Algorithms for Big Data in Advanced Communication Systems and Cloud Computing

Christos Stergiou, Student Member, IEEE, and Kostas E. Psannis, Member, IEEE

Abstract—Communication Systems are becoming significant in many areas of modern everyday life. Through this field several technologies grow and contribute to the improvement of people's everyday life. Big Data (BD) appears as a technology created and developed through communication systems. Big Data refers to the large-scale amounts of data used, transferred and managed through a network. Due to data usage and even large-scale amounts of data, the use of storage space without restrictions on its use becomes necessary. This storage space provided through a technology called Cloud Computing (CC). This technology mentions to a substructure in which data storage and data processing occur in real time outside of the user's device. This work surveys BD and CC and their basic features, with a focus on the security and privacy issues of both technologies. In addition to this, we will try to combine the functionality of BD and CC with the aim to examine the frequent features, and also to discover the benefits related in security issues of their integration.

Index Terms—Cloud Computing, Big Data, Algorithms, Big Data Applications, Cloud Computing Security.

I. CONTEXT & MOTIVATION

Big Data is a novel popular term, used to describe the surprisingly rapid raise in volume of data in structured and unstructured form. It is a broad term for data sets so large or complex that traditional data processing applications are inadequate. Rarely, it also adverts to a particular size of data set. Accuracy in big data may lead to more confident decision making, and better decisions can result in greater operational efficiency, cost reduction, and cut-rate risk [1]. From this scope we realize that the Big Data is now equally important both for business and internet. This happens because more information leads to more accurate analyses [2] [3].

A new generation of services, based on the sense of the "cloud computing", has made its appearance in the last few years with the scope of providing access to the information and the data from any place at any time, thus restricting or expunging the need for hardware equipment. The term "cloud computation" is defined as the use of computing logistical resources, as well as the software level, through the use of

Copyright (c) 2011 IEEE. Personal use of this material is permitted.

C. Stergiou is with the Department of Applied Informatics, University of Macedonia, Greece (e-mail: tm1163@uom.edu.gr).

K. E. Psannis is with the Department of Applied Informatics, University of Macedonia, Greece (corresponding author, phone: +302310891737; e-mail: kpsannis@uom.edu.gr).

services transported over the internet. Nowadays, Cloud Computing services compose one of the world's largest areas of competition between giant companies in the IT sector and software, such as Google, Amazon and Microsoft, which are struggling to take an advantageous position, to this rapidly growing industry [4] [5] [6].

Also, a few instances contain the restrictions of communication capabilities, processing, energy and storage. Such inadequacies exhort us to fuse the technology of CC and Big Data. As an upcoming technology, CC integrates a large number of technologies with the aim to maximize capacity and performance of the current infrastructure [7] [8] [9] [10].

II. BACKGROUND REVIEW

A. Big Data

BD is a more complicated world because the scale is much larger. The information is usually spread out over a number of servers, and the work of compiling the data must be coordinated among them. In the past, the work was largely delegated to the database software, which would use its magical JOIN [11] mechanism to compile tables, after add up the columns before handing off the rectangle of data to the reporting software that would paginate it. This was often harder than it sounds. Database programmers can tell you the stories about complicated JOIN commands that would lock up their database for hours as it tried to produce a report for the boss who wanted his columns just so [8] [12].

B. Big Data Characteristics

The three Vs of big data (figure 1), which are volume, variety, and velocity, constitute a comprehensive definition, and they bust the myth that big data is only about data volume.

Big Data Volume

The Data volume measures the amount of data available to an organization, which does not necessarily have to own all of it as long as it can access it [8] [12] [13].

Big Data Velocity

The Data velocity measures the speed of data creation, streaming, and aggregation. Data velocity management is much more than a bandwidth issue; it is also an ingest issue (extract-transform-load) [8] [12] [13].



Big Data Variety

The Data variety is a measure of the richness of the data representation – text, images video, audio, etc. Incompatible data formats, non-aligned data structures, and inconsistent data semantics represents significant challenges that can lead to analytic sprawl [8] [12] [13].

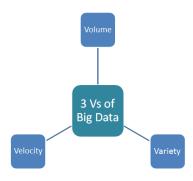


Figure 1: The Three Vs of Big Data.

C. Big Data Security & Privacy

New challenges and standards developed and created in data security issues through the development and the use of BD technology. This creates a growing need for further research on security technologies in order to be able to handle the huge amount of data and to ensure effective. Technologies for securing data are slow when applied to huge amounts of data.

D. Cloud Computing

Cloud computing provides computing, storage, services, and applications over the Internet (figure 2). In general, to render smartphones energy efficient and computationally capable, major changes to the hardware and software level required. This entails the cooperation of developers and manufacturers. [14] [15]. The technology of Cloud computing is the outcome of interdisciplinary approaches combining mobile computing with cloud computing. Thus, this transdisciplinary domain is also referred as mobile cloud computing [8] [14] [16] [17].



Figure 2: Cloud Computing.

E. Cloud Computing Features

As all technologies, so the CC technology has some

characteristics which determine its function. These features are analyzed and outlined subsequently.

Storage over Internet

Storage over Internet can be defined as a technology framework that uses Transmission Control Protocol/Internet Protocol (TCP/IP) networks to link servers and storage devices, and to facilitate storage solution deployment. The Storage over Internet technology is also known as Storage over Internet Protocol (SoIP) technology [18] [19] [20].

Service over Internet

The main objective of the Service over Internet is to be committed to help customers all over the world with the aim to transform aspirations into achievements by harnessing the Internet's efficiency, speed and ubiquity [18] [19] [21].

Applications over Internet

The programs which can be written to do the job of a current manual task, or virtually anything, and which perform their job on the server (cloud server) via an internet connection rather than the traditional model of a program that has to be installed and run on a local computer are the Cloud Applications, or as a scientific definition Applications over Internet [18] [19] [20].

Energy Efficiency

As a definition, the Energy Efficiency is a way of managing and restraining the growth in energy consumption. By delivering more services for the same energy input or for the same services for less energy input may be something more energy efficient [18] [19] [20].

Computationally Capable

The services of computational clouds are leveraging the computationally intensive and ubiquitous mobile applications which have been enabled by the technology of Mobile Cloud Computing [6] [17] [18] [19].

F. Cloud Computing Security Issues

CC and storage solutions provide users and enterprises with various capabilities to store and process their data in thirdparty data centers [22] [23]. Organizations use the Cloud in a variety of different service models (SaaS, PaaS, and IaaS) and deployment models (Private, Public, Hybrid, and Community) [15] [24] [25]. There are a number of security concerns associated with cloud computing. These issues fall into two broad categories: security issues faced by cloud providers (organizations providing software, platform, or infrastructureas-a-service via the cloud) and security issues faced by their customers (companies or organizations who host applications or store data on the cloud) [26] [27]. The responsibility is shared, however. The provider must ensure that their infrastructure is secure and that their clients' data and applications are protected while the user must take measures to fortify their application and use strong passwords and

authentication measures [8] [9].

III. RESEARCH OBJECTIVES

There is a survey of BD and CC technologies and their basic characteristics, with a focus on the security and privacy issues of both technologies. Moreover, I will try to combine the functionality of the two aforementioned technologies (i.e Big Data and Cloud Computing) in order to examine the common features, and also to discover the benefits related in security issues of their integration. The main goal of this paper is to try to combine the functionality of the BD and CC technologies in order to examine the common features, and also to discover the benefits related in security issues of their integration. This could be take place by the presentation of a new method of an algorithm that can be used for the purpose of improving Cloud Computing's security through the use of algorithms that can provide more privacy in the data related to Big Data technology. Furthermore, we survey the security challenges of the integration of those technologies. This can be the field of future research on the integration of those two technologies, and why not to have a huge improvement of their security and privacy issues in order to have a better use of them.

IV. RESEARCH APPROACH & METHODS

The purpose of this research proposal is to study the technologies for the "large-scale data" (big data) and "cloud computing" in order to analyze and manage the telecommunication systems.

This study will rely on the following items in order to examine all the existing and the future data in order to gather and produce all the necessary tools with the aim to come out:

- 1. Action Research
- 2. Development and implementation of the Unified Modeling Language (UML)
- 3. Software engineering
- 4. Implementation and model development
- 5. Usability evaluation

According to the aforementioned data, the survey will rely on the study of existing literature and the use of existing research on the fields I will study, in order to use all the necessary information. The next stage will be the recovery process of all the material will be collected and organized in the different steps. Based on the current situation regarding the scope of my research will develop simulation models in order to study and achieve my proposal. Finally, there will be a verification of the data that will be produced during the implementation of the model and the usability of the proposed model will be studied as well as the further improvement and evaluation.

V. CURRENT & EXPECTED CONTRIBUTIONS

As already mentioned, with this work I will try to find a

better security algorithm model for the BD in Cloud environments. For this purpose several existing security algorithms would be studied.

Table 1 lists a sample of the most popular security algorithms that would be studied in order to produce a new algorithm model. As regards the Table 1 we could conclude that even the most efficient algorithms give an encryption rate of 64.3MB/s. The information of those algorithms have been taken through related works [8] [28] [29] [30] [31].

Algorithm	Key length	Megabytes processed	Block size	Rounds
Blowfish	32-448 bits	256	64 bits	16
DES	56 bits	128	64 bits	16
3-DES	56, 112 or 168 bits	128	64 bits	48
AES	128, 192 or 256 bits	256	128 bits	10, 12 or 14
RSA	1025 – 4096 bits	300	512 bits	1

Table 1: Encryption Rates of popular Security Algorithms.

Through the table 1 we can conclude that in the sector of BD technology, in which the need of large amounts of data needs to be transferred we could observe an important bottle neck for encryption like huge amounts of data. This is harmful to the nature of BD that have real time processing and outcomes.

Subsequently, a correlation of characteristics of BD and CC can be made.

Big Data Features	Volume	Velocity	Variety
Cloud Computing Features			
Storage over Internet		X	
Service over Internet	X		X
Applications over Internet	X	X	X
Energy Efficiency	X	X	
Computational Capable		X	X

Table 2: Correlation of BD and CC characteristics.

By the table 2 can be exhibited the key characteristics of the two technologies which have been studied and used for the experimental proposal. Count on the study conducted, the key feature of BD technology which contributes more with the characteristics of CC technology is Velocity. Velocity contributes four from the five key characteristics of CC. Also, another thing that we can observe from table 2 is that the characteristic Applications over Internet contributed from all the key features of BD.

Moreover, a big part of this work would relay on the related works that have been made previously. Table 3 lists the findings and the concepts associated with problems and respective solutions indicating in a previous works. In table 3 the former work list in ascending chronological order starting from 2010 until today.

Author	Problems	Solutions	
H. Takabi et al [32]	• Specific characteristics worsen security & privacy challenges of Cloud Computing.	• Examines the possibilities of offering a trustworthy CC environment.	
H. T. Dinh et al [33]	 Detonating growth of mobile applications & resurgent of CC concept is considered advancement in mobile services. 	A survey of MCC, with focus on its definition, architecture & applications.	
N. Fernando et al [34]	• Intrinsic problems (e.g. resource scarcity, frequent disconnections) hinter the usage of mobile computing in its full scale.	 Categorizes the major issues in MCC & discusses different methods to solve these issues. Careful examination of problems which have not yet been addressed & put forward ideas for future research. 	
Sachdev & M. Bhansali [35]	The bigger the number of cloud users the most frequent the malicious activity in the cloud.Highly safe and persistent services needed.	A data encryption model which protects the privacy and security of the data before they are uploaded in the cloud.	
M. Ali et al [36]	 Third-party cloud services have more deficiencies and more vulnerable to security threads. Sharing the users' data outside the administrative control. 	Examines and shortly analyzes both internal and external security problems in the Mobile Cloud Computing.	
S. Bhavani et al [37]	 Load balancing is one of the cloud's issues. A reduction in the response time and optimization of the resource utilization can be achieved balancing the load. 	The best algorithm for balancing the load is Ant Colony Optimization.	
S. Sathya & R. Avinash [38]	How people adopt cloud as Cloud Technologies Mature.	An explanation of how BD and cloud responds for user's demand as a compelling combination.	
S. Rallapalli et al [39]	 The healthcare organizations face the critical challenge to analyze big data. Large amounts of data cannot be processed through conventional systems. 	Hadoop: An application which could prepare huge amounts of data in distributed environment could be deployed on cloud environment to prepare the big amount of healthcare data.	
O. Awodele et al [40]	 Security challenges are the most serious in cloud & big data services. Issues of service level agreement. 	 Shipping disk drives to cloud computing. Use of Data mining techniques. Use of Access control techniques. 	
N. R. Vajjhala & E. Ramollari [41]	 Contemporary methods in the field of BD using cloud resources. How the SMEs can take advantage of these technological trends. 	Cloud computing offers an alternative to SMEs shifting the burden of providing and maintaining expensive infrastructure to cloud service providers.	
P. Zhou et al [42]	 The increased usage of social media has created a new period, that of the BD. Privacy of users' contexts & video service sellers' repositories, that are remarkably sensitive & of important commercial value. 	An innovative "geometric differentially private" scheme, that could minimize the performance loss.	
A. A. Gnana Singh et al [43]	• Promote the research and development activities in the field of BD and CC.	• A method for storing the data on cloud using the cloudsim package.	

Table 3: Mapping problems against referenced solutions.

VI. CONCLUSION & NEXT STEPS

The CC technology provides many possibilities, but in addition to this places quite a lot of restrictions as well. This technology mentions to an infrastructure where both the data storage and processing occur outside of the user's device. In this work, we survey BD and CC technology and their basic characteristics, with a focus on the security and privacy issues of both technologies. Moreover, we have tried to combine the functionality of the two aforementioned technologies (i.e BD & CC) with the aim to examine the frequent characteristics, and moreover to discover the benefits related in security issues of their integration.

The main goal of this work is to find novel ways to achieve a better integration of BD and CC, with focus on security algorithms and all the challenges that the two aforementioned technologies faced on security level. This can be the field of future research on the integration of those two technologies. Regarding the rapid development of both technologies the security issue must be solved or reduced to a minimum in order to have a better integration model. These security challenges that surveyed in this paper could be the sector for further research as a case study, with the goal of minimizing them.

APPENDIX

Doctoral's advisor recommendation letter:

Dear Selection Committee,

I am writing this letter to strongly recommend Mr. Stergiou for your conference. I know Mr. Stergiou because we had cooperated in his dissertation for his Bachelor degree in Technology Management, University of Macedonia. I met with him and outlined a project. I gave him some background reading at our first meeting. By the time of our second meeting he had read what I had given him and prepared a two-page project description.

During his time in School of Technology Management, University of Macedonia, Mr. Stergiou demonstrated a good work ethic and interpersonal skills. We outlined a scope of work to be completed, and he successfully completed that work in the time required. He put in extra hours as necessary in order to meet specific deadlines that I set. I teamed him up with another student to work on the project. He seemed to work well with the other student, and I found him very personable. Mr. Stergiou put in sufficient work to be a co-author on a manuscript.

In summary, Mr. Stergiou is clearly the best student I have worked with in the last 10 years. I would very much like him match to our residency program. Even though I hope he stays here, I think he would be an outstanding asset to your conference. I give him my highest recommendation.

Sincerely,

Dr. Konstantinos E. Psannis



ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and feedback that was extremely helpful in improving the quality of the paper.

REFERENCES

- Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. Science, 332(6025), 60 –65. doi:10.1126/science.1200970.
- [2] C. Stergiou & K. E. Psannis, "Recent advances delivered by Mobile Cloud Computing and Internet of Things for Big Data applications: a survey," International Journal of Network Management, pp. 1-12, 11/3/2016.
- [3] Andreas P. Plageras et al, "IoT-based Surveillance System for Ubiquitous Healthcare," in *Industrial Electronics Society*, *IECON 2016* - 42nd Annual Conference of the IEEE, 22/12/2016
- [4] "The NIST Definition of Cloud Computing", National Institute of Standards and Technology, Retrieved 24 July 2011.
- [5] G. Skourletopoulos, C. X. Mavromoustakis, G. Mastorakis, J. Mongay Batalla and J. N. Sahalos, "An Evaluation of Cloud-Based Mobile Services with Limited Capacity: A Linear Approach". Soft Computing journal (2016). DOI: 10.1007/s00500-016-2083-4
- [6] C. Stergiou, K. E. Psannis, "Mobile Cloud Computing in 4G Networks (LTE)", (2015), 2nd Student Conference of Applied Informatics, 2/12/2015, Thessaloniki, Greece.
- P. Wayner, "7 top tools for taming big data," InfoWorld, 18/4/2012.
 [Online]. Available: http://www.infoworld.com/article/2616959/big-data/7-top-tools-for-taming-big-data.html. [Accessed 21/5/2016].
- [8] Stergiou Christos, Kostas E. Psannis, "Efficient and Secure Big Data delivery in Cloud Computing", Date Submitted: 23/12/2016, In Press.
- [9] D. Tomtsis, S. Kontogiannis, G. Kokkonis, I. Kazanidis, S. Valsamidis, "Proposed cloud infrastructure of wearable and ubiquitous medical services", 5th Int. Conf. on Digital Information Processing and Communications (ICDIPC 2015), pp. 213-218, Switzerland, Oct. 2015.
- [10] Vajjhala NR, Ramollari E (2016) Big data using cloud computing opportunities for small and mediumsized enterprises. Eur J Econ Bus Stud 1(4):129–137.
- [11] Y. Kryftis, G. Mastorakis, C. Mavromoustakis, J. Mongay Batalla, E. Pallis and G. Kormentzas, "Efficient Entertainment Services Provision over a Novel Network Architecture". To be published in IEEE Wireless Communications Magazine, 2016.
- [12] C. Buckler, "Understanding JOINs in MySQL and Other Relational Databases," sitepoint, 19/5/2011. [Online]. Available: https://www.sitepoint.com/understanding-sql-joins-mysql-database/. [Accessed 21/5/2016].
- [13] P. Russom, Big Data Analytics, USA: TDWI The Data Warehousing Institute, 2011.
- [14] D. Huang, "Mobile cloud computing," IEEE COMSOC Multimedia Communications Technical Committee (MMTC) E-Letter, vol. 6, no. 10, pp. 27–31, 2011.
- [15] Christos Stergiou, Kostas E. Psannis, Byung-Gyu Kim, Brij Gupta, "Secure integration of IoT and Cloud Computing", Elsevier, Future Generation Computer Systems, December 2016.
- [16] Mohammad A. Alsmirat; Yaser Jararweh; Islam Obidat; Brij B. Gupta, "Internet of Surveillance: A Cloud supported Large Scale Wireless Surveillance System," the Journal of Supercomputing, Springer.

- [17] Christos Stergiou, «Technologies of Internet of Things and Mobile Cloud Computing», Bachelor Dissertation, Technology Management, Information Technology, University of Macedonia, June 2016.
- [18] G. Md Whaiduzzaman et al, "A Study on Strategic Provision of Cloud Computing Services", The Scientific World Journal, pp. 1-8, 15/6/2014.
- [19] Garg SK, Versteeg S, Buyya R. "A framework for ranking of cloud computing services". Future Generation Computer Systems. 2013;29(4):1012–1023.
- [20] Georgios Skourletopoulos et al, "An evaluation of cloud-based mobile services with limited capacity: a linear approach," Soft Computing, pp. 1-8, 27/2/2016.
- [21] L. Villars et al, "The Critical Role of the Network in Big Data Applications", IDC Analyze th Future, pp. 1-12, 1/4/2012.
- [22] Mohammad Haghighat et al, "CloudID: Trustworthy cloud-based and cross-enterprise biometric identification," Expert Systems with Applications, vol. 11, no. 42, pp. 7905-7916, 30/11/2015.
- [23] Kriti Bhushan, B. B. Gupta, "Security Challenges in Cloud Computing: State-of-art," International Journal of Big Data Intelligence (IJBDI), Inderscience, 2015.
- [24] Madhan Kumar Srinivasan et al, "State-of-the-art cloud computing security taxonomies: a classification of security challenges in the present cloud computing environment," ICACCI '12 Proceedings of the International Conference on Advances in Computing, Communications and Informatics, pp. 470-476, 03/08/2012.
- [25] Gou Zhaolong, Shingo Yamaguchi, B. B. Gupta, "Analysis of Various Security Issues and Challenges in Cloud Computing Environment: A Survey," Handbook of Research on Modern Cryptographic Solutions for Computer and Cyber Security, IGI-Global's Advances in Information Security, Privacy, and Ethics (AISPE) series, USA, 2016.
- [26] Y. Mamoon, ""Swamp Computing" a.k.a. Cloud Computing," WEB Security Journal, 28/12/2009. [Online]. Available: http://security.sys-con.com/node/1231725. [Accessed 27/07/2016].
- [27] Rizwana Shaikha & Dr. M. Sasikumar, "Data Classification for achieving Security in cloud computing," Procedia Computer Science, no. 45, p. 493 – 498, 1/3/2015.
- [28] L. Villars et al, "The Critical Role of the Network in Big Data Applications", IDC Analyze the Future, pp. 1-12, 1/4/2012.
- [29] G. Md Whaiduzzaman et al, "A Study on Strategic Provision of Cloud Computing Services", The Scientific World Journal, pp. 1-8, 15/6/2014.
- [30] Garg SK, Versteeg S, Buyya R. "A framework for ranking of cloud computing services". Future Generation Computer Systems. 2013;29(4):1012–1023.
- [31] Badve OP, Gupta BB, Gupta S (2016) Reviewing the security features in contemporary security policies and models for multiple platforms, handbook of research on modern cryptographic solutions for computer and cyber security. IGI-Global's advances in information security, privacy, and ethics (AISPE) series, USA.
- [32] H. Takabi et al, "Security and Privacy Challenges in Cloud Computing Environments," IEEE Computer and Reliability Societies, pp. 24-31, 1/11/2010.
- [33] H. T. Dinh et al, "A survey of mobile cloud computing: architecture, applications, and approaches," Wireless Communications and Mobile Computing, no. 13, pp. 1587-1611, 11/10/2011.
- [34] N. Fernando et al, "Mobile cloud computing: A survey," Future Generation Computer Systems, no. 29, pp. 84-106, 6/6/2012.
- [35] A. Sachdev & M. Bhansali, "Enhancing Cloud Computing Security using AES Algorithm," International Journal of Computer Applications, vol. 9, no. 67, pp. 19-23, 1/4/2013.
- [36] M. Ali et al, "Security in cloud computing: Opportunities and challenges," Information Sciences, no. 305, pp. 357-383, 7/2/2015.
- [37] S. Bhavani et al, "Study on Cloud Computing and Different Load Balancing Algorithms in Cloud Computing," International Journal of Emerging Research in Management & Technology, vol. 5, no. 4, pp. 331-336, 1/5/2015.
- [38] S. Sathya & R. Avinash, "Big Data and Cloud Computing," in Rathinam College National Conference, Echanari, 2015.
- [39] S. Rallapallia et al, "Impact of Processing and Analyzing Healthcare Big Data on Cloud Computing Environment by Implementing Hadoop Cluster," International Conference on Computational Modeling and Security (CMS2016), pp. 16-22, 1/12/2015.

- [40] O. Awodele et al, "Big Data and Cloud Computing Issues," International Journal of Computer Applications, vol. 12, no. 133, pp. 14-19, 1/1/2016.
- [41] N. R. Vajjhala & E. Ramollari, "Big Data using Cloud Computing -Opportunities for Small and Medium-sized Enterprises," European Journal of Economics and Business Studies, vol. 1, no. 4, pp. 129-137, 1/4/2016.
- [42] P. Zhou et al, "Differentially Private Online Learning for Cloud-Based Video Recommendation With Multimedia Big Data in Social Networks," IEEE Transactions on Multimedia, vol. 6, no. 18, pp. 1217-1229, 1/6/2016.
- [43] R. Iqbal et al, "Big Data analytics: Computational intelligence techniques and application areas," International Journal of Information Management, pp. 1-11, 1/6/2016.