INTELLIGENT AUTOMATION & SOFT COMPUTING, 2017
https://doi.org/10.1080/10798587.2017.1327634

Hierarchical Optimization of Network Resource for Heterogeneous Service in Cloud Scenarios

Dong Huang[a,b], Yong Bai[b], Jingcheng Liu[c], Hongtao Chen[d], Jinghua Lin[d] and Jingjing Wu[d]

[a]ChongQing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, ChongQing, P.R. China; [b]Network Center, ChongQing Electromechanical Vocational Institute, ChongQing, P.R. China; [c]School of Petroleum Engineering, ChongQing University of Science and Technology, ChongQing, P.R. China; [d]Network Research Center, Ruijie Network Incorporation, FuJian, P.R. China

ABSTRACT

With limited homogeneous and heterogeneous resources in a cloud computing system, it is not feasible to successively expand network infrastructure to adequately support the rapid growth in the cloud service. In this paper, an approach for optimal transmission of hierarchical network for heterogeneous service in Cloud Scenarios was presented. Initially, the theoretical optimal transmission model of a common network was transformed into the hierarchical network with the upper and lower optimization transmission model. Furthermore, the computation simplification and engineering transformation were presented for an approximation method at the low cost of computational complexity. In the final section, the average delay in the engineering method shows its influence on the capability of access for common nodes.

KEYWORDS

Hierarchical network; Hierarchical optimization; Cloud service

1. Introduction

Cloud service as a spontaneous extension of computing resources provides scalable resources and economic facilities to consumers over the network. It can provide storage services and data access, which do not need the foresight of the specific physical location and the systems configuration that provides the computing resources. In recent years, Cloud service has become an indispensable technology facilitating a seamless integration of mobile users and high performance servers. It is a style of computing where scalable resources are supplemented as a service over the homogeneous or heterogeneous network. With the aid of cloud computing, much of available computing resources can be accessed via network connections. On one hand, as much computing capability as they require can be supplied for any instant time, there is no need for users to consider arrangement for the coming expansion. Meanwhile resources are no longer needed to be released back to the cloud system. On the other hand, users can dynamically provide resources, for avoiding disturbance and intermittent spikes in computing bandwidth requirements.

For infrastructure aspects, distributed terminals can extend their computing capabilities by providing service from computing clouds dynamically. In a traditional parallel and distributed computing model, mobile terminals are in distributed clients that could only provide the limited functionalities of interfaces or consoles to the complicated applications as depicted in Figure 1. The unsustainable computational processing like forecast of a climate or mathematical optimization in scientific simulations is actually processed in the cloud (Vingelmann, Pedersen, Fitzek, & Heide, 2010). Therefore, various mobile applications could become more complex while the actual devices could become negligible.

Since cloud resources become scalable to the requirements of smooth running, they allow users to customize their cloud-based computing environment dynamically to tackle unexpected enhanced resource requirements, and then release the supplementary resources after they are not needed any more. For the scalability of computing resources in the traditional method, by means of investing in additional hardware, much more expenditure will also be involved in a much more complex and time-consuming transversion (Deng, Zhou, Zou, & Zhang, 2014; Jung & Chung, 2010; Prevost, Manooj, & Jamshidi, 2013).

There still exist some factors that limit the pervasive use of cloud services.

(1) Connectivity: For the cloud system completely relies on the uninterrupted network connection, if the state of connection fails, the accessibility to the cloud services become invalid, so unstable connections may lead to poor cloud service performance and unavailability for random access to the complicated application environment.
(2) Link Usage: If the data is stored on local resources, there will be no resource consumption on link for customers. Even when consumers have a big data center, there will be no cost of Link Usage, because the cost of link usage to data availability from distributed resources in cloud system could be much more than the bandwidth cost of allocated resources quantitatively.
(3) Manageability of infrastructure: When consumers receive the allocated resources from the cloud system, and have little manageability over its corresponding infrastructure, they cannot do anything if a cloud provider changes the infrastructure, as

CONTACT Jingcheng Liu liujingcheng1980@126.com; Yong Bai cqbaiyong@163.com

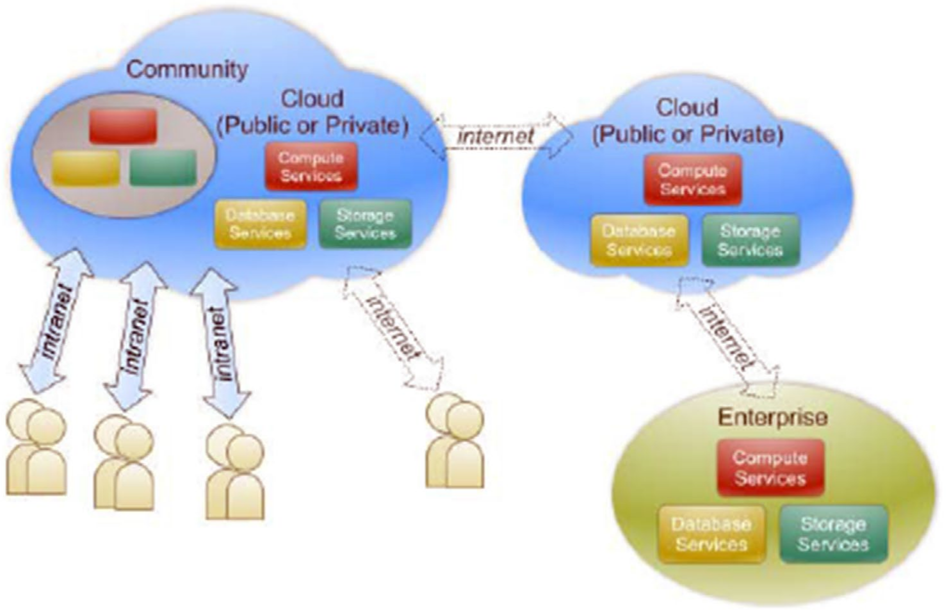


Figure 1. Hybrid Service in Cloud Computing.

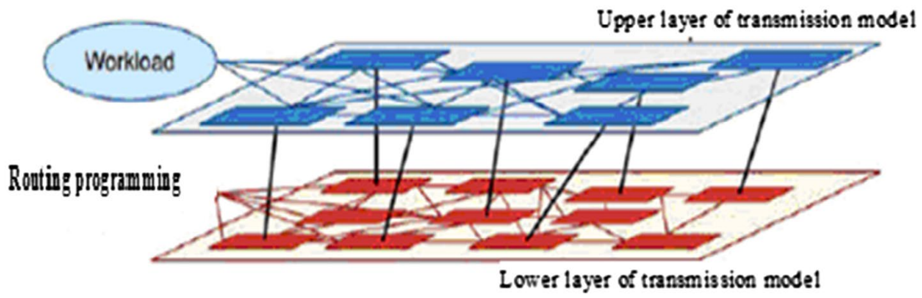


Figure 2. Hierarchical Model of Processing Units in Cloud Computing.

consumers have agreed on the cloud service agreement before allocating the service-oriented resource in the cloud system.

Therefore, development of cloud computing requires many functionalities that needs to be addressed to make it efficient. The performance optimization is the essential requirements when the multi-flows are transmitted in the cloud system.

2. Performance Formulation

For optimal scheduling of traffic flows, the hierarchical interconnection model was established, as depicted in Figure 2, and its optimization model can be divided into the upper optimization transmission model and lower optimization transmission model, respectively (Qian & Gagan, 2010).

Initially, we established the upper optimization transmission model, which has the tasks of traffic flow scheduling and link assignment. The constraints of the capacity for heterogeneous service at the instant time can be described as follows:

$$\theta_{j,l} - \theta_{i,j} \geq t^{u}_{j,l} - S(2 - z^{l}_{i,j,u}) + S(3 - 2\gamma^{l}_{i,u} - \gamma^{l}_{j,u})$$
$$\theta_{i,l} - \theta_{j,l} \geq t^{u}_{i,l} - Sz^{l}_{i,j,u} - S(2 - 2\gamma^{l}_{i,u} - 3\gamma^{l}_{j,u})$$

where $t^{u}_{i,l}$ and $t^{u}_{j,l}$ were constants, H the maximum tolerable duration, E the number of nodes, $M_{i,l}$ the number of the lower transmission model, $\theta_{i,l}$ the finishing time for the phenomenon that the ith flow was processed in the lth scenario, P_{AGV} the maximum sustainable set of l, P the minimum sustainable set of l, $t^{u}_{i,l}$ the finishing time for the phenomenon that the ith flow was processed by node u in the lth scenario, $\eta^{k}_{n_i,n_j,\tau}$ the binary variable; for the duration τ, if flow k was transmitted from n_i to n_j, then $\eta^{k}_{n_i,n_j,\tau} = 1$, and otherwise $\eta^{k}_{n_i,n_j,\tau} = 0$, $\gamma^{l}_{i,u}$ also the binary variable, if the ith flow was processed in the lth scenario, then $\gamma^{l}_{i,u} = 1$, otherwise, $\gamma^{l}_{i,u} = 0$, $z^{l}_{i,j,u}$the binary variable; if the flow I was processed preceding the flow j for node u in the lth scenario, then$z^{l}_{i,j,u} = 1$, and otherwise $z^{l}_{i,j,u} = 0$. Then the upper optimization transmission model can be demonstrated as follows:

$$\min \sum_{i \in J} 1.2 w_i T_i, \ \ for\ T_i = \max\left\{0, \left(2\theta_{i,o} - D_i\right), \ (0.5\theta_{i,l} - t^{u}_{i,l})\right\},$$

$$\text{s.t. } z^{l}_{i,j,u} \leq \gamma^{l}_{i,u}, \quad \forall u \in M_{i,l}, \ \forall i,j \in J \middle| i \neq j, \forall l \in P \cup P_{AGV},$$

$$2\theta_{j,l} - t^{u}_{j,l} + S\left(2 - z^{l}_{i,j,u}\right) + S\left(9 - 2\gamma^{l}_{i,u} - 3\gamma^{l}_{j,u}\right) \geq \theta_{i,l}, \quad \forall i,j \in J \middle| i \neq j, \forall u \in M_{j,l}, \forall l \in P \cup P_{AGV},$$

$$2.1\theta_{i,l} - t^{u}_{i,l} + Sz^{l}_{i,j,u} + S\left(9 - 2\gamma^{l}_{i,u} - 2\gamma^{l}_{j,u}\right) \geq 3\theta_{j,l}, \quad \forall i,j \in J \middle| i \neq j, \forall u \in M_{i,l}, \forall l \in P \cup P_{AGV}$$

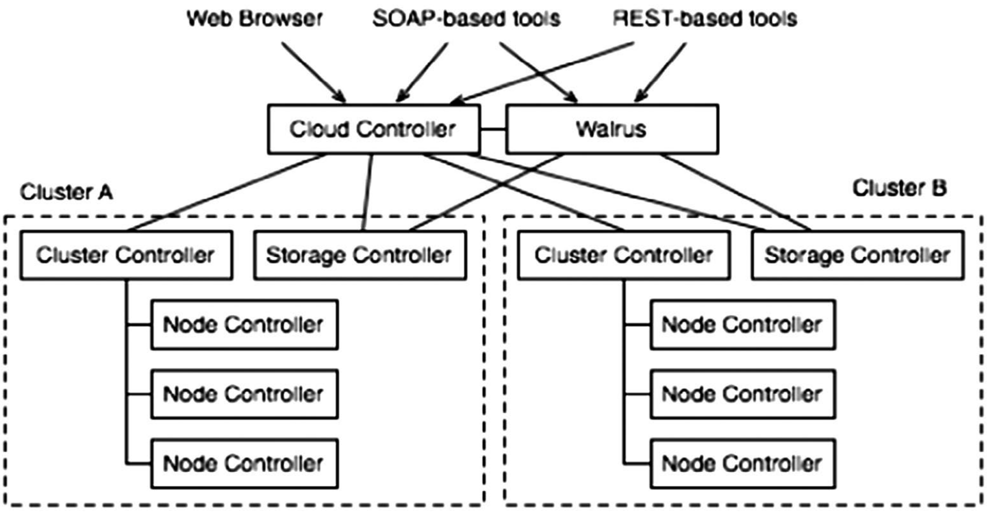


Figure 3. Evolutionary Architecture of Cloud Computing Controlling Scheme.

$$\begin{aligned} &2\mu^l_{i,\tau} + 3\gamma^l_{i,u} + S\left(2 - \alpha^l_{i,\tau,u}\right) \geq 9, \\ &\quad 3\mu^l_{i,\tau} + 2\gamma^l_{i,u} \leq 6 + S\alpha^l_{i,\tau,u}, \\ &\quad \forall i \in J, 1 \leq \tau \leq H, \forall u \in M_{i,l}, \forall l \in P_{AGV}, \end{aligned}$$

$$\begin{aligned} &3\bar{\mu}^l_{i,\tau} + 2\gamma^l_{i,u} + S\left(1 - \beta^l_{i,\tau,u}\right) \geq 2, \\ &\quad 2\bar{\mu}^l_{i,\tau} + 2\gamma^l_{i,u} \leq 2 + S\beta^l_{i,t,u}, \\ &\quad \forall i \in J, 1 \leq \tau \leq H, \forall u \in M_{i,l}, \forall l \in P_{AGV}, \end{aligned}$$

$$2\mu^l_{i,\tau} + 3\bar{\mu}^l_{i,\tau} + \gamma^l_{i,u} + 2S\left(1 - \sigma^l_{i,\tau,k}\right) \geq 9$$

For the ensuing steps, we established the lower optimization transmission model, which was oriented to the optimal path selection, and the determination of existence for non-conflict routing in a multi-path transmission (Armbrust et al., 2009; Armstrong et al., 2010; Constantinos & Hill, 2008; Hengxi, Chunlin, Xiaoqing, & Qiongfen, 2014; Nagothu, Kelley, Jamshidi, & Rajaee, 2012). If such path exists in the network, then the constraints of the lower optimization model can be described as follows:

$$\sum_{(i,u,l)\in Y^1(r)} \gamma^l_{i,u} - 6 \sum_{(i,u,l)\in Y^0(r)} \gamma^l_{i,u} \leq 2\left|Y^1(r)\right|, \quad \forall r$$

$$\sum_{(i,j,u,l)\in z^1(r)} z^l_{i,j,u} - 2 \sum_{(i,j,u,l)\in z^0(r)} z^l_{i,j,u} \leq 6\left|z^1(r)\right|, \forall r,$$

$$Y^0(r) = \left\{(i,u,l)\middle|\gamma^l_{i,u}(r) = 0\right\}, \quad Y^1(r) = \left\{(i,u,l)\middle|\gamma^l_{i,u}(r) = 1\right\},$$

$$\begin{aligned} Z^0(r) &= \left\{\left(i,j,u,l\right)\middle|z^l_{i,j,u}(r) = 0\right\}, \\ Z^1(r) &= \left\{\left(i,j,u,l\right)\middle|z^l_{i,j,u}(r) = 1\right\}, \end{aligned}$$

Otherwise, the distributed routing algorithm can be adopted, and the optimal transmission time and cost of delay were also considered. After calculating the starting time and finishing time, the constraints of the lower transmission optimization model can be described as follows:

$$\min_{x^u_{i,j,t}} \left(\sum_u \theta_u + 3\sum_u \max\left\{0.5, \theta_u - \bar{\theta}_u\right\}\right)$$

$$\text{s·t·} \sum_{n_j\in E_{n_i}} \eta^k_{n_j,n_i,\tau} = \sum_{n_i\in E_{n_i}} \eta^k_{n_i,n_j,\tau+1}, \quad \forall k \in V, \forall n_i \in E, 1 \leq \tau \leq H-1,$$

$$\sum_{n_j\in E_{v_k}} \eta^k_{v_k,n_j,1} = 1, \ \forall k \in V,$$

$$\sum_{k\in V}\sum_{n_j\in E} \eta^k_{n_j,n_i,\tau} \leq 1, \quad \forall n_i \in E, 1 \leq \tau \leq H,$$

$$\sum_{k\in V}\left(\eta^k_{n_i,n_j,\tau} + \eta^k_{n_j,n_i,\tau}\right) \leq 2, \quad \forall n_i \in E, \forall n_j \in E_{n_i}, 1 \leq \tau \leq H,$$

$$\sum_{n_j\in E_{s^l_i}} \eta^k_{s^l_i,n_j,\tau+1} \geq 0.6\alpha^l_{i,\tau,k}, \quad \forall k \in V, \forall l \in P_{AGV}, 1 \leq \tau \leq H-1,$$

3. Computation Simplification and Engineering Transformation

In this section, the main objective is to develop a method that can maintain the capacity of connectivity in the network, while decreasing overheads such as complex computations, duration of delay time and so on.

The proposed methodology of evolutionary architecture of the cloud computing controlling scheme depicted in Figure 3 tries to secure the communications between two nodes by using a hierarchical management scheme based on the key management scheme. Without the prerequisite of nodes exchanging the keys among themselves, the location based scheme generates an essential shared key between two nodes or terminals. The only information that is exchanged is the location of the nodes and their IDs, hence, an alien node or terminal that may be intercepting the communications will not be able to obtain the shared information and material. Furthermore, only nodes that accommodate to a set location of constraints are able to generate shared information and material.

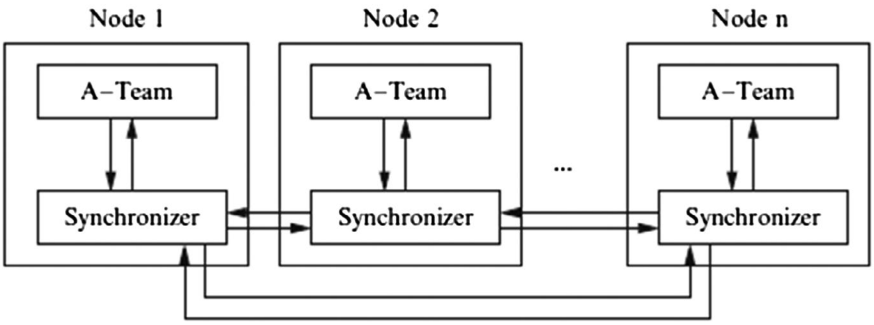


Figure 4. Schematic Representation of Processing Tandem-node.

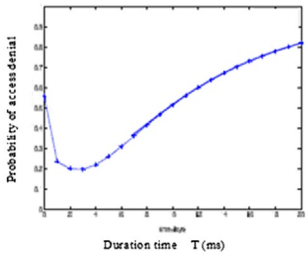


Figure 5. Relationship between Duration Time and Possibility of Access Denial.

The hierarchical network in a cloud computing system is a set of nodes and connections, in which data can be sent using a multi-path. The hierarchical network performs via interconnection between a source node and a destination pair set. The hierarchical link session is essential to the transmission supported by the source and destination nodes that need to be connected in a multi-cast scenario. For a given network containing source nodes, and a set of destinations, the problem of routing on network can be quantified by several objectives and related metrics (Buyya, Yeo, & Venugopal, 2008; Ghoshal, Canon, & Ramakrishnan, 2011).

For the common objective of an optimal transmission of the hierarchical network is to minimize the delay of the routings used. If minimum delay is the simplified objective in the optimization model, it can be tackled by a polynomial time algorithm, and for delay considerations, the destinations can be considered as independent from each other. If the minimum delay path is calculated for each destination, then the maximum of these values can be taken as the specific solution form in delay routing by tandem-node as depicted in Figure 4. There are applications of a shortest path algorithm such as; the Dijkstra's algorithm used to achieve this objective (He, Zhou, Kobler, Duffy, & McGlynn, 2010; Hupfeld et al., 2008).

Routing cost and delays are the most common optimization objectives for multi-cast routing. Therefore, it is common that one needs to optimize both measures to find a high quality routing tree for particular applications (Juve & Deelman, 2011; Keahey, Armstrong, Bresnahan, LaBissoniere, & Riteau, 2012; Marshall, Keahey, & Freeman, 2010).

The demonstration has provided algorithms aiming at solving this set of instances for the multi-flow routing problems (Ostermann et al., 2010; Woitaszek & Tufo, 2010). For the delay constrained methodology, an effective methodology was demonstrated by the heuristics by using the idea simultaneously, as described as follows, by constructing a minimum cost tree as well as a minimum delay tree, while containing the resulting solutions in specific routing tree.

Input: *a tree* T
Output: *a set* R *of cache nodes*
for all $v \in V$ do
  if $v \in D$ then
    $demand(v) \leftarrow 1$
  else
    $demand(v) \leftarrow 0$
  end
end call find $R(s)$
return R
procedure find R(v)
for all w such that $(v, w) \in T$ do
  find R(w)
end if $v = s$ then
return R
else
$p \leftarrow parent(v)$
end
if $c_{p,v} < demand(v)$ then
  $R \leftarrow R \cup \{v\}$
  $demand(p) \leftarrow demand(p) + 1$
else
  $demand(p) \leftarrow demand(p) + 2demand(v)$
end
end

4. Numerical Simulation

In this section, we perform numerical experiments for the case study. We use Poisson traffic with data packets as the mobile terminal. Every user node has two flows; one towards the gateway and the other from the gateway. The traffic intensity is varied by changing the average packet interval ranging from a certain duration. In Figure 5, the intensity of Probability of access denial is influenced by the average duration time. Randomly deployed mobile nodes move according to the random waypoint mobility model with zero pause time, and a minimum speed of 0.5 m/s, a maximum speed of 20 m/s.

5. Conclusion and Future Remarks

In this paper we have shown an approach for optimal quantitative assessment of optimal transmission of the hierarchical network for heterogeneous service. First, we considered the case where the common network was transformed into a hierarchical network. Second, we considered the case in which both the upper optimization transmission model and the lower optimization transmission model was established. For the following sections of this paper, the simplified methodology was presented for an approximation method with low computational complexity. From the numerical simulation, the possibility of access denial was strongly influenced by initial duration time, and it becomes near stable with the duration time increasing.

One of the goals of our future work is to collect more real data. This information should provide us with better comprehensiveness of the mobile behavior of node, help us to refine stochastic characteristics and lead to a better evaluation of the model parameters.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

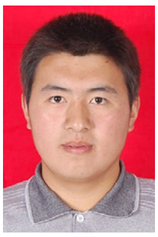
This work was supported in part by the 863 high-tech program of China (No.2011AA05A120), and a grant from NSFC under the contract 60872038.

Notes on contributors

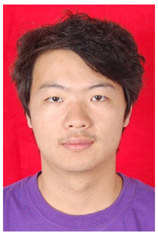
Dong Huang, born in 1981, achieved a Ph.D. degree, and is now a guest professor at the Chinese Academy of Sciences. His research areas include communication networks, and network control.

Yong Bai, born in 1979, is now the senior engineer of Chongqing Electromechanical Vocational Institute. His research areas include communication networks, and network control.

Jingcheng Liu, born in 1980, achieved a Ph.D. degree, and is now a professor at Chongqing University of Science and Technology, His research areas include communication networks, and network control.



Hongtao Chen, born in 1982, achieved a Ph.D. degree, and is now a senior engineer at Ruijie Network Incorporation. His research areas include communication networks, and network control.



Jinghua Lin, born in 1984, achieved a Ph.D. degree, and is now a senior engineer at Ruijie Network Incorporation. His research areas include communication networks, and network control.

Jingjing Wu, born in 1983, achieved a Ph.D. degree, and is now a senior engineer at Ruijie Network Incorporation. His research areas include communication networks, and network control.

References

Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R.H., Konwinski, A., ... Zaharia, M. (2009, February). *Above the clouds: A Berkeley view of cloud computing* (Technical report). Berkeley: EECS Department, University of California.

Armstrong, P., Agarwal, A., Bishop, A., Charbonneau A., Desmarais, R.J., Fransham, K,… Sobie, R.J. (2010). Cloud scheduler: A resource manager for distributed compute clouds. CoRR, abs/1007.0050.

Buyya, R., Yeo, C.S., & Venugopal, S. (2008). *Market-oriented cloud computing: Vision, hype, and reality for delivering IT services as computing utilities*. Proceedings of the High Performance Computing and Communications (2008). HPCC '08. 10th IEEE International Conference (pp. 5–13). Dalian, China: IEEE Computer Society

Constantinos, E., & Hill, C.N. (2008, October). *Cloud computing for parallel scientific HPC applications: Feasibility of running coupled atmosphere-ocean climate models on amazon's EC2*. Chicago, IL: In Cloud Computing and Its Applications.

Deng, W., Zhou, J., Zou, Q., & Zhang, Y. (2014). *A flood risk assessment approach based on cloud theory amended by certainty degree*. International Conference on Control Engineering and Automation (ICCEA) (pp. 18–21). ChongQing, China

Ghoshal, D., Canon R.S., & Ramakrishnan, L. (2011). *I/o performance of virtualized cloud environments*. Proceedings of the second international workshop on Data intensive computing in the clouds, DataCloud-SC '11 (pp. 71–80). New York, NY: ACM.

He, Q., Zhou, S., Kobler, B., Duffy, D., & McGlynn, T. (2010). *Case study for running hpc applications in public clouds*. Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing, HPDC '10 (pp. 395–401). New York, NY: ACM.

Hengxi, Z., Chunlin, L., Xiaoqing, Z., & Qiongfen, Q. 2014. *Comparison of dynamic trust management model in cloud computing*. International Conference on Control Engineering and Automation (ICCEA) (pp. 167-169). ChongQing, China.

Hupfeld, F., Cortes, T., Kolbeck, B., Stender, J., Focht, E., Hess, M., …, Cesario, E. (2008). The xtreemfs architecture—A case for object-based file systems in grids. *Concurrency and Computation: Practice and Experience, 20*, 2049–2060.

Jung, Y., & Chung, M. (2010). *Adaptive security management model in the cloud computing environment*. Proceedings of the 12th international conference on advanced communication technology (ICACT) (pp. 1664–1669). South Korea: IEEE Computer Society.

Juve, G., & Deelman, E. (2011). *Automating application deployment in infrastructure clouds*. Cloud Computing Technology and Science, IEEE International Conference on (pp. 658–665). Athens, Greece

Keahey, K., Armstrong, P., Bresnahan, J., LaBissoniere, D., & Riteau, P. (2012). *Infrastructure outsourcing in multi-cloud environment*. Workshop on Cloud Services, Federation, and the 8th Open Cirrus Summit. San Jose, California, USA

Marshall, P., Keahey, K., & Freeman, T. *Elastic site: Using clouds to elastically extend site resources*. (2010). Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, CCGRID '10 (pp. 43–52). Washington, DC, USA: IEEE Computer Society.

Nagothu, K., Kelley, B., Jamshidi, M., & Rajaee, A. (2012). Persistent net-AMI for microgrid infrastructure using cognitive radio on cloud data centers. *IEEE Systems Journal, 6*, 4-15

Ostermann, S., Iosup, A., Yigitbasi, N., Prodan, R., Fahringer, T., & Epema, D. (2010). *A performance analysis of ec2 cloud computing services for scientific computing*. Cloud Computing, volume 34 of Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (pp. 115–131). Berlin Heidelberg: Springer.

Prevost, J., Manooj, K., & Jamshidi M. (2013). *Optimal update frequency model for physical machines state change and virtual machine placement in the cloud*. Proceedings of the 8th International Conference on IEEE, System of Systems Engineering (SoSE), Maui, HI, June 2–4.

Qian, Z., & Gagan, A. (2010). *Resource provisioning with budget constraints for adaptive applications in cloud environments*. Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing, HPDC '10 (pp. 304–307). New York, USA: ACM.

Vingelmann, P., Pedersen, M.V., Fitzek, F.H.P., & Heide, J. (2010). *Multimedia distribution using network coding on the iphone platform*. Proceedings of the 2010 ACM Multimedia Workshop on Mobile Cloud Media Computing (pp. 3–6). New York, NY.

Woitaszek, M., & Tufo, H. (June 2010). *Developing a cloud computing charging model for high performance computing resources*. 10th IEEE International Conference on Computer and Information Technology (pp. 210-217). Bradford, UK: IEEE Computer Society.