PHYSICAL SERVER AND VIRTUAL SERVER: THE PERFORMANCE TRADE-OFFS

Monjur Ahmed, BSc (Hons), MSc, MS

Daffodil Institute of Information Technology, Bangladesh.

Abstract

Virtualization and virtual computing was one of the core interests of the researchers and the professionals in terms of exploration and deployment of the virtualization technology in IT and computing infrastructure. It was an established fact by then that the blessings of the concept were in the minimization of the tangible infrastructural requirement if virtualization were deployed. But it was always a point of interest and argument whether the performance of the virtual servers would place the virtual computing as a competent substitute compared to its physical server counterpart. Thus the performance tradeoffs between the virtual servers and the physical servers were an issue worth investigating. This paper focuses on this issue where the presentation has addressed on a number of experiments with physical server and virtual server to explore the performance efficiency between them in a comparative way.

Keyword: Virtualization, Virtual Computing, Virtual Servers, Physical Servers, Server Performance

Introduction

The introduction of virtualization had the potentiality to consider computing as well as IT infrastructure with a dynamic viewpoint. In traditional tangible IT infrastructural meaning, it was assumed that having one server dedicated to one sole purpose or functionality would help to load balance the tasks among different servers where a single server was destined to act as the performer of one single task (Shields, 2009). For example, it was a common trend to have two different physical servers for serving the http requests processing and database query processing; one server for each single functionality. This approach, fairly enough, ensured no resource lacking in terms of hardware as well as processor load balancing where it was accepted that the messy and huge pile of hardware had to be a normal consequent of the deployment of servers on a large corporate scale with

different processing purposes. But as time progressed and so did the technology, one of the significant law known as Moore's law had been successfully reflected within the processing capability of the computing devices where the capability was doubled approximately in every 18 months (Huff, 2008). A time came when everyone had desktop computers with the capability that the supercomputers were supposed to have in line with the conceptual and the factual realization of the definition of capacity and capability of desktop computing machines and the supercomputing machines or supercomputers. Once supercomputers-in-terms-of-performance started to be on the home desks in the form of personal computers which subsequently raised one eventual question regarding the aspect explained earlier; that is, if it was still such that one physical server was incapable of handling loads which the earlier servers could not. The fact behind this questioning issue was based on the realization that the processing capability of the computing machines had been growing exponentially when the processing requirements in the environmental setting where the 'supercomputing home computers' were deployed were growing geometrically. This led to the situation where the physical servers were not being used optimally while used as a single-purpose server as the processing requirements were far below than the total capacity of the physical servers. As the so called 'servers' pointed to rather the piece of software that served the purpose instead of the physical computers which had been considered as a server rather than just another computing machine, it was then at least theoretically logical to consider the situation where more than one server software could reside in one physical computer. Thus it was the piece of software which defined what a server was, not the physical computer itself. The 'one computing machine multiple servers' would thus mean having more than one server in one physical computing machine which could then be considered as a computer with multiple servers installed on it (Ali & Meghanathan, 2011). It was also apparent that any server software was essentially an operating system plus something more which was normally the embedded features for which it was called that specific server. For example, in the contemporary era, there had been Linux operating systems only with firewall functionalities which made them known as firewall servers (Wikipedia, 2013). There had been the concept of dual booting where more than one operating system could reside on one physical computer but only one operating system could be operational at any given time, leaving the problem of resource and capacity wastage of the physical servers as a topic of research. Further exploration to find the solution was the introduction of virtualization, also known as virtual computing. By means of virtualization, it was possible to have more than one server on one physical computer that could run in parallel as if they were running on different physical

computers concurrently (Uddin & Rahman, 2011; Kizza, 2012). This ensured further and optimal usage of the physical server resources and capacity as well as eliminating the need for a massive collection of hardware and associated infrastructural needs. It was counted that by means of virtualization, the cost minimization could be achieved to a significant level (Singh & Jangwal, 2012). This indicated the significance of having successful virtualization in the field of IT and computing infrastructures. It was important to realize the performance of different servers in a virtualized environment compared to their tangible or physical counterpart for the reason that the main goal of virtualization could not be the cost reduction at all if performance had to be sacrificed. In some real time situations, the performance were most crucial than the expenditure and thus the success of the virtualization was largely dependent on its ability to compete with its physical counterpart. This is the main point of concentration in this paper. The study that has been presented within this paper had a goal to determine the performance of the servers with the same functionality in virtual and physical environment to compare and subsequently finding whether the virtualized servers were worthy or not in terms performance when compared to physical servers.

Related work

A significant level of studies had been carried out in the field of virtualization. Different aspects had also been addressed where one of the core interests was the performance trade-offs between physical and virtual server. Some studies had found a significant level of fluctuation between the performance of virtual servers and the physical servers where the physical servers were 50% to 100% efficient in performance than that of the virtual servers (Jung, Bae & Soh, 2011). It was also found that the virtualization performance was dependent on the underlying virtualization technique or architecture as a different virtualization platform exhibited different levels of efficiency and performance level. It was also argued that the performance of the virtual servers was largely dependent on the total number of virtual operating systems that were being used on one single virtualized platform (Jung, Bae & Soh, 2011; Ali & Meghanathan, 2011; Prakash, Anala & Shobha, 2011). The management of the data centers may become efficient but the associated performance issues of virtualization needed to be taken into serious consideration as the SLAs for the business were mandatory to maintain; this requirement was from the fact that it was the virtual servers that were under the question of performance issues where physical servers were established as the reference and benchmarking point in terms of performance (Wood, et. al., 2009). It was also found that, if planned and implemented properly, the virtualization approach and thus the virtual servers could potentially help the computing

infrastructures to facilitate innovative computing like cloud computing without compromising the performance and thus the virtualization techniques were becoming more and more popular in data centers and in the field of cloud computing (Praveen & Vijayrajan, 2011; Singh & Jangwal, 2012; Berl et. al., 2010; Bento & Bento, 2011). There had been a number of different approaches by which the deployment of server virtualization could be achieved; it was something related to the performance of the total virtualized infrastructure including the cloud based infrastructures (Wood et. al., 2009; Kumar & Petal, 2012). The performance of the virtualized environment had been directly related to a number of factors including the total number of virtual appliances occupied within the virtual environment as well as the processing capability of the virtual environment (Bhukya & Ramachandram, 2009). Server consolidation in terms of virtualized servers could emerge for the data centers as a blessing if the performance tuning were given a priority while implementing the virtual infrastructure (Uddin & Rahman, 2010). The choice of hardware for the virtual infrastructure had always been a core issue as the performance degradation of the virtual computing had been observed in the scenarios of high performance networks where low latency and high level of throughput were crucial (Ali & Meghanathan, 2011). When the specific factor of performance was taken into account, a little deviation of the virtual servers compared to their physical counterparts were quite acceptable from the viewpoint that by means of virtualization - which subsequently facilitated cloud computing - the availability of the servers and thus the network availability were increased to a significant level to make the data centers more robust (Foster, Kesselman & Tuecke, 2001; Singh & Jangwal, 2012; Berl et. al., 2010; Kumar & Petal, 2012). From performance point-of-view, the successful implementation of virtualization infrastructure depended on a number of factors among which the most crucial were choosing the appropriate approach and severs for virtualization (Uddin & Rahman, 2011). As cloud computing was becoming an unavoidable part of the total computing infrastructure which in turn was facilitated by means of virtualization, the performance of the virtual servers was a definite concern where load balancing had been considered to be one of the core processing aspects (Bhaskar, Deepu & Shylaja, 2012; Barham et. al., 2003; Sapuntzakis et. al., 2002). One of the most significant aspect while streamlining the virtualization platform was the security aspects of the virtual servers; the approach to deploy virtualization infrastructure were needed to be considered not only from performance point-of-view, but also from the security viewpoint as it had been another burning question for virtual servers along with performance (Solms, Chaudhuri & Chaudhuri, 2011). When implemented for cloud computing, there had been approached to load balance

the virtual processing among a number of servers which helped the virtualization to gain a performance level to supersede its physical counterparts (Kizza, 2012). The performances of the virtualized machines, according to some studies, were not inferior at all compared to the physical servers (Clark et. al., 2005; Casazza, Grienfield & Shi, 2006; Kumar & Petal, 2012). One of the distinguished features of virtualization was that, while it was not inferior in performance compared to physical servers – it was more environment friendly than the physical servers (Dawson, 2008) for which the adaptation of virtualization technology was an expected one.

Virtual and Physical server performance

The study that had been conducted to determine the performance tradeoffs between physical and virtual server was associated with setting up environment with virtual server as well as physical servers. As the purpose was to measure the performance of the virtual server as well as the physical server, the similar services were 'played' on both the environment to determine the performance in each environment. The physical server and the virtual server environment were created to run applications like DNS server, mail Server, database server and web server. While all the servers were implemented on a single virtual platform, there were different physical servers for each of the aforementioned server functionality for the experiment. The same configuration of the physical server had been used so that the exact comparable load and performance could be determined which was the main purpose of the study. The following table enlists the hardware and software that were used in the experimental environment:

Server (Hardware) Model	HP ProLiant ML350
Server Operating System	Windows Server 2008 Enterprise Edition
Virtual Platform	Hyper-V

Table – 1: Hardware and Software for experimental environment

First of all, the response time of the database server on the physical server and the virtual server was measured over a continuous trending time period. The database server response trend on the physical server and the virtual server is shown in the following graph:

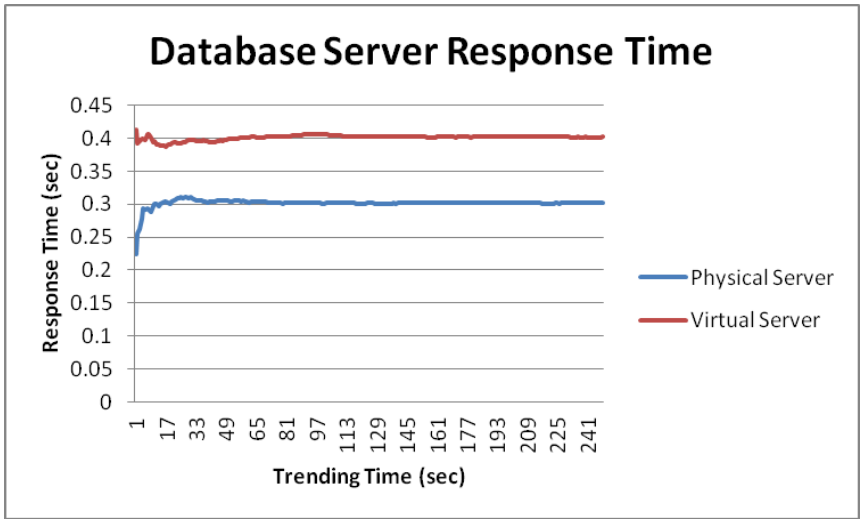

Figure – 1: Database server response time for physical and virtual server

The above illustration revealed that the response time from the virtual server was higher than the physical server at any given time. Next, the traffic load on the Ethernet needed to be determined for each case of the virtualization environment and the physical server environment to determine the level of optimal resource utilization and to ascertain whether any of the testing environments was associated with overload issues. The depiction in figure 2 shows that the capacity of the Ethernet could be more optimally utilized in a virtualized environment or in other words, the virtualized environment produced more traffic in terms of Ethernet load than that of the physical server:

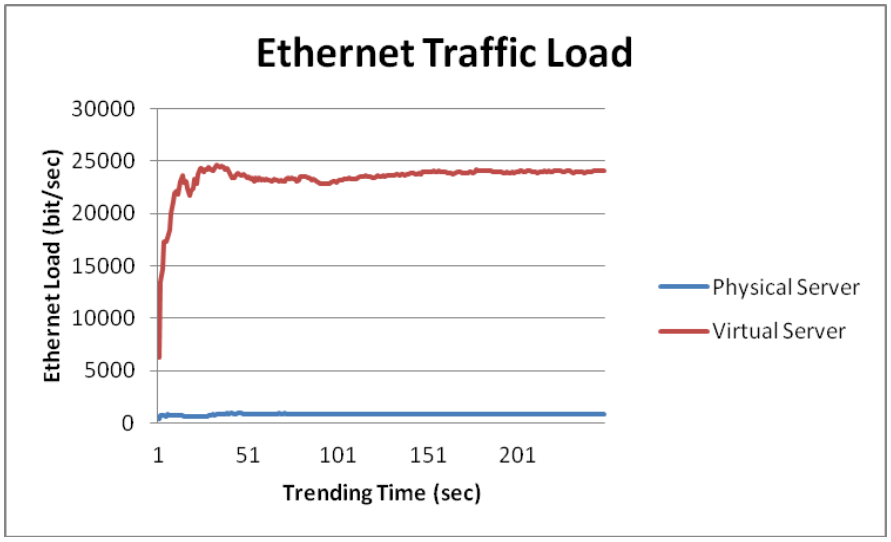

Figure – 2: Ethernet traffic load

The response time of the FTP server was another point of interest to determine how the physical server and the virtualized server handled the ftp requests. This is presented in the following figure:

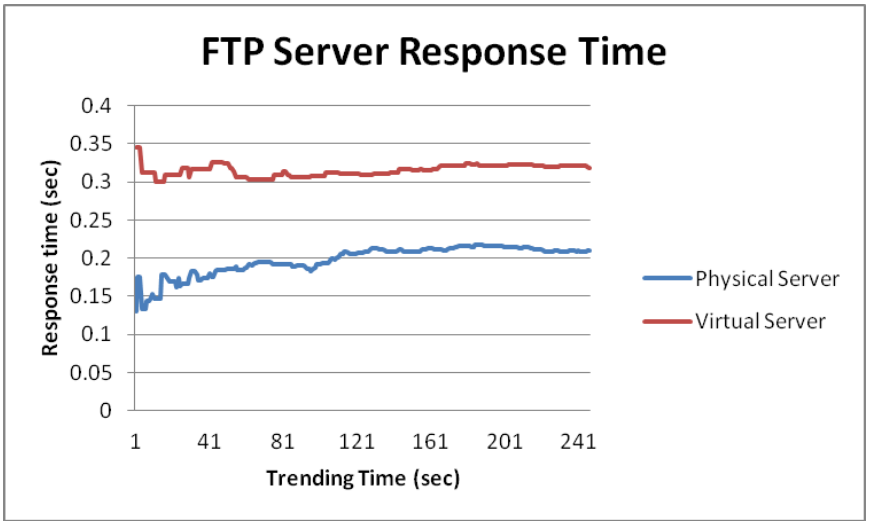

Figure – 3: FTP server response time

As seen on the above figure, the response time for the virtual server had been constantly higher than the physical server. It was part of the experiment to observe the http request responses of the servers respectively. The factual representation in figure 4 demonstrated that the response time for the http requests were also higher for the virtual server.

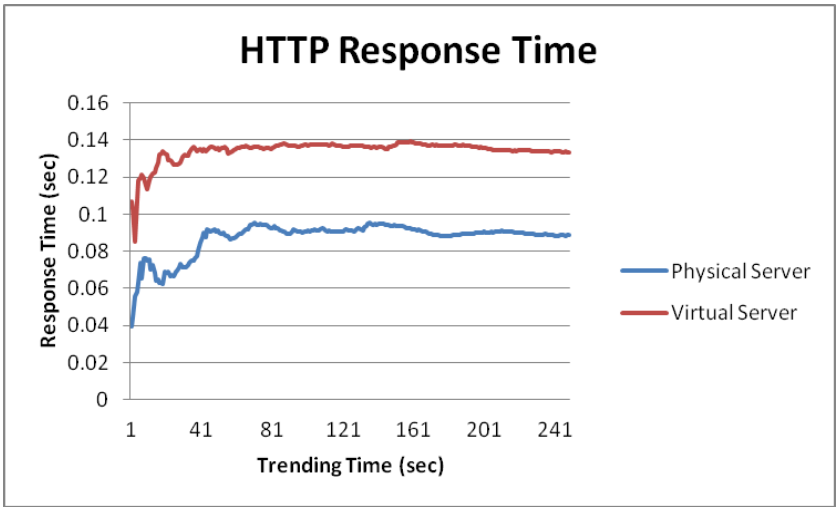

Figure – 4: HTTP response time

As performance was related to the usage of the resources specifically the CPU load, both the servers were tested to find out the total percentage of CPU usage in idle state when no traffic was generated or no additional user-initiated tasks had been carried out. The depiction in figure 5 revealed that the virtual server demanded more CPU processing cycle in the idle state compared to its physical counterpart.

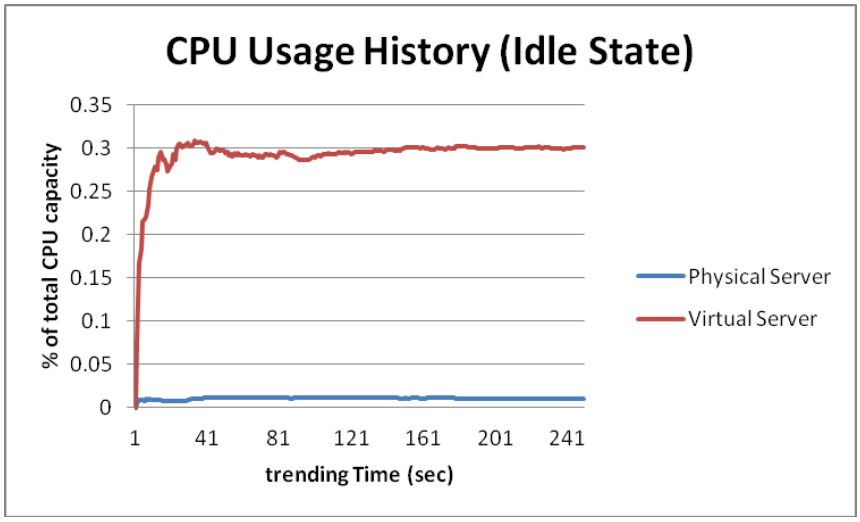

Figure – 5: CPU Usage History (Idle State)

Analysis of results

The conducted study gave an insight into the performance trade-offs when compared between physical and virtual server. To specifically start with the database server, the performance degradation was observed in the virtual server in the form of higher response time. The same was equally applicable to the ftp and http server response time. The higher response time of the virtual servers indicated the significance of the fact that the deployment of the virtual servers and the virtualized infrastructure needed to be carefully planned where the applicability of the virtual servers was to be scrutinized on a case-by-case basis as in all situations the deployment of virtualized infrastructure could not be feasible.

If the optimum resource utilization were taken as a factor, the virtual servers would be in the winning situation provided that the total demand in terms of resource and capacity had been correctly outlined prior to the virtualization infrastructure implementation. This important aspect was also supported by the CPU capacity and processing load demanded by the physical and the virtual servers where the virtual servers always demanded more resources compared to those of the physical servers. This aspect was quite acceptable as the load on the virtual server could be several times depending on the scenario where they were deployed. However, the performance degradation with the specific phenomenon of higher response time of the virtual servers compared to that of the physical servers was evident from the conducted study which questioned the credibility of deploying virtualized servers and infrastructures in the places where real-time and very low-latency processing were of utmost priority; the safety critical computing and operating environment could be an example of such scenario in this regard.

Conclusion

The literature review revealed the fact that any isolated experiment cannot be taken as the performance benchmark of virtual server performance from a holistic point-of-view. This indicated to the facts that any experiments cannot be relied upon in an isolated manner to determine the performance scale of a virtualized environment; rather, a number of different studies and their outcome could potentially indicate the trend of performance of the virtual servers compared to that of the physical servers. From this perspective, the conducted study was to add another feather to the relevant hat. As found in the study, the performances of the virtual servers in some cases were slightly deviated than the physical servers. But considering the resource usage, the little deviations were acceptable as the virtualized environment had a very low rate of wastage of the server resources compared to the physical server environment. It might be well argued whether virtualization could supersede the performance of the physical servers, but considering the total aspects including deployment, cloud computing and resource utilization – the virtualized infrastructure was the choice over the physical server deployment where research on improving virtual server performance was in demand.

References:

Ali, I. & Meghanathan, N. (2011), Virtual Machines And Networks – Installation, Performance, Study, Advantages and Virtualization Options, *International Journal of Network Security & Its Applications (IJNSA)*, Vol.3, No.1, pp. 1- 15

Barham, P., Dragovic, B., Fraser, K., Hand, S., Harris, T., Ho, A., Neugebauer, R., Pratt, I. & Warfield, A. (2003), Xen and the Art of Virtualization, Proceedings of the 19th ACM symposium on Operating Systems Principles, Bolton Landing, NY, USA, pp 164-177

Bento, A. & Bento, R. (2011), Cloud Computing: A New Phase in Information Technology Management, Journal of Information Technology Management, Vol. XXII, no. 1, pp. 39-46

Berl, A. et. al., (2010), Energy-Efficient Cloud Computing, *The Computer Journal*, vol. 53 no. 7, pp. 1045-1051

Bhaskar. R, Deepu. S.R, & Shylaja, D. B. S. (2012), *Dynamic Allocation Method for Efficient Load Balancing in Virtual Machines for Cloud Computing Environment*, Advanced Computing: An International Journal (ACIJ), Vol.3, No.5, pp. 53-61

Bhukya, D. P. & Ramachandram, P. S. (2009), Performance Evaluation of Virtualization and Non Virtualization on Different Workloads using DOE Methodology, *IACSIT International Journal of Engineering and Technology*, Vol.1, No.5, pp. 404-408

Casazza, J. P., Greenfield, M. & Shi, K. (2006), Redefining Server Performance Characterization for Virtual Benchmarking, *Intel Technology Journal*, Vol. 10, Issue 3, pp. 243-252

Clark, C., Fraser, K., Hand, S., Hansen, J. G., Jul, E., Limpach, C., Pratt, I., & Warfield, A. (2005), *NSDI '05: 2nd Symposium on Networked Systems Design & Implementation*, pp. 273-286

Dawson, C. (2008), The Environmental Benefits of Virtualization, *Virtualization*, Vol. 1, issue 3, pp. 20-22

Foster, I., Kesselman, C. & Tuecke, S. (2001), The Anatomy of the Grid: Enabling Scalable Virtual Organizations, *International Journal of High Performance Computing Applications*, Vol. 15, No. 3, pp. 200-222

Huff, H. (2008), *Into the Nano Era: Moore's Law Beyond Planner Silicon CMOS*, Springer, pp. 11-20

Jung, S., Bae, Y., & Soh, W. (2011), Web Performance Analysis of Open Source Server Virtualization Techniques, *International Journal of Multimedia and Ubiquitous Engineering*, Vol. 6, No. 4, pp. 45-51

Kizza, Joseph. M., (2012), Africa Can Greatly Benefit from Virtualization Technology – Part I., *International Journal of Computing and ICT Research*, Vol. 6, Issue 1, pp 6-10

Kumar, K. & Petal, A. K. (2012), Innovation Idea of Virtualization in Cloud Computing, *International Journal of Research Review in Engineering Science and Technology*, vol. 1, issue 2, pp. 92-95

Prakash H R, Anala M.R. & Shobha, Dr. G, (2011), Performance Analysis Of Transport Protocol During Live Migration Of Virtual Machines, Indian Journal of Computer Science and Engineering, Vol. 2, No. 5, pp. 715-722

Praveen, G. & Vijayrajan, P. (2011), Analysis of Performance in the Virtual Machines Environment, *International Journal of Advanced Science and Technology*, Vol. 32, pp. 53-64

Sapuntzakis, C. P., Chandra, R., Pfaff, B., Chow, J., Lam, M. S. & Rosenblum, M. (2002), Optimizing the Migration of Virtual Computers, *Operating System Design and Implementation*, pp. 377 – 390

Singh, S. & Jangwal, T. (2012), Cost breakdown of Public Cloud Computing and Private Cloud Computing and Security Issues, *International Journal of Computer Science & Information Technology*, vol 4, no 2, pp. 17-31

Shields, G. (2009), *The Shortcut Guide to Implementing Virtualization in the Small Environment*, Relatime Publishers, pp. 2-7

Solms, S. H. V., Chaudhuri, A. & Chaudhuri, D. (2011), Auditing Security Risks in Virtual IT Systems, *ISACA Journal*, Volume 1, pp. 1-10

Uddin, M. & Rahman, A. A. (2011), Virtualization Implementation Model for Cost Effective & Efficient Data Centers, *International Journal of Advanced Computer Science and Applications*, Vol. 2, No.1, pp. 69-74

Uddin, M. & Rahman, A. A. (2010), Server Consolidation: An Approach to Make Data Centers Energy Efficient & Green, *International Journal of Scientific & Engineering Research*, Vol. 1, Issue 1, pp. 1-7

Wikipedia, 2013. *List of Router or Firewall Distributions.* [Online] Available at <https://en.wikipedia.org/wiki/List_of_router_or_firewall_distributions> [Accessed 28 February 2013]

Wood T. et al., (2009), Sandpiper: Black-box and gray-box resource management for virtual machines, Computer Networks, doi:10.1016/j.comnet.2009.04.014