

Towards Efficient and Lightweight Security Architecture for Big Sensing Data Streams

by

Deepak Puthal

M. Tech. (National Institute of Technology Rourkela)

**A thesis submitted to
Faculty of Engineering and Information Technology
University of Technology, Sydney**

**for the degree of
Doctor of Philosophy**

April 2017

To my family and friends

CERTIFICATE OF ORIGINAL AUTHORSHIP

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Student:

Date:

Acknowledgement

I sincerely express my deep gratitude to my principle coordinating supervisor, Prof. Jinjun Chen, for his experienced supervision and continuous encouragement throughout my PhD study. And I want to show my most honest appreciation to my co-supervisors, Dr. Surya Nepal and Dr. Rajiv Ranjan from CSIRO, for their supervision and encouragement. Without their consistent support and supervision, I would not have been able to complete this thesis. I express my hearty gratitude to Dr. Ranjan for his financial support, without him it may have been difficult for me to travel to Australia for PhD study.

I thank the Commonwealth Scientific and Industrial Research Organisation (CSIRO) for offering me a full Scholarship throughout my doctoral program. I also thank University of Technology Sydney (UTS) and the Faculty of Engineering and IT (FEIT) for providing me an IRS Scholarship throughout my doctoral program.

My thanks also go to staff members, research assistants, previous and current colleagues, and friends at UTS, and CSIRO for their help, suggestions, friendship and encouragement; in particular, Dr. Priyadarsi Nanda, Prof. Sean He, Eryani Tjondrowalujo, Chang Liu, Xuyun Zhang, Chi Yang, Adrian Johannes, Nazanin Borhan, Ashish Nanda, Jongkil Kim, Nan Li, Danan Thilakanathan, Mian Ahmed Jan, and Usman Khan.

Last but not least, I am deeply grateful to my parents Karitk Ch. Puthal, Shakuntala Puthal, my brother, sisters and brothers-in-law for supporting me to study abroad, understanding, encouragement and help. Most importantly, I would like to sincerely express the deepest gratitude to almighty god.

Abstract

A large number of mission critical applications from disaster management to health monitoring are contributing to the Internet of Things (IoT) by deploying a number of smart sensing devices in a heterogeneous environment. Resource constrained sensing devices are being used widely to build and deploy self-organising wireless sensor networks for a variety of critical applications. Many such devices sense the deployed environment and generate a variety of data and send them to the server for analysis as data streams. The key requirement of such applications is the need for near real-time stream data processing in large scale sensing networks. This trend gives birth to an area called big sensing data streams. One of the key problems in big data is to ensure end-to-end security where a Data Stream Manager (DSM) must always verify the security of the data before executing a query to ensure data security (i.e., confidentiality, integrity, authenticity, availability and freshness) as the medium of communication is untrusted. A malicious adversary may access or tamper with the data in transit. One of the challenging tasks in such applications is to ensure the trustworthiness of collected data so that any decisions are made on the correct data, followed by protecting the data streams from information leakage and unauthorised access. This thesis considers end-to-end means from source sensors to cloud data centre. Although some security issues are not new, the situation is aggravated due to the features of the five Vs of big sensing data streams: Volume, Velocity, Variety, Veracity and Value. Therefore, it is still a significant challenge to achieve data security in big sensing data streams. Providing data security for big sensing data streams in the context of near real time analytics is a challenging problem.

This thesis mainly investigates the problems and security issues of big sensing data streams from the perspectives of efficient and lightweight processing. The big data streams computing advantages including real-time processing in efficient and lightweight fashion are exploited to address the problem, aiming at gaining high scalability and effectiveness. Specifically, the thesis examines three major properties in the lifecycle of security in big data streams environments. The three properties include authenticity, integrity and confidentiality also known as the AIC triad, which is different to CIA triad used in general data security. Accordingly, a lightweight security framework is proposed to maintain data integrity and a selective encryption technique to maintain data confidentiality over big sensing data streams. These solutions provide data security from source sensing devices to the processing layer of cloud data centre. The thesis also explore a further proposal on a lattice based information flow control model to protect data against information leakage and unauthorised access after performing the security verification at DSM. By integrating the access control model, this thesis provides an end-to-end security of big sensing data streams i.e. source sensing device to the cloud data centre processing layer. This thesis demonstrates that our solutions not only strengthen the data security but also significantly improve the performance and efficiency of big sensing data streams compared with existing approaches.

The Author's Publications

So far, I have published nine refereed papers including one book chapter, one IEEE magazine, one ERA ranked A*¹ journal paper, one ERA ranked A journal paper, three ERA ranked A conference papers and one ERA ranked B conference paper and other papers. The publications as well as one paper that is under review are listed below in detail. The impact factor (IF)² of each journal paper is also stated.

Book Chapter:

1. **Deepak Puthal**, Surya Nepal, Rajiv Ranjan, and Jinjun Chen. "End-to- End Security Framework for Big Sensing Data Streams." in *Big Data Management, Architecture, and Processing*, CRC Press, to be published 2017.

Journal Articles:

2. **Deepak Puthal**, Surya Nepal, Rajiv Ranjan, and Jinjun Chen. "A Dynamic Prime Number Based Efficient Security Mechanism for Big Sensing Data Streams." *Journal of Computer and System Sciences (JCSS)*. Vol. 83(1), pp. 22-42, 2017. (A*, IF: 1.583)
3. **Deepak Puthal**, Surya Nepal, Rajiv Ranjan, and Jinjun Chen. "DLSeF: A Dynamic Key Length based Efficient Real-Time Security Verification Model for Big Data Streams." *ACM Transactions on Embedded Computing Systems*

¹ ERA ranking is a ranking framework for publications in Australia. Refer to http://www.arc.gov.au/era/era_2010/archive/era_journal_list.htm for detailed ranking tiers. The 2010 version is used herein. For journal papers: A* (top 5%); A (next 15%). For conference papers (no A* rank): A (top 20%).

² IF: Impact Factor. Refer to <http://wokinfo.com/essays/impact-factor/> for details and query.

(TECS), Vol. 16(2), pp. 51:1-51:24, 2016. (A*, IF: 1.19)

4. **Deepak Puthal**, Surya Nepal, Rajiv Ranjan, and Jinjun Chen. "Threats to Networking Cloud and Edge Datacenters in the Internet of Things." *IEEE Cloud Computing*. Vol. 3(3), pp. 64-71, 2016.
5. **Deepak Puthal**, Surya Nepal, Rajiv Ranjan, Xindong Wu, and Jinjun Chen. "SEEN: A Selective Encryption Method to Ensure Confidentiality for Big Sensing Data Streams." *IEEE Transactions on Big Data (TBD)*, Minor revision, February 2017.

Conference Papers:

6. **Deepak Puthal**, Surya Nepal, Rajiv Ranjan, and Jinjun Chen. "A Synchronized Shared Key Generation Method for Maintaining End-to-End Security of Big Data Streams." in *50th Hawaii International Conference on System Sciences (HICSS-50)*, Hawaii, USA. 2017. (A)
7. **Deepak Puthal**, Surya Nepal, Rajiv Ranjan, and Jinjun Chen. "IoT and Big Data: An architecture With Data Flow and Security Issues." in *2nd international conference on Cloud, Networking for IoT Systems (CN\$IoT)*, Brindisi, Italy, 2017.
8. **Deepak Puthal**, Surya Nepal, Rajiv Ranjan, and Jinjun Chen. "A Secure Big Data Streams Analytics Framework for Disaster Management on Cloud." in *18th IEEE International Conferences on High Performance Computing and Communications (HPCC 2016)*, Sydney, Australia. 2016 (B)
9. **Deepak Puthal**, Surya Nepal, Rajiv Ranjan, and Jinjun Chen. "A Dynamic Key Length based Approach for Real-Time Security Verification of Big Sensing Data Streams." in *16th International Conference on Web Information System Engineering (WISE 2015)*, Miami, Florida, USA. 2015. (A)
10. **Deepak Puthal**, Surya Nepal, Rajiv Ranjan, and Jinjun Chen. "DPBSV – An Efficient and Secure Scheme for Big Sensing Data Streams." in *14th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (IEEE TrustCom-15)*, Helsinki, Finland. 2015. (A)

11. **Deepak Puthal**, Surya Nepal, Cecile Paris, Rajiv Ranjan, and Jinjun Chen.
"Efficient Algorithms for Social Networks Coverage and Reach." in *IEEE BigData Congress*, New York, USA, 2015.

Table of Contents

Figures **xiii**

Tables **xv**

Algorithms **xvi**

Chapter 1	Introduction	1
1.1	Background	1
1.1.1	Big Data with Security Issues	3
1.1.2	Cloud Computing	5
1.2	Motivation: Securing Big Sensing Data Streams	6
1.3	Overview of the Work	9
1.3.1	Methodology	9
1.3.2	Contributions	11
1.4	Thesis Organisation	13
Chapter 2	Background Studies and Related Work	15
2.1	General Research Trend	15
2.2	Review of Reviews	17
2.2.1	Data Centre Security	17
2.2.2	Network Security	19
2.2.3	IoT Security	21
2.3	IoT Generated Data Stream Architecture	23
2.3.1	IoT Architecture	23
2.3.2	Security Threats of Each Layer	28
2.4	Big Data Stream Security	38
2.4.1	Security Requirements	40
2.4.2	CIA Triad Properties	41

2.4.3	Confidentiality of Big Data Streams	42
2.4.4	Integrity of Big Data Streams	45
2.4.5	Availability of Big Data Streams	51
2.5	Comparison	56
2.6	Summary	59
Chapter 3 Security Verification Framework for Big Sensing Data Streams		61
3.1	Introduction	62
3.2	Preliminaries to the Chapter	64
3.3	Research Challenges and Research Motivation	65
3.3.1	Research Challenges	66
3.3.2	Research Motivation	67
3.4	Dynamic Prime-Number Based Security Verification	70
3.4.1	DPBSV System Setup	70
3.4.2	DPBSV Handshaking	72
3.4.3	DPBSV Rekeying	72
3.4.4	DPBSV Security Verification	74
3.5	Security Analysis	76
3.5.1	Security Proof	76
3.5.2	Forward Secrecy	81
3.6	Experiment and Evaluation	81
3.6.1	Sensor Node Performance	82
3.6.2	Security Verification	83
3.6.3	Performance Comparison	86
3.6.4	Required Buffer Size	87
3.7	Summary	88
Chapter 4 Lightweight Security Protocol for Big Sensing Data streams		89
4.1	Introduction	89
4.2	Preliminaries to the Chapter	92
4.3	Research Challenges and Research Motivation	94
4.3.1	Research Challenges	94
4.3.2	Research Motivation	96
4.4	DLSeF Lightweight Security Protocol	96
4.4.1	DLSeF System Setup	97
4.4.2	DLSeF Handshaking	100

4.4.3	DLSeF Rekeying	101
4.4.4	DLSeF Key Synchronisation	104
4.4.5	DLSeF Security Verification	108
4.5	Security Analysis	110
4.5.1	Security Proof	111
4.6	Experiment and Evaluation	115
4.6.1	Sensor Node Performance	115
4.6.2	Security Verification	117
4.6.3	Performance Comparison	120
4.6.4	Required Buffer Size	121
4.7	Summary	123
Chapter 5 Seletive Encryption Method to ensure Confidentiality of Big Sensing Data Streams		124
5.1	Introduction	125
5.2	Design Consideration	127
5.2.1	System Architecture	128
5.2.2	Adversary Model	130
5.2.3	Attack Model	131
5.3	Research Challenges and Research Motivation	132
5.3.1	Research Challenges	132
5.3.2	Research Motivation	134
5.4	Selective Encryption Method for Big Data Streams	135
5.4.1	Initial System Setup	136
5.4.2	Rekeying	138
5.4.3	New Node Authentication	139
5.4.4	Reconfiguration	141
5.4.5	Encyption/Decryption	142
5.4.6	Tradeoffs	143
5.4.7	Requirement Resources for SEEN	144
5.5	Theoretical Analysis	147
5.5.1	Security Proof	147
5.5.2	Forward Secrecy	150
5.6	Experimental and Evaluation	150
5.6.1	Security Verification	151

5.6.2	Performance Comparison	153
5.6.3	Required Buffer Size	154
5.6.4	Network Performance	155
5.7	Summary	157
Chapter 6	Access Control Framework for Big Sensing Data streams	158
6.1	Introduction	158
6.2	Background Studies	161
6.2.1	Stream Processing	161
6.2.2	Stream Security	162
6.2.3	Chinese Wall Policy	163
6.3	Design Consideration	163
6.3.1	System Architecture	163
6.3.2	Defination	166
6.3.3	QoS Requirements	167
6.3.4	Adversary Model	169
6.4	Access Control Model	170
6.5	Experimental Evaluation	173
6.5.1	System Setup	173
6.5.2	Results Discussion	175
6.6	Summary	176
Chapter 7	Conclusion and Future Work	177
7.1	Conclusion	177
7.2	Future Work	181
Bibliography		183

Figures

Figure 1-1 Typical Lifecycle of Security Framework for Big Sensing Data Streams	6
Figure 2-1 Cloud computing security architecture	19
Figure 2-2 Layer wise IoT Security architecture	22
Figure 2-3 layer wise IoT architecture from IoT device to cloud data centre	26
Figure 2-4 Communication protocol in IoT	28
Figure 2-5 Cloud computing security threats, attacks and vulnerabilities	38
Figure 2-6 CIA triad of data security either data in transit or in rest	41
Figure 3-1 A simplified view of a DSMS to process and analyse input data stream	62
Figure 3-2 Overlay of our architecture from sensing device to data centre	65
Figure 3-3 Pair of dynamic relative prime number generation	68
Figure 3-4 The sensors used for experiment	81
Figure 3-5 Estimated power consumption during the key generation process	83
Figure 3-6 Scyther simulation environment result page	84
Figure 3-7 Performance of the security scheme comparison	85
Figure 3-8 Performance comparison of minimum buffer size required	87
Figure 4-1 High level of architecture from source sensing device to big data processing centre	93
Figure 4-2 Secure authentication of Sensor and DSM	100
Figure 4-3 Neighbour node discovered to get the key generation properties	105
Figure 4-4 Neighbour discovery with all possible conditions	107
Figure 4-5 Performance computation of two different sensors	116
Figure 4-6 Energy consumption by using COOJA in Contiki OS	116
Figure 4-7 Scyther simulation environment result page	118

Figure 4-8 Security verification results of Scyther during neighbour authentication	119
Figure 4-9 Performance comparison.....	121
Figure 4-11 Efficiency comparison of minimum buffer size required to process ..	121
Figure 5-1 High level architectural diagram for SEEN protocol	130
Figure 5-2 Initial authentication methods with 4 steps process	138
Figure 5-3 Key Selection	139
Figure 5-4 Shared key management for robust clock skew	140
Figure 5-5 Method to the data sensitivity level	141
Figure 5-6 Scyther simulation result page of security verification	152
Figure 5-7 Performance comparison SEEN method	153
Figure 5-8 Efficiency comparison by comparing required buffer size	154
Figure 5-9 Energy consumption	155
Figure 6-1 Overview of access control of big data streams using lattice model ...	166
Figure 6-2 Lattice model for data access	171
Figure 6-3 Experiment Setups	172
Figure 6-4 Mapping time for HT Sensor Dataset	173
Figure 6-5 Mapping time for Twin Gas Sensor Dataset	174

Tables

Table 2-1 Network layer security threats	31
Table 2-2 Possible threats of IoT generated Big Dat streams in CIA triad representation.....	57
Table 2-3 Comparison of IoT generated big data stream security threats and solutions according to CIA triad method	58
Table 3-1 DPBSV Notations.....	69
Table 3-2 Notations Symmetric key (AES) algorithm takes time to get all possible keys using most advanced Intel i7 Processor	77
Table 4-1 Notations used in this DLSeF model	98
Table 5-1 SEEN Notations	135
Table 5-2 Performance and Properties of Security Solutions	156
Table 5-3 Communication overhead of SEEN protocol	156
Table 6-1 Machine specification	174
Table 6-2 Dataset information	174

Algorithms

Algorithm 3-1 Security Framework for Big Sensing Data Stream	74
Algorithm 3-2 Dynamic Prime Number Generation	78
Algorithm 4-1 Synchronisation of Dynamic Key Length Generation	102
Algorithm 4-2 Key Generation (Rekeying) Process.....	107
Algorithm 4-3 Lightweight Security Protocol for Big Sensing Data Stream	109
Algorithm 5-1 Rekeying process	140
Algorithm 5-2 Selective encryption method for big sensor data streams	145