LOW LATENCY, ELASTIC AND PRIVACY PRESERVING DATA STREAM PROCESSING

Buddhima Arosha Rodrigo

148237H

Degree of Master of Science

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

May 2018

LOW LATENCY, ELASTIC AND PRIVACY PRESERVING DATA STREAM PROCESSING

Buddhima Arosha Rodrigo

148237H

This dissertation submitted in partial fulfillment of the requirements for the degree Master of Science in Computer Science Specializing Parallel Computing

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

May 2018

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement of any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works.

Candidate

....

Buddhima Arosha Rodrigo Date

I certify that the declaration above by the candidate is true to the best of my knowledge and he has carried out research for the Masters Dissertation under my supervision.

Supervisors

..... ….....

Prof. Sanath Jayasena Dr. Miyuru Dayarathna

....

Date Date

Abstract

Prevalence of the Infrastructure-as-a-Service (IaaS) clouds has enabled organizations to utilize compute services on demand via elastic scaling of their applications. Data stream processing is one such area which is benefited by elastic scaling. The main drawback of using these IaaS clouds is the security risks on sensitive data in the aspect of data stream processing. It will be a great solution if we can preserve the privacy of data of data-sensitive applications, while using them in IaaS clouds with minimized security risks.

The aim of this research is to implement elastic scaling mechanism in a private/public cloud environment by preserving the privacy of the data in the aspect of stream processing. To enable the privacy preserving on data, we use the concept of Homomorphic Encryption (HE) which can perform computations on encrypted data. We designed and implemented several functions which support Homomorphic Encryption using a well-known library HElib. We extended existing Elastic Switching Mechanism (ESM) to support newly implemented HE based functions. This Homomorphic Encryption based Elastic Switching Mechanism (HomoESM) operates between the boundaries of a private and a public cloud while preserving data security.

Using two real-world data stream processing scenarios, which include an email data set and a web server access log processor data set (EDGAR), we derive four benchmark applications. Several experiments on those benchmarks indicate that, the proposed approach for Homomorphic Encryption based equal operation provides significant results which are 10% and 17% improvement of average latency when compared to private Stream Processor (SP) only case for the scenarios of Email Filter benchmark and EDGAR Filter benchmark respectively. The HE operations which consume more computations such as greater-than and less-than comparison operations, add and subtract operations, also provide beneficial results but not much as equal operation's results. Therefore, this HomoESM performance directly depends on the complexity of HE computations. In this work we use data batching technique in our HomoESM implementation by creating a composite event using several plain events in order to address Single-Instruction-Multiple-Data (SIMD) support given by HElib. This approach is the key advancement in our HomoESM which enables to realize the elastic stream processing with HomoESM. Mainly our work addresses the feasibility and limitations of using HE operations under the aspect of data stream processing in a private/public cloud environment.

Keywords: Cloud computing, Elastic data stream processing, Homomorphic Encryption, IaaS, System sizing and capacity planning.

ACKNOWLEDGEMENTS

This project would not have been possible without the support of many people. Specially thanks to my supervisors, Prof. Sanath Jayasena and Dr. Miyuru Dayarathna, for their valuable guidance and endless support. Many thanks to our M.Sc. research project coordinators, Dr. Shehan Perera, Dr. Dilum Bandara and Dr. Malaka Walpola, for their dedication and support. Thanks to all the lecturers at the Faculty of Computer Science and Engineering, University of Moratuwa, for their valuable advices. I also thank to my parents and my wife, Manorika Athukorala who always offering me support and love. Finally, I thank to my colleagues Prabath Weerasinghe, Oshan Deshapriya and numerous friends who endured this long process with me.

TABLE OF CONTENTS