UNIVERSITÉ DE GENÈVE

FACULTY OF SCIENCE DEPARTMENT OF COMPUTER SCIENCE

MASTER'S THESIS

Confidential Computing of Machine Learning using Intel SGX.

Submitted By :

Prasad Koshy Jose

18341164

Supervisor :

Dr Eduardo Solana

Contents

Acknowledgements

List of Figures

List of Abbreviations

Chapter 1

1 Introduction

The need for data security is always on the rise. With the rise of cloud technologies, security and privacy are the most significant concerns for both enterprises and consumers. Additionally, cloud computing piles on the stress as private data is stored and processed in multi-tenant hardware providers.

Traditional systems performs access control by separating software roles into multiple user modes. Privileged system software like hyper-visors and operating systems control that manage machine resources and other vital components of the system, the user developed applications rely not only on the correctness of the system software, ie, they are free of bugs that are exploitable, but also on the honesty of system softwares. The correctness of the system software is vastly exploited by hackers as this is very hard to verify. The main reason for this is a software system comprises of millions of lines of code (LOC) and very complex control and data flows. Most of the operating systems including Windows and Linux have a lot of legacy codes which makes them susceptible to various malicious attacks. The analysis done by [1] shows more than 10000 computer systems vulnerabilities are discovered each year. It is not a far fetched assumption that these vulnerabilities still linger in cloud technologies and that an adversary would exploit them to gain access to confidential processes and data.

The situation gets trickier when it comes to 'honesty'. With the emergence of cloud computing, many applications and systems are deployed in third party infrastructure providers. While the pros of this approach are many, like cost of operations and maintenance, the cons are rather dire. The main questions that arise are 1. where is the processing unit located? 2. How many people are sharing this hardware? 3. How many have access to the data/code? The last

question is quite complex. How does one trust the infrastructure provider? even if one trusts the infrastructure provider and all of its employees with physical access or administrative access to the machines, still one should wonder about the other tenants running on the same platforms and are not malicious.

The solution is safe data processing and execution environments inside untrusted or compromised computers. They can be used for confidential data processing in public clouds where the need of protected encrypted code and data from malicious administrators and hackers.

1.1 Definition and Explanation of Key Terminologies

1.2 Structure of the Thesis

Chapter 2

2 State of the Art

2.1 Introduction

2.2 Background Information

2.3 Research Questions

2.4 Conclusion

3 Methods

4 Analysis and Discussion

5 Conclusion

References

[1] MITRE Corporation,Common vulnerabilities and exposures (CVE) details: The ultimate security vulnerability datasource. browse vulnerabilities by date, 2019. Accessed on 10/10/2019. [Online]. Available:https://www.cvedetails.com/browse-by-date.php.