

A Review on Data Security in Cloud Computing with Knowledge Discovery

Mr. Ankush Chopra
M.Tech.(Computer Science)
Kurukshetra University,
Geeta Engineering College,
Panipat, Haryana, India

Ms. Komal
Assistant Professor
Computer Science & Engineering,
Geeta University, Panipat,
Haryana, India

Mr. Kapil Saini
Assistant Professor
Computer Science & Engineering,
Geeta University, Panipat,
Haryana, India

Abstract— Cloud-computing, sometimes referred to as Cloud or SaaS (Software-as-a-Service), is one of the newest trends in the IT sector. Computing is evolving towards a paradigm that consists of commodities that are distributed similarly to utilities like water, power, gas, and telecommunications. This way, independent of the location of the hosting infrastructure, users obtain services according to their needs. As compared to dedicated infrastructures, it offers complete scalability, stability, great performance, and a relatively affordable option. Cloud has numerous advantages but there is a high gamble of information getting secret data getting spilled. To profit the advantages of cloud, the security of information being moved between the client and client should be secured. Security is the key for the Cloud achievement, security in the cloud is presently the principal challenge of distributed computing. Until a couple of years prior all the business cycles of associations were on their confidential framework and, however it was feasible to re-appropriate administrations, it was typically non-basic information/applications on confidential foundations. Presently with distributed computing, the story has changed. The conventional organization border is broken, and associations feel they have let completely go over their information. New assault vectors have showed up, and the advantage of being open from anyplace turns into a major danger. The Blood Bank, Indian Railway, Gmail Services, and Public Data Verifier are some of the primary generic areas of interest at work that are being improved by current cloud services. [1]

Keywords: *Cloud computing, data security, knowledge discovery, public data verifier*

I. INTRODUCTION

1.1. C. Gentry et al. [19] Cloud computing is probably the most recent advancement in the IT business otherwise called on-request processing. Processing is being changed into a model comprising of administrations that are commoditized and conveyed in a way like utilities like water, power, gas, and communication. In such a model, clients access administrations in light of their prerequisites, paying little mind to where the administrations are facilitated. It gives the full versatility, unwavering quality, elite execution and moderately minimal expense possible arrangement when contrasted with devoted frameworks. It is the application given as administration over the web and framework equipment in the server farms that gives these administrations. Distributed computing is the latest arising worldview

promising to turn the vision of "figuring utilities" into a reality. Distributed computing is join no coherent progression that spotlights on the manner in which we configuration registering frameworks, foster applications, and influence existing administrations for building programming. Whenever you store your information some data computerized or e-information like photographs online rather than on your home PC, or use webmail or a long range interpersonal communication website, you are utilizing a "distributed computing" administration. Assuming you are an association, and you need to use, for instance, a web based invoicing administration as opposed to refreshing the in-house one you have been utilizing for a long time, that internet invoicing administration is a "distributed computing" administration. Distributed computing alludes to the conveyance of registering assets over the Internet.

1.2. Kuyoro S. O., Ibikunle F. & Awodele O et al. [4] Rather than keeping information on your own hard drive or refreshing applications for your requirements, you utilize an assistance over the Internet, at another area, to store your data or utilize its applications. So, distributed computing takes into account the sharing and adaptable sending of administrations, depending on the situation, from practically any area, and for which the client can be charged in view of real use. It depends on the idea of dynamic provisioning, which is applied not exclusively to administrations yet additionally to process ability, capacity, systems administration, and data innovation (IT) infra-structure overall. Assets are made accessible through the Internet and presented on a compensation for each utilization premise from distributed computing sellers.

1.3. C. Wang et al. [6] Cloud computing was authored for what happens when applications and administrations are moved into the web "Cloud." Cloud processing isn't something that abruptly showed up for the time being; in some structure it might follow back to when PC frameworks remotely time-shared registering assets and applications. All the more as of now however, distributed computing alludes to the a wide range of kinds of administrations and applications being conveyed in the web cloud, and the way that, generally speaking, the gadgets used to get to these administrations and applications require no extraordinary applications .Cloud

Computing is a developmental stage, has been filled in as a cutting edge foundation of the business. It is a model which empowers wide organization access, asset pooling, and fast versatility. With the rising interest of safety the servers are not adequately secure to satisfy client's need. Consequently the cloud stage is planned in such a way so it meets every one of the prerequisites of the client. According to the definition gave by the National Institute to Standards and Technology (NIST) "Distributed computing is a model for empowering helpful, on-request network admittance to a common pool of configurable figuring assets (e.g., networks, servers, capacity, applications, and administrations) that can be quickly provisioned and delivered with insignificant administration exertion or specialist co-operation".

2. CHARACTERISTICS OF CLOUD

The various aspects of Cloud Computing characteristics include services models, deployment models, benefits, and challenges. [11]

The characteristics of cloud computing include on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service. On-demand self-service means that customers (usually organizations) can request and manage their own computing resources. Broad network access allows services to be offered over the Internet or private networks. A simple meaning of Pooled resources a user draw from a pool of computing resources, usually in remote data centers. Services can be make for larger or smaller depends of user requirement and use of a service is measured and customers are billed accordingly. [12]

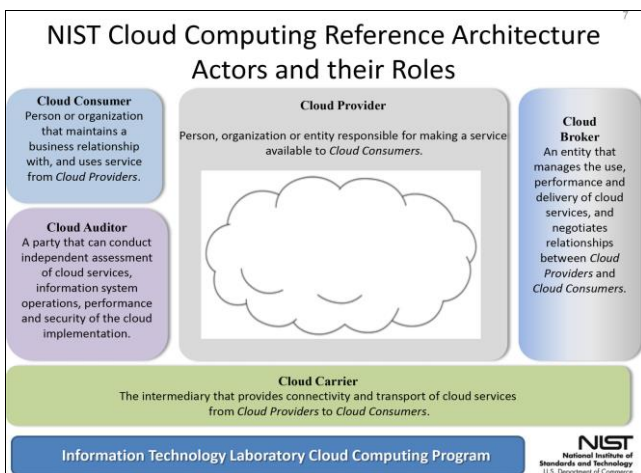


FIGURE 1.1-CLOUD ARCHITECTURE

3. LITERATURE REVIEW

3.1. Susanta Nanda et al. [1] During a literature survey we collect some of information about the blood bank management system located in city and rural area, some of the hospital have its own blood bank unit with each and all technical facilities in city but this conduction is poor in rural area.

3.2. R. J. Adair et al. [2] Some of the country maintain an online blood bank system like in this project have

combination of three sub modules which is blood module, patient module, donor module. In this project blood bank staff has authorized access permission to maintain the all module. Blood module can manage the types, quantity and expiry dates for each category of blood that stored in blood transfusion unit.

3.3. Kuyoro S. O et al. [4] India total blood collection is 7.5 million units yearly, 2% of blood is discarded (minimum) due to various reasons. If we deduct 2% of discarded blood, the total usable whole blood or red cells will be 6460,000 units in India. For blood components, let us take a conservative estimate that only 25% blood is separated into components. In that situation, we will have about 1,365,000 components for patients. Now to find out the total revenue generation across the country, let us take the service charge ceiling laid down by the National AIDS Control Organization (NACO). NACO has prescribed ₹ 850 per unit of whole blood or RBC and 6460,000 units will generate ₹ 549, 1000,000. On the other hand, components will attract revenue of ₹ 68, 2500,000 (@ ₹ 500 per component on an average). Total revenue generated by whole blood/red cells and components is ₹ 617, 3500,000 (or US\$123270000 @ 1 USD = ₹ 50). We have four types of blood banks/centers (from the administrative point of view) in India. They are managed by the public (government) sector, Indian Red Cross Society (IRCS), nongovernment organizations (NGOs, on not for profit basis) and corporate or commercial sectors. Let us discuss today how efficiently more than 2,460 blood banks in India are managed. Roughly, about 55% blood banks are from the government sector, 5% from the IRCS, about 20-25% are from the NGO sector and the rest are from corporate or profit-making sectors. In this article author present a one major problem every year our nation requires about 4 Crore units of blood.

3.4. Yang Tang, Patrick P.C. Lee al et. [5] Out of which only average 5 Lakh units of blood are available. It is not that, people do not want to donate blood. Often they are unaware of the need and also they do not have a proper facility to enquire about it. As a result, needy people end up going through a lot of pain. India has many blood banks, all-functioning in a decentralized fashion. In the current system, individual hospitals have their own blood banks and there is no interaction between blood banks. The management is adhoc with no resemblance of organization or standard operating procedures. Donors cannot access blood from blood banks other than the bank where they have donated blood. In present system all the blood banks are attached to hospitals and there is no stand-alone blood bank as we have seen during COVID-19 pandemic in 2020-2021.

3.5. C. Wang et al. [6] Indian Railway Reservation related concept we are included in this part of literature survey , this is include the some existing research work its concept already done by the researcher apart form that we are also find the some limitation in the same field. Indian Railways is the principal mode of transport in the country. It is one of the world's largest rail networks under a single management. The route length is around 63,332 km with more than 8000 stations [18]. As it is the backbone of

nation's transport system, IR owns more than 25,000 wagons, 45,000 different types of coaches and 8000 locomotives. The system carries about 5,000 million passengers generating a traffic output of 340 billion passenger kilometers.

3.6. Q. Wang et al. [7] The passenger reservation system was given the name Integrated Multi-train Passenger Reservation System (IMPRESS) which could handle the functional requirements of reservation, enquiry, accounting and charting.

3.7. K. Popovic and Z. Hocenski et al. [8] This paper explores the need, development of computerized passenger reservation system for Indian railways. The passenger reservation system is based on online transaction processing. It was given the name Integrated Multitrain Passenger REServation System (IMPRESS) which later developed into single image system called Countrywide Network Of Computerized Enhanced Reservation (CONCERT). Then this paper goes onto the general architecture of the system which is 3-tier client server architecture. Apart from the obvious advantage of being computerized over manual reservation and enquiry, this paper explores other advantage of implementing this system and its future.

3.8. S. Ramgovind, M.M. Eloff, and E. Smith et al. [9] Computerized Passenger Reservation System for Indian Railways .This paper presents the Railway Reservation System which is at Bapat Chourah, Indore, M.P, India, after studying advantages and disadvantages of system structure, this paper recommends changing the present queuing system to alternate queuing system, to avoid the inconvenience of passengers. It was proved that this model of the queuing system is feasible and the results are effective and practical.

3.9. T. Ristenpart et al. [13] Alternate queuing system for tatkal railway reservation system AASRFC Digital business is the need of the hour, 'digitalise or perish' is the slogan for every business of physical world businesses whether small, medium or large sized. The present study is on the study of e-ticketing services in India with special reference to the IRCTC (Indian Railway Catering & Tourism Corporation Ltd). The study also covers the consumers' perspectives towards IRCTC'S e-ticketing services. Analysis shows that the consumers' perception towards IRCTC e-ticketing services is positive, however there are some glitches which need to be overcome.

3.10. B. Grobauer, T. Walloschek, and E. Stöcker et al. [14] This paper proposes the Dynamic Seat Allocation (DSA) system considering the advantage of QR code processing along with one of the standards of wireless communication. Moreover, dynamic authentication to the wireless device is incorporate which is passenger specific. Through this research paper our approach is to make fair processing in seat reservation or allocation in Indian Railway.

3.11. M.A. Morsy, J. Grundy, and I. Müller et al. [15] In the current scenario of Indian Railway a device called palmtop is given to ticket checker. This device will replace the reservation list carried by TTEs. Palmtop is connected to central server. From the Palmtop TTE update the

passenger's presence in the train to central server. From this, the status of availability of seats would be shown on the railway network and that could be booked by anyone who is willing to travel in the train at the upcoming stations. But, here we proposed that, if there are waiting list passengers and if any seat is available during their journey then the seat is provided to that passenger which is having waitlist 1, which means available seat is allocated on first come first serve basis.

3.12. H. K. Maji, M. Prabhakaran, and M. Rosulek et al. [23] This system has been developed by the Center for Railway Information System (CRIS), a railway public sector company which provides consultancy and IT services to the human transport system. Presently, this system is operational in Amritsar, Ajmer and Dehradun Shatabdi Express. They suggested an auction based mechanism for selling the tickets to the in need passengers. They used the fact that during TATKAL bookings for different distances fixed price is charged due to which Indian Railway loses a significant amount of money. But we are focusing on asking a higher fixed price that too for a fixed time interval so that the other regular customers do not get affected due to this change.

3.13. F. Zhao, T. Nishide, and K. Sakurai et al. [24] In this literature survey I am used the data from Gmail blog [16], this is an official site that provide the information about the Gmail services. As we all know that the facilities provide by the Gmail these are:-Gmail account, Gmail apps is a Google Apps is a suite of communication and collaboration tools, including Gmail, Google Calendar, and Google Docs. For the security the Gmail is provide the custom signature services in this services Email signatures are automatically inserted at the bottom of every message you send, and can be a great place to add your title, contact information, and even the latest news from your company. Just go to Settings at the top of your inbox and enter your signature text in the box at the bottom. Another facilities is provide by the Gmail is manage the multiple account If you're like a lot of business owners, you probably regularly receive email in several different accounts.

3.14. Divya bharathy S, Ramesh T et al. [17] By centralizing your correspondence in Gmail, you'll be able to keep track of it all more easily. Embrace labels. Folders are familiar, especially when it comes to work email. If you want to organize your emails in a similar way, make sure you're using Gmail labels. Combined with filters, they can be a powerful tool to manage your mail. Create labels for projects, vendors, customers, weekly reports, launches, to-do --the list goes on. You can also add custom colors to your labels, order them based on priority, and search the contents of specific labels. And don't forget that you can drag messages into labels, just like you can with folders. Use offline Gmail anytime you're not online. Despite having Internet access almost everywhere, work may take you to places where you just can't get online. Turn on offline Gmail from the Offline tab under Settings, and Gmail will download a local cache of your mail which synchronizes with Gmail's servers while you're connected. When you lose connectivity, Gmail automatically switches to offline mode, so you can continue to work, and your

replies are automatically sent the next time Gmail detects a connection. Create scanned responses and quickly reply to common questions. When it comes to emailing at work, you're probably used to sending out weekly reports, or answering the same questions from customers or colleagues multiple times. That's where canned responses can save precious time: turn on this feature in Gmail Labs, compose your response once, save it, and then use it over and over again.

3.15. C. Wang, Q. Wang, K. Ren, N. Cao and W. Lou et al. [18] A careful analysis of literature on the variants and methodologies of privacy preserving in cloud computing reveals the following: So many method are already exiting for auditing cloud content before storing cloud Environment, this will be done via third person or some time called TPA .The user might give his/her identity of proof certificate [19] This paper includes the Problems of misuse of the proof of identity (POI) certificate if fallen into unauthorized person. However, public auditing on the integrity of shared data with these existing mechanisms will inevitably reveal confidential information identity privacy-to public verifiers.

3.16. T.Schneider, and M. Winandy et al. [20]-decentralized key management work for providing a security to cloud data. Security and privacy protection in clouds are being explored by many researchers.

3.17. R. Lu, X. Lin, et al. [21] Many holomorphic encryption techniques have been suggested [22] to ensure that the cloud is not able to read the data while performing computations on them. Token based [23] Secure provenance has been studied in [24]. User Based Access Control (UBAC), Role Based Access Control (RBAC), and Attribute Based Access Control (ABAC). Access control in online social networking has been studied in [25] Attribute Based Signature (ABS) has been applied. ABS was proposed by Maji et al. [20]. In ABS, users have a claim predicate associated with a message. The claim predicate helps to identify the user as an authorized one, without revealing its identity. Other users or the cloud can verify the user and the validity of the message stored. ABS can be combined with ABE to achieve authenticated access control without disclosing the identity of the user to the cloud. Earlier work by Zhao et al. [25] provides privacy preserving authenticated access control in cloud. However, the authors take a centralized approach where a single key distribution center (KDC) distributes secret keys and attributes to all users.

3.18. S. Jahid, P. Mittal et al. [22] proposed a decentralized approach; their technique does not authenticate users, who want to remain anonymous while accessing the cloud.

3.19. A.-R. Sadeghi et al. [20] proposed a distributed access control mechanism in clouds. However, the scheme did not provide user authentication new protocol known as Attribute Based Signature (ABS) has been applied. In ABS, users have a claim predicate associated with a message. The claim predicate helps to identify the user as an authorized one, without revealing its identity. Other users or the cloud can verify the user and the validity of the message stored. ABS can be combined with ABE to

achieve authenticated access control without disclosing the identity of the user to the cloud.

4. RESEARCH GAP

- 4.1. Understanding of the cloud computing concept in relation to user privacy and security. [5]
- 4.2. Classification of cloud components, threats, and security implementations based on the STRIDE model. [7]
- 4.3. Providing security and privacy classifications based on attack mitigation and addictiveness. [10]
- 4.4. Providing different approaches to what and how existing works in the literature have provided solutions to cloud computing security and privacy. [18]

5. PROPOSED SOLUTION & METHODOLOGY

People lived in rural area not have knowledge about computer but those people have a knowledge about mobile phone .In this paper, we proposed the solution for improving the services of conventional blood bank management system, using a new technology cloud computing. [7] This technology explore the mechanisms of decision making support in blood bank information systems in rural area. Methodology used -Firstly, the properties of data and decisions in a blood bank are examined carefully; then, we introduce the development of computerized decision making support with special concerns on blood donation and transfusion service; This project is very help full for casualty case like Accidental cases deliver cases etc. [8]

6. Comparative Analysis of Other Method Table for Blood Bank System-

Objective of Paper/Article/ Blog	Information of Donor for Public	Applicable for Rural Area	Paper Domain
Only Donors Age Group Cluster	NO	NO	Technical
Blood Management in all aspects	NO only report are in the SANBS web site	NO only in SOUTH AFRICAN NATIONAL BLOOD SERVICE NPC	Annual Report Biomedical information
Donor Criteria as per medical aspects, so some reason for deferral donor	NO	NO	Biomedical Information and case study Blog, Link are mention, data available in the web site
News related the Illegal Market of Blood in India	NO	NO	New Blog Article, Link are mention, data available in the web site
Blood Factory in India	NO	NO	New Blog Article, Link are mention, data available in the web site
Blood donor sale blood up to 500-1500 per unit	NO	NO	New Blog Article, Link are mention, data available in the web site
Survey of Some of the Indian place that involve in the black market of blood	NO	NO	New Blog Article, Link are mention, data available in the web site
Blood Component Storage, and in this paper author also show the various reason (Biomedical) due to that a blood components are wastage.	NO	NO	Biomedical research paper
Cloud technology used for providing information of blood donor.	Yes with some condition	Yes	Technical paper
Data Mining Technique used for blood bank field with KDD	NO for the public use	NO	Technical paper
Provide security to conventional blood bank system	NO	NO	Technical
Web based e-health blood bank system of Kenyan health sector	NO donor registration process only in the website	NO	Technical
Analysis the large data set of blood bank using k means algorithm	NO	NO	Technical
Cloud Technology used for providing the blood donor information form the blood donation centre BTS and BDS proposed	Yes only for the stakeholders	NO	Technical

7. CONCLUSION

The Blood Bank, Indian Railway, Gmail Services, and Public Data Verifier are some of the primary generic areas of interest at work that are being improved by current cloud services. During the literature review, we discovered a few issues with blood bank administration. Without the proper tools, blood cannot be kept for an extended period of time; therefore, expensive technologies are necessary, all of which are readily available in large cities but not in countryside. Secondly, with respect to railway services, the facility of waiting ticket confirmation during the journey is not available without the feedback given by the passenger to improve the particular field and provide better services. In the G-Mail Services, switching between G-Mail to Yahoo, vice-versa is not possible. In the context of Public Data Verifier, data is not checked during exchange capabilities at these facilities before being uploaded to the cloud as it is common knowledge that the cloud is used for data sharing and secure data retrieval. There are security issues/concerns related to public clouds & private clouds too.

8. REFERENCES

- [1] A Survey on Virtualization Technologies Susanta Nanda Tzi-cker Chiueh {susanta, chiueh}@cs.sunysb.edu Department of Computer Science SUNY at Stony Brook Stony Brook, NY 11794-4400
- [2] R. J. Adair, R. U. Bayles, L. W. Comeau, R. J. Creasy, "A Virtual Machine System for the 360/40", IBM Corporation, Cambridge Scientific Center Report No. 320-2007, 1966
- [3] NOVELL, Virtualization in Data Center, 2006, [Electronic resource] - www.novell.com/collaboration
- [4] Kuyoro S. O., Ibikunle F. & Awodele O. International Journal of Computer Networks (IJCN), Volume (3) : Issue (5) : 2011 247 Cloud Computing Security Issues and Challenges
- [5] Secure Overlay Cloud Storage with Access Control and Assured Deletion Yang Tang, Patrick P.C. Lee, Member, IEEE, John C.S. Lui, Fellow, IEEE, and Radia Perlman, Fellow, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 9, NO. 6, NOVEMBER/DECEMBER 2012.
- [6] C. Wang et al., "Toward Publicly Auditable Secure Cloud Data Storage Services," IEEE Network, vol. 24, no. 4, 2010, pp. 19-24.
- [7] Q. Wang et al., "Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing," IEEE Trans. Parallel and Distributed Systems, vol. 22, no.
- [8] K. Popovic and Z. Hocenski, "Cloud Computing Security Issues and Challenges," Proc. 33rd Int'l Convention on Information and Comm. Technology, Electronics and Microelectronics (MIPRO 10), IEEE Press, 2010, pp. 344-349.
- [9] S. Ramgovind, M.M. Eloff, and E. Smith, "The Management of Security in Cloud Computing," Proc. Information Security for South Asia (ISSA 10), IEEE Press, 2010, pp. 1-7.
- [10] H. Takabi, J.B.D. Joshi, and G.-J. Ahn, "SecureCloud: Towards a Comprehensive Security Framework for Cloud Computing Environments," Proc. 2010 IEEE 34th Ann. Computer Software and Applications Conf. Workshops, IEEE Press, 2010, pp. 393-398.
- [11] S. Subashini and V. Kavitha, "A Survey on Security Issues in Service Delivery Models of Cloud Computing," J. Network and Computer Applications, vol. 34, no. 1, 2010, pp. 1-11.
- [12] M. Zhou et al., "Security and Privacy in Cloud Computing: A Survey," Proc. 6th Int'l Conf. Semantics, Knowledge and Grids, IEEE Press, 2010, pp. 105-112
- [13] T. Ristenpart et al., "Hey, You, Get Off of My Cloud: Exploring Information Leakage in Third-Party Compute Clouds," Proc. 16th ACM Conf. Computer and Communications Security (CCS09), ACM Press, 2009, pp. 199-212.
- [14] B. Grobauer, T. Walloschek, and E. Stöcker, "Understanding Cloud-Computing Vulnerabilities," IEEE Security and Privacy, vol. 9, no. 2, 2011, pp. 50-57
- [15] M.A. Morsy, J. Grundy, and I. Müller, "An Analysis of the Cloud Computing Security Problem," Proc. 17th Asia Pacific Software Eng. Conf. 2010 Cloud Workshop (APSEC 10), IEEE Press, 2010
- [16] Virtualization's Impact on Cloud Security Hsin-Yi Tsai, IT Pro January/February 2012 published by the IEEE Computer Society 1520-9202/12/\$31.00 © 2012 IEEE.[19] Bertino, E.; Paci, F.; Ferrini, R. 2009 Privacy-preserving Digital TPA Identity Management for Cloud Computing, IEEE Computer Society Technical Committee on Data Engineering.
- [17] Divya bharathy S, Ramesh T 2014 IEEE International Conference on Innovations in Engineering and Technology (ICIET'14) on 21st& 22nd March Organized by K.L.N.
- [18] C. Wang, Q. Wang, K. Ren, N. Cao and W. Lou, "Toward Secure and Dependable Storage Services in Cloud Computing", IEEE T. Services Computing, vol. 5, no. 2, pp. 220-232, 2012
- [19] C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Stanford University, 2009, <http://www.crypto.stanford.edu/craig>.
- [20] A.-R. Sadeghi, T. Schneider, and M. Winandy, "Token-based cloud computing," in TRUST, ser. Lecture Notes in Computer Science, vol. 6101. Springer, pp. 417-429, 2010.
- [21] R. Lu, X. Lin, X. Liang, and X. Shen, "Secure rovenance: The Essential of Bread and Butter of Data Forensics in Cloud Computing," in ACM ASIACCS, pp. 282-292, 2010.
- [22] S. Jahid, P. Mittal, and N. Borisov, "EASiER: Encryption-based access control in social networks with efficient revocation," in ACM ASIACCS, 2011.
- [23] H. K. Maji, M. Prabhakaran, and M. Rosulek, "Attribute-based signatures: Achieving attributeprivacy and collusion-resistance," IACR Cryptology ePrint Archive, 2008.
- [24] F. Zhao, T. Nishide, and K. Sakurai, "Realizing fine-grained and flexible access control to outsourced data with attribute-based cryptosystems," in ISPEC, vol. 6672. Springer, pp. 83-97, 2011.
- [25] Kan Yang, Xiaohua Jia and Kui Ren, "DAC-MACS: Effective Data Access Control for Multi-Authority Cloud Storage Systems", IACR Cryptology ePrint Archive, 419, 2012.