

DNA Sequencing 의 사례를 이용한 빅데이터 처리 클라우드 하드웨어 플랫폼의 성능 비교 연구

홍보의, 김한이, 서태원

*고려대학교 컴퓨터 학과

e-mail: boyhong@korea.ac.kr, hanyeemy@korea.ac.kr, suhtw@korea.ac.kr

A Comparative Study on the Performance of Cloud Hardware Platform for Big Data Processing using DAN Sequencing Case

BoUye Hong, Hanyee Kim, and Taeweon Suh

Dept. of Computer Science and Engineering, Korea University

요약

본 연구에서는 클라우드 컴퓨팅 환경에서 운용되는 빅데이터 처리 프로그램에 ARM 과 Intel 의 하드웨어 보안이 어떠한 방식으로 적용되는지 비교 및 분석한다. 비교를 위하여 클라우드 서비스 모델을 제시하고, 실제 빅데이터 처리 알고리즘을 ARM 과 Intel CPU 를 갖춘 기기에서 작동시켜 수행 시간을 비교하였다. 연구 결과, ARMv7 의 취약점인 하드웨어 암호화 모듈과 메모리 암호화의 부재를 도출하였고, 그 대안 방안으로서 FPGA(Field Programmable Gate Array)의 사용과 그 발전 방향을 제시하였다.

1. 서론

IT 중소기업의 입장에서 빅데이터 처리를 위한 고가의 인프라 구축은 부담되는 부분이다. 이러한 상황에서 클라우드 서비스는 대량의 데이터 처리를 요하는 비즈니스 프로그램의 처리속도를 증가시킬 수단을 제공한다. 이 과정에서 클라우드 서비스 사용자는 자신의 데이터를 서비스 제공자에게 제공하게 되며, 이렇게 클라우드 환경에 노출된 정보는 해킹의 목적으로 이용될 수 있다.

특히 최근에는 데이터 처리가 이루어지는 하드웨어 보안의 중요성이 점점 더 커지고 있으며 관련 하드웨어 제조사들은 하드웨어 보안 기술의 연구 및 개발에 적극적으로 나서고 있다. CPU 시장을 선도하는 ARM 과 Intel 은 이러한 하드웨어 보안 기술로 TrustZone 과 SGX(Software Guard Extensions)를 각각 선보이고 있다.

ARM 과 Intel 의 하드웨어 보안 기술은 클라우드 환경에서 대량의 데이터를 처리하는데 있어서 성능과 보안성에 차이를 보인다. 이는 ARM 과 Intel 의 기본적인 수행능력 차이에서 비롯되기도 하지만, 각 하드웨어 보안 기술의 특징에서 성능 차이의 이유를 찾을 수 있다.

본 연구에서는 실제 빅데이터 처리를 위한 클라우드 환경 모델을 산정한다. 그리고 두 하드웨어의 보안성에 대한 실험을 통해서 클라우드 환경에서 ARMv7 의 TrustZone 이 취약성이 있음을 제시한다. 이어서 이에 대한 해결책으로 FPGA 를 이용한 하드웨

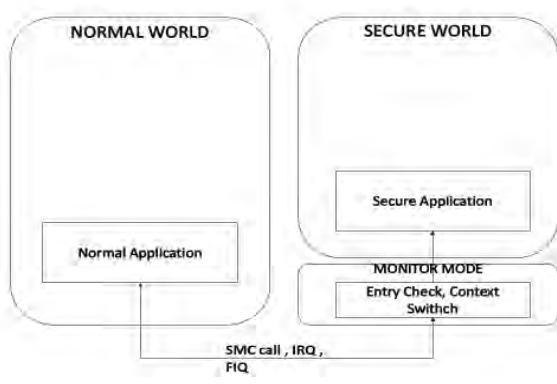
어 설계에 대해 논의한다.

2. 클라우드 환경에서의 ARM 과 Intel

클라우드 환경에서 데이터는 인터넷 네트워크를 통해서 전달된다. 이때 네트워크 상의 모든 데이터는 패킷 스니핑과 같은 MITM(Man-In-The-Middle) 공격의 목표가 된다. 또한 전달된 데이터는 데이터 센터의 저장소에서 발생하는 물리적 데이터 추출 공격이나 악의적 코드의 실행으로부터 취약하다. 따라서 이를 위한 데이터의 암호화는 필수적이다.

Intel 은 별도의 하드웨어 AES(Advanced Encryption Standard) 모듈을 통해서 암호화 성능을 가속화 한다. 또한 Intel SGX 기술은 별도의 명령어를 통해서 특정 애플리케이션에 종속적인 암호화된 메모리 영역을 생성하게 한다.

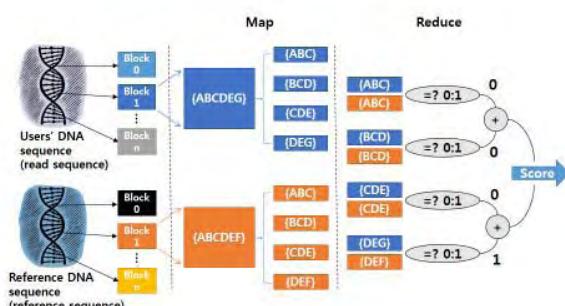
ARMv7 아키텍처 또한 TrustZone 기술을 통하여 ARM 기반 SOC 를 Secure World 와 Normal World 로 구분한다. 두 구간은 특별한 모드인 Monitor Mode 를 이용하여 접근되며, Normal World 의 애플리케이션은 Secure World 에 접근하지 못한다. (그림 1)은 이 과정을 설명한다. 하지만 이는 각 SOC 구성 성분으로의 접근을 제어할 뿐 별도의 DRAM 암호화 기능을 제공하지 않는다. 또한 TrustZone 은 AES 를 지원하는 하드웨어 가속기가 없다.



(그림 1) ARM TrustZone

3. DNA Sequencing 를 통한 ARM 과 Intel 의 비교

실험은 대량의 데이터 처리에 대한 애플리케이션이 클라우드 환경에서 제공되며, 이때 서비스를 요청하는 측의 데이터는 개인정보와 같은 민감한 정보임을 가정한다. 본 연구에서 사용한 Target 애플리케이션은 DNA Sequencing 이다. DNA Sequencing 은 개인의 DNA 문자열을 Reference DNA 에 비교하여 유사도를 측정하는 애플리케이션이며, DNA 정보의 특성상 대량의 문자열을 포함하고 있다.



(그림 2) DNA Sequencing

(그림 2)는 DNA Sequencing 의 전 과정을 나타낸다. 먼저 DNA 문자열을 Substring 단위로 자르는 맵 단계 와 각 Block 을 Substring 단위로 비교를 수행하는 리듀스 단계로 나뉜다. 네트워크를 통해 데이터가 전송되는 과정에서 정보 보호를 위해 모든 데이터는 암호화 과정이 필요하고, 이를 위하여 맵 단계와 리듀스 단계 사이에 AES 암호화를 사용하였다.

본 연구에서는 개인의 DNA 는 사용자가 직접 암호화하여 AES Key 와 함께 암호화된 데이터들을 클라우드 서비스 제공자에게 전달한다고 가정한다. 또한, 데이터 센터 내의 DNA Sequencing 처리는 맵과 리듀스 두 작업을 수행하는 노드들이 물리적으로 분리되어 동작하는 모델이라 가정한다. 따라서 네트워크 상의 분리된 노드들은 데이터 전송을 위해 암호화 및 복호화가 필요하다. 이를 위해 맵 노드는 사용자의 암호화된 데이터를 복호화 하여 이를 처리한 후 결과를 다시 암호화하여 리듀스 작업이 준비된 노드로 전송

해야 한다. 리듀스 노드는 맵 노드와 유사하게 받은 데이터를 복호화한 후 처리하고, 결과를 다시 암호화하여 데이터센터 루트를 통해 서비스 요청자에게 전달한다.

실험은 DNA Sequencing 의 변수인 Block 크기와 Substring 크기에 변화를 주며 이루어 졌으며 별도의 AES 하드웨어가 지원되는 Intel 의 경우 하드웨어 가속기의 사용과 미사용의 경우를 비교하였다.

3.1 ARM 의 서비스

ARMv7 기반 애플레이션 도구인 ZedBoard[2]로 수행된 DNA Sequencing 실험은 해당 아키텍처가 별도의 AES 하드웨어를 지원하지 않으므로 소프트웨어로 진행 되었다. AES 알고리즘은 공개 소프트웨어 중 공인된 방법에 의해 검증된 알고리즘을 사용하였다[3].

<표 1>은 ZedBoard 에서 수행된 DNA Sequencing 에서 Block 의 크기와 Substring 의 크기에 따른 수행시간을 보여준다. 하지만 수행시간의 대부분은 맵 리듀스 과정이 아닌 수행 DNA String 의 암호화 과정에서 소모됨을 알 수 있다.

Module	Parameter	ARM(ns)
Mapper	16block 8sub	2795
	16block 10sub	2762.5
	32block 8sub	6967.5
	32block 10sub	9172.5
	48block 8sub	11735
	48block 10sub	14217.5
Reducer	16block 8sub	3447.5
	16block 10sub	3432.5
	32block 8sub	7112.5
	32block 10sub	10395
	48block 8sub	14330
	48block 10sub	17387.5
128-bit AES	encryption	51365
	decryption	305110

<표 1> ARMv7 DNA Sequencing

하나의 예로 32 개의 문자열을 갖는 Block 을 10 개의 길이의 Substring 으로 절단한 결과의 경우 230 개의 문자로 이루어진 문자열간 비교가 이루어 지며, 1block 의 암호화에 사용되는 128-bit AES 는 총 15 번이다. <표 1>을 통해 128bit 의 AES 암호화와 복호화에 소모되는 시간이 350μs 임을 확인하면 총 0.5ms 의 시간이 암호화에 소모된다. 이에 비하여 한 블록의 맵 과 리듀스에 소모되는 시간은 10μs 에 불과하다.

3.2 Intel 의 서비스

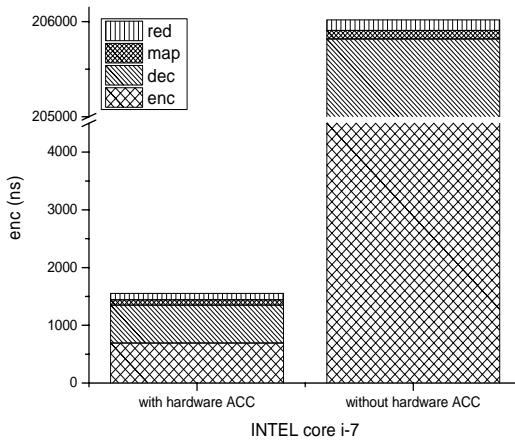
ARMv7 기반의 실험과 달리 Intel i7 CPU 를 탑재한 PC 에서의 실험은 AES 하드웨어 가속기의 지원을 받을 수 있다. (그림 3)은 Intel i7 에서 오픈소스 AES 소프트웨어를 사용한 것과 AES 가속기를 사용한 것의 수행시간을 비교한 그래프이다. 실험 결과, Intel AES 가속기는 오픈소스 AES 에 비하여 수행 시간을 100 배 이상 감소 시켰다.

클라우드 환경에서 네트워크를 통한 데이터 전송이

AES 암호화를 요구할 경우에, Intel 이 제공하는 하드웨어 가속기는 높은 성능 향상을 나타낸다.

4. ARMv7 아키텍처의 취약점

ARMv7 아키텍처는 Intel i7 과 달리 별도 AES 가속기의 부재로 인해 전체 DNA Sequencing 의 수행시간이 지연된다. 암호화 과정이 클라우드 시스템에서 네트워크의 이용과정에서 빈번히 이용됨을 고려할 때, 암호화 하드웨어 모듈의 부재로 인해 소프트웨어 암호화 작업은 전체 클라우드 서비스의 성능을 저해한다.



(그림 3) Intel i7 DNA Sequencing

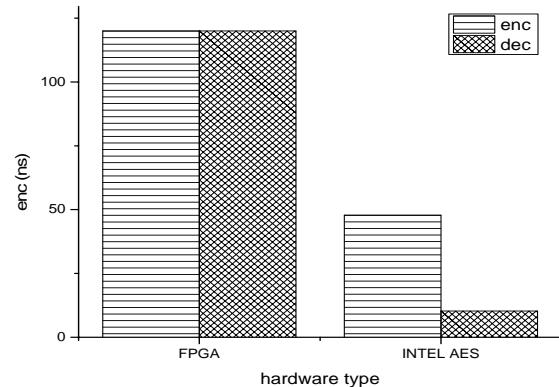
또한 TrustZone 은 데이터를 메모리에 저장할 때 하드웨어 수준의 암호화를 지원하지 않는다. 그렇기 때문에 TrustZone 에서 암호화 과정을 진행한다고 하더라도, 악의적 유저가 커널권한을 지닌다면 TrustZone 에서 암호화 이전 혹은 이후의 데이터가 추출될 가능성은 존재한다. 이에 대한 사례로서 TrustZone 에서 실행되는 디바이스 드라이버가 해킹되어 지문 이미지가 추출되었고[4], Integer Overflow 공격에 의하여 TrustZone 의 진입 절차가 해킹되어 TrustZone 으로의 Code Injection 공격이 성공 하였다[5]. 따라서 메모리 암호화를 지원하지 않는 ARMv7 TrustZone 기술의 특징상 유저 데이터의 해킹 가능성을 배제할 수 없다.

5. FPGA 를 이용한 ARMv7 하드웨어 보안

ARMv7 의 취약점에 대한 대안으로 FPGA 와 ARM CPU 와의 협동을 통한 보안 방법론이 있다[6]. 맵 리듀스의 과정에서 FPGA 내의 암호화 모듈은 서비스 이용자의 민감성 정보를 FPGA 상에서 복호화하고 해당 정보를 메모리 상에 노출시키지 않은 상태로 맵 혹은 리듀스에 해당하는 작업을 FPGA 상에서 처리한다.

이 방법은 복호화부터 맵 리듀스, 그리고 결과의 암호화까지 서비스 제공자가 설계한 FPGA 내의 하드웨어를 통해 이루어 지기 때문에 메모리 공간에서의 데이터 노출을 원천 차단한다. 또한 ARMv7 에서 FPGA 를 통한 AES 알고리즘의 병렬 수행은 하드웨

어 가속으로 인해 큰 성능 향상을 얻을 수 있다. (그림 4)는 Intel i7 의 AES 모듈과 FPGA 에 설계된 모듈간의 128-bit AES 암호화 성능을 비교한 것이다.



(그림 4) Encryption module of FPGA and Intel

FPGA 는 Intel i7 CPU 가 제공하는 암호화보다 느리지만 비슷한 수준의 성능을 갖춘 모듈의 설계가 FPGA 에 가능하며, 이를 통해 ARMv7 의 암호화에 의한 성능 저하를 어느 정도 상쇄시켰다.

6. 결론

본 연구에서는 ARMv7 과 Intel i7 CPU 를 사용한 빅데이터 처리 하드웨어 클라우드 플랫폼의 성능을 비교 및 분석하였다. 연구 결과, ARMv7 아키텍처를 이용하여 클라우드 서비스 환경을 구축할 경우에는 암호화 모듈에 의한 성능 저하와 TrustZone 해킹 가능성이라는 취약점이 있음을 확인하였다. 그리고 FPGA 로 설계한 암호화 모듈은 데이터를 입력받아서 다시 암호화된 결과를 모듈 밖으로 출력한다는 점에서 사용자 데이터가 외부에 노출되지 않는다는 장점을 가지고 있으며, 성능 또한 상용화된 암호화 모듈과 비슷한 성능을 보였다.

FPGA 를 이용하여 빅데이터 애플리케이션에 밀착된 하드웨어 가속기를 설계한다면, 다양한 종류의 애플리케이션이 성능과 보안 측면에서 향상을 볼 수 있을 것이다. 향후 클라우드 서비스를 필요로 하는 대량의 데이터 처리 애플리케이션들에 대한 하드웨어 가속기의 연구 및 개발이 지속되기를 기대해 본다.

참고문헌

- [1] S. Subashini, V.Kavitha. A survey on security issue in service delivery models of cloud computing. Journal of Network and Computer Applications 34. 2011. 1-11.
- [2] SoC with cortex-A9 available form <http://zedboard.org/product/zedboard>
- [3] <https://hub.com/kokke/tiny-AES128-C>
- [4] Di Shen. Exploiting Trustzone on Android.
- [5] Dan Rosenberg. QSEE TrustZone Kernel Integer Overflow Vulnerability.
- [6] Han-Yee Kim. Optimizing Hardware-Based Privacy-Preserving MapReduce.

- [7] ARM Ltd. TEE reference documentation.
<http://www.arm.com/products/processors/technologies/trustzone/tee-reference-documentation.php>, 2014.
- [8] ARM Ltd. Trustzone.
<http://www.arm.com/products/processors/technologies/trustzone/index.php>, 2014.