷입니다[2]. ONNX 를 이용해서 애플리케이션이 사용하고 있는 모델을 ONNX 포맷으로 바꾼 후 그것을 암호화하는 작업이 이루어 진다면 어떤 모델을 사용하든지 그 모델은 보호될 수 있을 것입니다.

2. 선행연구 및 기술현황

2. 1 신뢰 실행 환경(TEE)

머신 러닝 과정을 실행할 때 일반적으로 가장 많은 메모리를 차지하는 것은 model weights입니다.[3] 높은 계산 능력을 가진 모델은 그만큼 heavy하기 때문에 메모리를 많이 사용하게 됩니다. 따라서, 이용자들의 디바이스에서 직접 inference하는 것이 힘들어

지고 이를 해결하기 위해서는 클라우드를 이용하는 것이 굉장히 효율적입니다. 하지만, 기기에서 클라우드로 데이터를 전송하고 받는 과정에서 악의적인 공격에 의해 개인정보가 유출이 될 수 있는 위험 역시 존재합니다. 따라서 애플리케이션의 높은 성능과 memory의 효과적인 사용을 위해 클라우드 서비스를 이용하되, 개인정보 유출의 위험은 방지할 수 있는 연구가 진행되어야 합니다.

신뢰 실행 환경(TEE)에서 inference가 이루어 진다면, 개인정보 유출을 막을 수 있을 것입니다. TEE란 메인 프로세서의 보안 영역으로 기밀성과 무결성 측면에서 내부에 로드되는 코드와 데이터가 보호될 수 있도록 보장합니다. 표1의 하드웨어 기술들이 TEE 구현을 위해 사용될 수 있습니다.

AMD	플랫폼 시큐리티 프로세서 (PSP)
	AMD Secure Execution Environment
ARM	트러스트존
IBM	IBM 시큐어 서비스 컨테이너
Intel	신뢰 실행 기술(Trusted Execution Technology)
	SGX Software Guard Extensions
	Silent Lake

표 1. TEE 구현을 위해 사용될 수 있는 하드웨어 기술 [1]

이와 관련된 선행연구로는 Intel에서 개발한 SGX를 이용하여 TEE를 구현한 프레임워크인 Occlumency[4]가 있습니다. Occlumency는 model inference 과정을 클라우드 내부 enclave (공격으로부터 강력하게 보호되는 클라우드 상의 구역)에서 실행하여 inference 시에 발생할 수 있는 input과 output 데이터의 공격 가능성을 예방하고, 보다 더 작은 메모리 상에서 머신 러닝을 실행할 수 있게 합니다. 해당 논문에서는 메모리 사용량의 대부분을 차지하는 model weights를 enclave 외부의 구역에 저장함으로써 메모리를 더 효율적으로 사용하는 방법을 쓰고 있는데, 그 이유는 머신 러닝 과정에서 사용되는 model 들은 대부분 오픈 소스로서 인터넷 상에 이미 알려져 있기 때문에 굳이 기밀성을 유지할 필요가 없기 때문이라고 해당 논문에서는 언급하고 있습니다.

하지만, 이 model 자체를 공격하는 경우에는 상황이 달라지게 되는데, 예를 들어 임의의 input 데이터를 입력했을 때 output을 반복적으로 확인함으로써 model의 parameter를 역추적하는 형태의 공인 Model extraction attack(모델 추출 공격)을 이용하면 머신 러닝의 model을 탈취하는 것이 가능합니다.

이처럼, model 역시 보호하지 않으면 다른 누군가가 그 것을 이용해서 악의적으로 공격을 하는 것이 가능합니다. 따라서 Occlumency처럼 클라우드 상에서 model inference를 실행하여 전송되는 데이터를 보호하고 메모리를 효율적으로 사용할 수 있는 것과 동시에 외부에 두어 보호하지 않는 모델을 같이 보호하는 기법 또한 연구가 필요합니다.

2. 2 ONNX

ONNX는 Microsoft와 Facebook이 공동으로 개발하고 있는 오픈소스 라이브러리로 ONNX 표준 포맷을 정의하여 다른 다양한 머신러닝 프레임워크들을 ONNX 포맷으로 변환이 가능하게 합니다. 예를 들어, tensorflow를 이용해 만든 모델이 필요에 의해 pytorch로 만든 모델로 변경되어야 하는 상황을 가정합니다. 일반적으로는 기존에 만들었던 방식 그대로 다른 프레임 워크를 이용해 다시 모델을 만들어야 합니다. 하지만 ONNX를 이용하면 tensorflow -> ONNX format -> pytorch가 가능해집니다. 이는 개발자로 하여금 상황에 맞는 프레임워크를 사용할 수 있게 하여 모델의 성능 향상에 집중할 수 있게 해줍니다.

ONNX는 그림1과 같이 널리 쓰이고 있는 대부분의 머신러닝 프레임워크들과 호환이 되기 때문에, ONNX를 이용하면 대부분의 머신러닝 서비스를 제공하는 애플리케이션 모델을 공통된 포맷으로 변환이 가능할 것입니다. 그리고 ONNX 포맷으로 변환된 모델을 암호화할 수 있다면 클라우드 기반 서비스에서 발생할 수 있는 모델 관련 정보 유출을 방지할 수 있을 것입니다.

그림1. ONNX가 지원하는 머신러닝 프레임워크 [4]

3. 논문 전체 진행계획, 구성 및 평가

3. 1. 머신 러닝 모델 파악

연구에 앞서 가장 먼저 해야 할 것은 애플리케이션에서 사용하는 머신 러닝 모델이 어떻게 작동하는지 정확히 파악하는 것입니다. 이를 통해 해당 모델이 클라우드 상에서 작동할 때 어떤 보안 취약점이 생길 수 있는지를 알 수 있고, 앞으로 개발할 보안 프레임워크를 적용했을 때 메모리 사용량과 속도 측면에서 어떤 문제점이 발생할 수 있을지 예측이 가능할 것입니다.

3. 2. TEE 구현을 위한 효율적인 하드웨어 기술 찾기

Occlumency 논문에서는 Intel SGX를 이용하여 그의 특성 중 하나인 enclave를 활용하여 클라우드 상에서 model inference 과정을 보호하는 기법을 설명하였습니다. 이와 비슷하게, 저희가 앞으로 진행할 첫 번째 계획은 Occlumency가 Intel SGX를 사용한 것처럼 저희가 추구하고자 하는 목표를 가장 잘 만족시켜 줄 수 있고, 가장 효율이 좋다고 여겨지는 하드웨어 기술을 찾아 TEE를 구현하는 것입니다. 표1에 정리된 하드웨어 기술들이 어떻게 TEE를 구현하는데 사용될 수 있는지 조사하고 각각의 장단점을 비교하여 최적의 TEE를 구현할 것입니다.

3. 3. ONNX 암호화 알고리즘 적용

ONNX를 이용하여 다양한 머신러닝 프레임워크를 하나의 포맷으로 만들었다면 이를 보호하기 위한 암호화 과정이 필요합니다. 3. 2.의 결과로 나온 TEE와 병합하여 하나의 프레임워크로서 작동할 수 있는 암호화 알고리즘을 적용해야 합니다.

3. 4 디자인 시 고려사항

가장 우선되어야 할 것은 model inference 과정을 클라우드 상에서 실행하였을 때 input과 output 데이터, 그리고 model의 파라미터를 실제로 보호할 수 있는지 여부입니다. 그 다음으로 고려해야 할 것은 개발된 프레임워크가 효율적인 메모리 사용과 빠른 연산 속도를 가능하게 만드는 것입니다. 일반적으로 model inference 과정에서 가장 많은 메모리를 차지하는 것은 model weights입니다. 또한, model weights 뿐만 아니라 머신 러닝 과정 상에서 많은 메모리를 차지하고 연산 속도를 느리게 하는 요소들이 존재할 것입니다. 따라서 저희는 프라이버시를 보호하면서 보다 더 효율적으로 메모리를 사용하고, 최대한 높은 연산 속도를 가질 수 있게 하기 위한 연구를 할 계획입니다.

3. 5 평가

개발된 프레임워크의 성능을 확인해야 합니다. 그 것을 위해서 이 프레임워크가 적용되지 않은 상태의 model inference 과정과 비교해서, 실제로 이것이 사용자의 input과 output 데이터, model weights를 보호해 주는지, 그리고 원래에 비해서 사용되는 메모리 용량과 연산 속도가 얼마나 차이 나는지, 만약 더 많은 메모리를 사용하고 느린 연산 속도를 가진다면 메모리와 속도를 개선하지 않고 보안 기법만 적용한 상태의 model inference 과정에 비해서 얼마나 효율적인지를 비교하면서 평가할 예정입니다. 이를 통해 개발한 프레임워크가 실제로 사용 가능한 지 판단하고 강점과 약점을 파악할 수 있을 것입니다.

4. 팀원 간의 역할분담

성명: 서한결

역할: 클라우드와 머신러닝에 초점을 맞춰서 연구를 진행할 예정입니다.

1. 다양한 클라우드 기반 애플리케이션의 작동 원리를 분석하여 취약점 조사를 위한 기반을 다집니다.
2. 머신러닝 프레임워크가 ONNX 포맷으로 변환되는 과정을 자세히 조사하고 변환 과정 중 ONNX 포맷으로 변환된 모델을 암호화할 수 있는 방법을 연구합니다.

성명: 강동윤

역할: 주로 보안 쪽에 초점을 맞춰서 연구를 진행할 예정입니다. 우선 현재 학계에서 관심을 가지고 있는 보안 관련된 이슈 중 논문에 필요한 연구와 기술현황에 대해 알아볼 것입니다.

1. TEE를 구현할 수 있는 하드웨어 기술에는 Intel SGX, ARM TrustZone, AMD Secure Execution Environment 등이 있습니다. 그 것들 각각에는 장단점이 존재할 것인데, 이 하드웨어 기술들을 연구해서 그들의 장단점을 비교해서 결과적으로 연구에 어떤 하드웨어 기술을 적용할지를 결정할 것입니다.
2. 현재 머신 러닝에 사용되는 보안 기법에 대해서 연구를 하고, 좋은 점이 있다면 유지하고 안 좋다고 생각되는 점이 있다면 어떻게 개선할 지를 연구할 예정입니다.

5. 참고문헌

[1]Wikipedia TEE. https://ko.wikipedia.org/wiki/%EC%8B%A0%EB%A2%B0_%EC%8B%A4%ED%96%89_%ED%99%98%EA%B2%BD

[2] ONNX. https://onnx.ai/supported-tools.html

[3] Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", arXiv:1409.1556

[4] Taegyeong Lee, Zhiqi Lin, Saumay Pushp, "Occlumency: Privacy-preserving Remote Deep-learning Inference Using SGX", MobiCom '19, October 2019.