International Journal of Intelligent Computing and Information Science

A NOVEL SECURITY MODEL FOR DATA MANAGEMENT IN CLOUD COMPUTING

S. S El Ashry M. A. El-dosuky M. Z. Rashad H. S. Abdelkader

: Faculty of computer, Mansoura University, Egypt

Faculty of computer, Menofia University, Egypt

Sobhank_raby@yahoo.com Dr_dos_ok@ yahoo.com Magdi_z2011@yahoo.com Hatem6803@yahoo.com

Abstract: *Data security is always the focus of huge possible cloud clients, also a big obstacle for its extensive applications. Till now there is no proficient mechanism for data security accustomed for the cloud environment, and various service types need different solutions for data protection. The objective of this research is to develop a new data security model for the cloud environment. The proposed model is decomposed of four modes of operations denoting the different ways for providing the service for better user convenience. The four modes are: Scheduling mode, Through-off mode, Virtual machine mode, and Batch mode. Analyzing the proposed model shows the advantages it has over previous models. Based on extensive experiments, the proposed model covers many security cases, and is robust in handling security threats while gaining user convenience.*

Keywords*: Cloud Computing, Data Management, Security Model, Simulation.*

1. Introduction

Cloud Computing, abbreviated as CC, offers services to consumers according to the provided service level. There are three service categories as follows: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and software as a Service (SaaS) [1, 22]. For these three types, security and protecting personal privacy have been ranked as risk [2,23], However using the conventional methods of cryptography become inappropriate to provide a proficient technique for the clouds environments [3]. In conventional software development, security issue has not a main concern, but user privacy protection in CC is a major issue. This is because the user data often in unencrypted pattern are located in some machines and this user data maybe contains several company operators, commercial sensitive data, as well as potential damage in materials privacy [4]. So, safety of client sensitive data that hosts in the virtual servers of the cloud is extremely significant. During the establishment of the cloud service system that achieve control of data security challenge, the cloud providers have to consider of all factors types and improves client trust level in the laws case, has to evaluate and obtain the system measure with principle of each phase design [5].

CC on World Wide Web is a significant improvement application, the cloud systems serve consumers without need to master control power hardware [6], the customer of the software service that the cloud offers, processing and storing data like individual profiles, credit cards, health and finance in addition to other data. The characteristics of cloud are for offering data storage services, processing, and platform

service outflow to universal consumers, however this material set result is data protection, leading to the popular reason for the enterprises otherwise individuals worried about using the clouds services solutions packages. Also, the data security in CC still needs more researches are done on this topic as it has several problems with no solution. Till now there is no proficient mechanism for CC data security system [2], and several CC service types need several solutions for data protection. This research objective is developing a new data security model for CC.

This paper is organized as the following: section 2 reviews the cloud working principle, CC management, and previous models. Section 3 shows the proposed model which is decomposed of four modes of operations denoting the different ways for providing the service for better user convenience. The four modes of operations are: Scheduling mode, Through-off mode, Virtual machine mode, and Batch mode. In section 4, the used tool and different configuration of the tested scenarios .In section 5, the results are described. Finally in section 6 the conclusions.

2. Related Work

2.1 The Clouds Working Principle.

The clouds systems consist of the network infrastructure provider (IaaS), and the platform service provider (PaaS) and software service provider (SaaS) [7,26]. From Google CC Trends analysis chart can be viewed that CC has well-respected degree grown year after year. Also, the database manufacturers' products are joining the purpose of the clouds to support the database. For example: Oracle currently is operating a service of clouds computing platform (EC2) directly in Amazon [8]. So launching more and more data in the clouds service, causing that the data safety will still has a serious concern, as these data often include significant sensitive information for the organizations otherwise the individuals.

2.2 CC Management Approach.

Hardware devices of Clouds environment is offered via an external third party agency entrusted with clouds i.e. TTP Clouds otherwise an interior cloud. Clouds environments maybe limited in several organizations share the public cloud i.e. Public Clouds otherwise in private cloud enterprises i.e. Private Clouds [9, 10, 24].

2.3 The Classification of Security Requirements.

The of classification level for the security challenge must be considered as the following [7]: (1) security of server/s, (2) security of Internet (3) security of database (4) data security (5) security of applications.

2.4 Previous Models.

The basic data model of the distributed system is Client-server [11, 12, 13], There are many models are proposed for incorporating security in the distributed systems such as: Kerberos [14-17, 25] and GARS [18]. Among the problems facing these previous models are the delivery of service which is a significant aspect of quality of service (QOS), is postponed until verification process is performed. Another critique is that the service is not provided in a continues manner.

The [19] threat modeling process is as follows:

1. Identify the known threats to the system.
2. Rank the threats in order by decreasing risk.
3. Determine how you will respond to the threats.

4. Identify techniques that mitigate the threats.
5. Choose the appropriate technologies from the identified techniques.

2.4.1 STRIDE Model

A model you may find useful for data security in clouds environment is STRIDE by Microsoft, derived from an acronym for the follow six threat categories: S stands for "Spoofing identity", T stands for "Tampering with data", R stands for "Repudiation", I stands for "Information disclosure", D stands for "Denial of service" and E stands for "Elevation of privilege" [20].

3. Proposed Model

Definition 1: Overall required security (ORS)

Overall required security (ORS) is the required security that managed by the provider.

Definition 2: Computing Capacity(C)

Computing Capacity(C) is the all ability of the datacenters.

For heterogeneous

$$C = \sum_{k=1}^{d} v_k c_k \quad (1)$$

Where d is number of the data centers, v_k is number of the virtual machines in datacenter k, and c_k is the computing capacity of virtual machine v_k .
For homogeneous

$$C = d.\ v \quad (2)$$

Where d is number of the data centers, and v is number of the virtual machines.

Notes: For simplicity, we will use equation (2). At design time C is varying, but once it is in operation is considered a constant threshold, because it does make any difference if it is homogeneous or heterogeneous.

Definition 3: Total number of requests(R)

Total number of requests(R): the whole number of the users' requests.
Similarity the total requests for heterogeneous

$$R = \sum_{i=1}^{u} \frac{(r_i.t_i.c_i)}{g_i} \quad (3)$$

Where u is number of the user bases, ri is rate of the requests per hour, ti is the duration of the simulation session, ci is computational size of request i , gi is grouping factor of user bases.

For homogeneous

$$R = \frac{u.r.t}{g} \quad (4)$$

Where u is number of the user bases, r is rate of the requests per hour, t is the duration of the simulation session, g is grouping factor user bases.
For simplicity, we will use equation (4).

Definition 4: Security Measure (M):

Security Measure (M): is the measure of the total security threats, based on the following assumptions:

1) The ORS is proportional to size of user bases that assess the cloud services.
2) The ORS is proportional to the value of data.

We have two dimensions that are determine the required security:

1. Value of data (more value, more required security)
2. User base size (more requests, more required security) to decide the appropriate mode type for clients among the four components of the proposed model. This does not mean to ignore less-value data .these less-value data need to be secured but not as the security level of the more-value data.

3.1 The Model

After a deep study and analysis of the related existing systems, we designed our proposed model for clouds applications. This model is decomposed four modes of operations denoting the different ways for providing the service for better user convenience as in our suggested model we tried to avoid the mentioned problems in section 2.4 that faced implementing data security in distributed systems. This optimized model consists of four modes of operations as follows:

1. Scheduling Mode.
2. Through-off Mode.
3. Batch Mode.
4. Virtual machine Mode

Each mode of operation has specific characteristics

Table 1: Mode of operation characteristics.

Mode Name	User base size	Value of data
Scheduling Mode	$< C$	$< \theta$
Through-off Mode	$> C$	$> \theta$
Batch Mode	$< C$	$> \theta$
Virtual Machine Mode	$> C$	$< \theta$

Table 1 shows the four modes of operation with their features .In scheduling mode the user base size is less than computing capacity and the value of data is less than a specific threshold .In through-off mode the user base size is greater than computing capacity and the value of data is greater than a specific threshold. In batch mode the user base size is less than computing capacity and the value of data is greater than a specific threshold. In virtual machine mode the user base size is greater than computing capacity and the value of data is less than a specific threshold.

3.2 The Algorithm

We have two dimensions that are determine the required security:
1-Value of data (more value, more required security)
2-User base volume (more requests, more required security) to decide the appropriate mode type for clients among the four components of the proposed model. Figure1shows the main algorithm

4. Evaluation

4.1 Tool

4.1.1 CloudAnalyst Simulator

It is a simulator which supports simulating large-scale Clouds applications to study these applications behavior with a variety of configuration. This tool can helping developers to understand applications distribution ways across clouds infrastructures as well as evaluate the services like the providers incoming by using Services Brokers and application performance optimization [21].

4.2 Configuration

We used CloudAnalyst software in our simulation with the following configuration:
D=1, v=10, so C=1*10=10, u=1, scheduling techniques is RR, t = 1 hour.

Algorithm 1 main ( )

d is number of the data centers
v is number of the virtual machines
C is the computing capacity
u is number of the user bases
r is rate of the requests per hour
g_u is the users grouping factor
g_v datacenter grouping factor
t is duration of the simulation session
θ is a constant that represents a threshold for security measure
M is the security measure

C=d.v

$R = \frac{u.r.t}{g}$

If R-C<0 &M< θ& g_v >1 **Then**
Scheduling mode
Else If R-C>0 &M> θ& **Then**
Through-off mode ( )
Else If R-C<0 &M< θ& g_v >1 **Then**
Batch mode
Else If R-C<0 &M< θ& g_v >1 **Then**
Virtual Machine mode ( )
Else
Error
End If

Figure1: The Main Algorithm.

4.3 Test Scenarios:

The first scenario consists of one Data center with 10 virtual machines, R=C=10 & gu=1,gv=1, we notice that the overall response time is 300.24ms and Data Center processing time: is 0.26 ms. The second scenario consists of one Data center with 10 virtual machines, R=15 & g_u=1,gv=1, we notice that the overall response time is 300.92 ms and Data Center processing time: is 0.25 ms. The third scenario consists of one Data center with 10 virtual machines, R=5 & g_u=1,gv=1, we notice that the overall response time is 300.14 ms and Data Center processing time: is 0.26 ms. In the next scenario there is one Data center with 10 VMs, R=10, &g_u=10,gv=10, we notice that the overall response time is 300.48 ms and Data Center processing time: is 0.36 ms. Whereas in the fifth scenario consists of one Data center with 10 virtual machines, R=15 & g_u=10,gv=10, we notice that the overall response time is 299.69 ms and Data Center processing time: is 0.37ms. The sixth scenario consists of one Data center with 10 virtual machines, R=5 & g_u=10,gv=10, we notice that the overall response time is 300.09 ms and Data Center processing time: is 0.37 ms. In the next scenario, there is one Data center with 10 virtual machines, R=10 & g_u=10,gv=1, we notice that the overall response time is 300.38 ms and Data Center processing time: is 0.26 ms. The eighth scenario consists of one Data center with 10 virtual machines, R=15 & g_u=10,gv=1, we notice that the overall response time is 299.58 ms and Data Center processing time: is 0.26 ms. Finally, in the last scenario one Data center with 10 virtual machines, R=5 & g_u=10,gv=1, we notice that the overall response time is 299.98 ms and Data Center processing time: is 0.26 ms. Table 2 shows the settings of tested scenarios.

NOTE: As shown in table 2, the difference between the response times (ms) of the pervious tested scenarios is so little. Clearly, the difference is in decimal percentages. Thus the settings maps of the pervious scenarios look like the same and also the figures of response times of tested scenarios seem the same. So, presenting the map of only one scenario (to show the GUI of the used simulator and implementing for this scenario) will be sufficient. Figure 2 shows CloudAnaylst GUI & settings map of test scenario #1. And there is no need to show the maps of the rest scenarios. Figure 3 shows the response time of test scenario #1, we notice that the simulation time is concentrated in the duration from 0-2 hours(the first two hours) as the simulation duration is t = 1 hour.

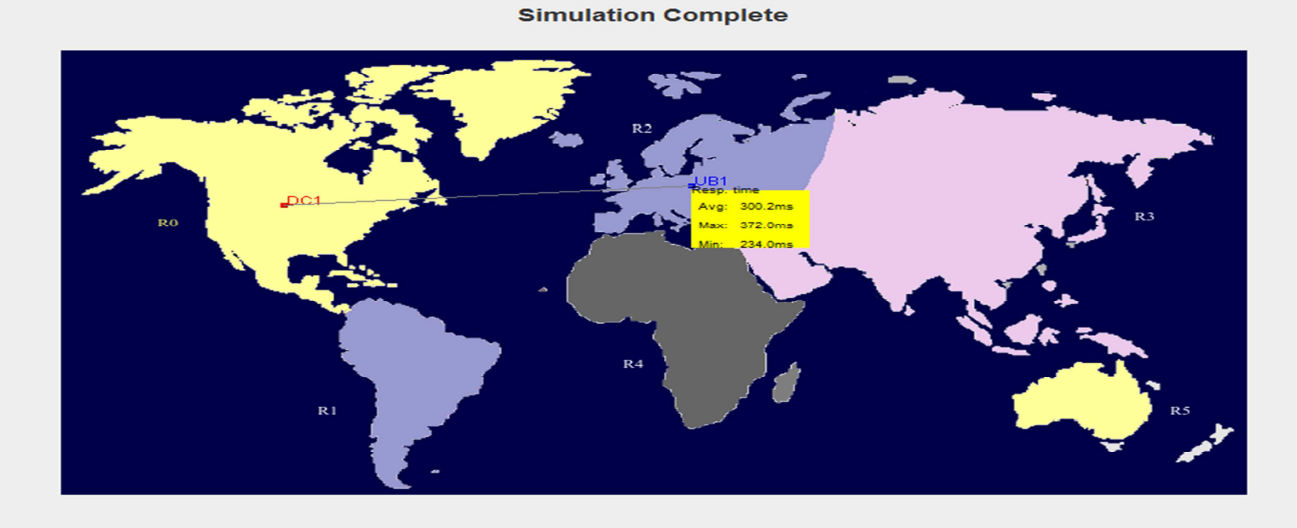


Figure 2: CloudAnaylst GUI & settings map of test scenario #1.

Table 2: Simulation settings & results of experiments.

	Simulation settings	Overall average response time (ms)	Data Center processing time for a request (ms)	VM Cost
1	One Data center with 10 Virtual Machines , R=10 & g_u=1,gv=1	300.24	0.26	$ 1.11 1
2	One Data center with 10 Virtual Machines , R=15 & g_u=1,gv=1	300.92	0.25	$ 1.02
3	One Data center with 10 Virtual Machines , R=5 & g_u=1,gv=1	300.14	0.26	$ 1.18
4	One Data center with 10 Virtual Machines & R=C &g_u=10,gv=10	300.48	0.36	$ 1.11
5	One Data center with 10 Virtual Machines & R>C &gu=10,gv=10	299.69	0.37	$ 1.02
6	One Data center with 10 Virtual Machines & R<C &g_u=10,gv=10	300.09	0.37	$1.18
7	One Data center with 10 Virtual Machines & R=C &gu=10,gv=1	300.38	0.26	$ 1.11
8	One Data center with 10 Virtual Machines & R>C &gu=10,gv=1	299.58	0.26	$ 1.02
9	One Data center with 10 Virtual Machines & R<C &gu=10,gv=1	299.98	0.26	$1.18

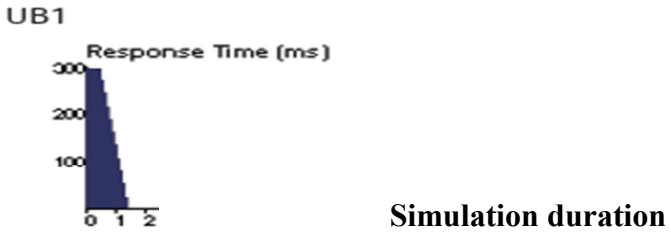


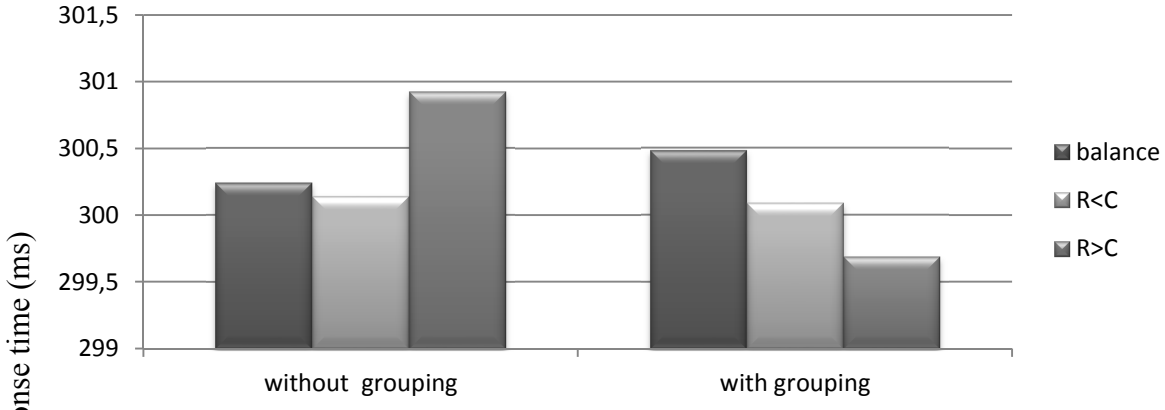
Figure 3: The response time of test scenario #1.

5. Results and Discussion

Table 3 summarized the results.

	without grouping	with grouping
balance	300.24	300.48
R<C	300.14	300.09
R>C	300.92	299.69

Figure 4 shows the summary of experiments based on whether there is grouping or not.



gure 4: The summary of experiments based on whether there is grouping or not.

gure 5 shows a comparison between the three cases of balanced(R=C), (R<C), and(R>C).

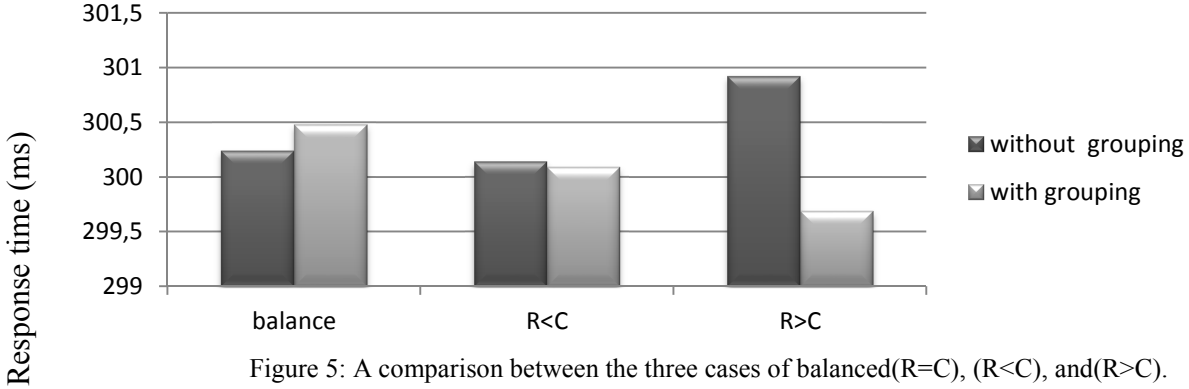


Figure 5: A comparison between the three cases of balanced(R=C), (R<C), and(R>C).

- Three cases are tried without grouping for the users & virtual machines, three cases with grouping for both the users & virtual machines, three cases with grouping for the users & without grouping for the virtual machines. CloudAnaylst doesn't accept that the grouping factor for the virtual machines greater than grouping factor of the users.
- In the first three scenarios, we notice that overall average response time (ms) increase in scenario #2 when we increase the total requests of users (R) compared with scenario#1 and the overall average response time (ms) decrease when the total requests of users (R) decrease as in scenario #3.

- When R=C=10(balance mode) , the overall average response time (ms) increase in case of grouping for the users & virtual machines(i.e. grouping factor of the users & Datacenter =10) and also increase in case of grouping for the users & without grouping for the virtual machines(i.e. grouping factor of the users=10 & grouping factor of the Datacenter =1) when it is compared with the case (scenario#1) of without grouping for the users & virtual machines (i.e. grouping factor of the users & Datacenter =1).
- When R=5(less than balance mode) , the overall average response time (ms) decrease in case of grouping for the users & virtual machines(i.e. grouping factor of the users & Datacenter =10) and also decrease in case of grouping for the users & without grouping for the virtual machines(i.e. grouping factor of the users=10 & grouping factor of the Datacenter =1) when it is compared with the case(scenario#3) of without grouping for the users & virtual machines (i.e. grouping factor of the users & Datacenter =1).
- When R =15(greater than balance mode). the overall average response time (ms) decrease in case of grouping for the users & virtual machines(i.e. grouping factor of the users & Datacenter =10) and also decrease in case of grouping for the users & without grouping for the virtual machines(i.e. grouping factor of the users=10 & grouping factor of the Datacenter =1) when it is compared with the case(scenario#2) of without grouping for the users & virtual machines (i.e. grouping factor of the users & Datacenter =1). This may be counter-intuitive, but there is some logic behind that. If there is a grouping, the response time decreases regardless that the total user requests is less than or greater than computing capacity.

6. Conclusions

In this research we discuss the data security challenge to cloud and propose an optimized secure model for data management in cloud computing environment. We can summarize the main contributions as follows:

- The model hypothetical foundation is the overall required security is determined by two factors which are the total number of requests (R) with focusing in measuring it and the security measure (M). Determine mode of operation is based on M and R.
- This hypothesis leads to dividing the model into four modes of operation namely: Scheduling mode, Through-off mode, Virtual machine mode, and Batch mode.
- Based on extensive experiments, the proposed model covers many security cases, and is robust in handling security threats while gaining user convenience.
- Among interesting observations is the effect of the grouping on the overall required security. If there is a grouping, the response time decreases regardless that the total user requests is less than or greater than computing capacity.

A possible future work direction is implementing the proposed model by incorporating different combinations of symmetric encryption and hashing algorithms. A clustering algorithm such as K-mean [27] can be used in calculating value of data (M). Then, Hidden Markov Model (HMM) [22] may be used in determining the appropriate mode for users dynamically.

References

1. M. Armbrust, A. Fox, R. Griffith et al., Above the Clouds: ABerkeley View of CC, 2011.
2. L. Gu and S.-C. Cheung, "Constructing and testing privacy aware services in a CC environment—challenges and opportunities," in Proceedings of the 1st Asia-Pacific Symposiumon Internetware (Internetware '09), October 2009.
3. M. Jensen, J. Schwenk, N. Gruschka, and L. L. Iacono, "On technical security issues in CC," in Proceedings of the IEEE International Conference on CC (CLOUD '09), pp. 109–116, September 2009.
4. M. Mowbray and S. Pearson, "A client-based privacy manager for CC," in Proceedings of the 4th International ICST Conference on Communication System Software and Middleware (COMSWARE '09), Dublin, Ireland, June 2009
5. S. Pearson, "Taking account of privacy when designing CC services," in Proceedings of the ICSE Workshop on Software Engineering Challenges of CC (CLOUD '09), pp. 44–52,Vancouver, Canada, May 2009.
6. B. R. Kandukuri, P. V. Ramakrishna, and A. Rakshit, "Cloud security issues," in Proceedings of the IEEE International Conference on Services Computing (SCC '09), pp. 517–520, September 2009.
7. L.-J. Zhang and Q. Zhou, "CCOA: CC open architecture," in Proceedings of the IEEE International Conference on Web Services (ICWS '09), pp. 607–616, Los Angeles, Calif, USA, July 2009
8. Amazon Elastic Compute Cloud—EC2, http://aws.amazon.com/ec2/.
9. L. Robert and G. Yunhong, *On the Varieties of Clouds for Data Intensive Computing*, 2009
10. Wikipedia, Cloud Computing, http://en.wikipedia.org/wiki/ CC#cite note-idc-28.
11. www.saylor.org/site/wp-content/uploads/2012/12/PRDV251_WikiBooks_A-Bit-History-of-Internet-Chapter-5-Client-Server.pdf
12. Mary C. Lacity, Leslie P. Willcocks, Ashok Subramanian, A strategic client/server implementation: new technology, lessons from history, The Journal of Strategic Information Systems, Volume 6, Issue 2, June 1997, Pages 95-128, ISSN 0963-8687, http://dx.doi.org/10.1016/S0963-8687(97)00009-7. (http://www.sciencedirect.com/science/article/pii/S0963868797000097)
13. Indu Shobha Chengalur-Smith, Peter Duchessi, The initiation and adoption of client–server technology in organizations, Information & Management, Volume 35, Issue 2, 8 February 1999, Pages 77-88, ISSN 0378-7206, http://dx.doi.org/10.1016/S0378-7206(98)00077-9. (http://www.sciencedirect.com/science/article/pii/S0378720698000779)
14. Frederick Butler, IlianoCervesato, Aaron D. Jaggard, Andre Scedrov, Christopher Walstad, Formal analysis of Kerberos 5, Theoretical Computer Science, Volume 367, Issues 1–2, 24 November 2006, Pages 57-87, ISSN 0304-3975, http://dx.doi.org/10.1016/j.tcs.2006.08.040. (http://www.sciencedirect.com/science/article/pii/S0304397506005743)
15.]F. Pereniguez, R. Marin-Lopez, G. Kambourakis, S. Gritzalis, A.F. Gomez, PrivaKERB: A user privacy framework for Kerberos, Computers & Security, Volume 30, Issues 6–7, September–October 2011, Pages 446-463, ISSN 0167-4048, http://dx.doi.org/10.1016/j.cose.2011.04.001. (http://www.sciencedirect.com/science/article/pii/S0167404811000617)
16. Anish Prasad Shrestha, Dong-You Choi, Goo Rak Kwon, Seung-Jo Han, Kerberos based authentication for inter-domain roaming in wireless heterogeneous network, Computers & Mathematics with Applications, Volume 60, Issue 2, July 2010, Pages 245-255, ISSN 0898-1221, http://dx.doi.org/10.1016/j.camwa.2010.01.019. (http://www.sciencedirect.com/science/article/pii/S0898122110000416)
17. Denis Russell, High-level security architectures and the Kerberos system, Computer Networks and ISDN Systems, Volume 19, Issues 3–5, November 1990, Pages 201-214, ISSN 0169-7552, http://dx.doi.org/10.1016/0169-7552(90)90073-2.

(http://www.sciencedirect.com/science/article/pii/0169755290900732
18. Chih-Yung Chen and Jih-Fu Tu,“ A Novel CC Algorithm of Security and Privacy”, Hindawi Publishing Corporation ,Mathematical Problems in Engineering,2013
19. https://msdn.microsoft.com/enus/library/ee810542(v=cs.20).aspx.
20. https://msdn.microsoft.com/enus/library/ee823878(v=cs.20).aspx.
21. Wickremasinghe B, Calheiros R, Buyya R. CloudAnalyst: A CloudSim-based visual modeller for analysing cloud computing environments and applications. Proceedings of the 24th IEEE International Conference on Advanced Information networking and Applications (AINA 2010), Perth, Australia, 20–23 April 2010; 446–452.
22. Harsha Banafar, Sanjay Sharma ,” Intrusion Detection and Prevention System for Cloud Simulation Environment using Hidden Markov Model and MD5”, International Journal of Computer Applications (0975 – 8887) Volume 90– No.19, March 2014.
23.] Namita N. Pathak , Meghana Nagori ,” Enhanced Security for Multi Cloud Storage using AES Algorithm” , International Journal of Computer Science and Information Technologies, Vol. 6 (6), 2015.
24. Arda Sezen, Ali Yazıcı , İbrahim Akman ,” cloud computing security issues and selection of deployment model and service model according to security requirements”, January 2015.
25. Yaser Fuad Al-Dubai , Khamitkar S. D,” Kerberos: Secure Single Sign-on Authentication Protocol Framework for Cloud Access Control”, Volume 14 Issue 1 Version 1.0 Year 2014.
26. Arash Mahjani , “Security issues of virtualization in cloud computing environments” , 2015
27.] Pardeep Kumar1 , Nitin2 , Vivek Sehgal3 , Kinjal Shah4 , Shiv Shankar Prasad Shukla Durg Singh Chauhan,” A Novel Approach for Security in Cloud Computing using Hidden Markov Model and Clustering”, IEEE, 2011.