Fingerprint Verification using Cloud Services with Message Passing Interface over PC Clusters

Fazal Noor, Majed Alhaisoni*, Antonio Liotta+*
*Computer Science and Software Engineering Department
University of Hail, Saudi Arabia
+Department of Electrical Engineering and
Department of Mathematics and Computer Science
Eindhoven University of Technology, The Netherlands
f.noor@uoh.edu.sa, majed.alhaisoni@gmail.com, aliotta@tue.nl

***Abstract*—Nowadays cloud-computing services are being offered by various organizations. Peer-to-Peer (P2P) networks can be used as a collaborative computing environment to solve computationally intensive problems. In this work, we use a PC cluster to simulate a P2P network and present results of a computationally intensive image matching algorithm (fingerprint verification). Collective communications are used to transfer images to destination peers over a network. Communication to computation time ratio are calculated of transferring of fingerprint images of various sizes on the internet. As transfer of raw images are communication intensive, a proposed method is to use FBI approved Wavelet Scalar Quantization (WSQ) compression method at the source before transmitting to the destination nodes. We study the viability of fingerprint identification and/or verification service offered by cloud computing. In particular, we present a distributed fingerprint verification algorithm.**

Keywords-PC Cluster; Normalized Correlation Method; Minutia Method; Cloud Services; Latency; Bandwidth; Message Passing Interface; Phase Correlation method; Log-Polar; communication to computation ratio.

I. INTRODUCTION

Cloud-computing services are becoming common nowadays [1],[2]. There are many types of services being offered to customers of cloud-computing. As the demand for cloud computing grows, different types of applications appear, which require intensive computing power. In such cases, PC clusters are more affordable and a cheaper alternative for buying supercomputers, which are very costly.

Authentication of a person is required to access places of high security and can be done in several ways, verification by knowledge such as passwords, verification by possession such as id cards or passports, or verification by biometrics such as fingerprints [3],[4]. Authentication can be done locally, such as access to restricted areas, Personal Digital Assistants (PDAs), computers, etc. Here we look at authentication done remotely via cloud computing and therefore security issues which arise are of major concern, especially in keeping the data secure. Remote verification is necessary when the original fingerprint is stored at a remote site. One such application is verification by possession and by biometrics (fingerprints), which may be required at airports, border points, checkpoints, etc. All these check points require verification by possession and for further confirmation verification by biometrics may be done with the use of mobile wireless devices. Travel safety is a major concern not only for the governments but also for any person for example boarding a plane. Security has been a great issue at airports due to terrorist activities. Currently, verification is by possession of a valid passport. Consider the following scenario at an airport, verification of passengers' identities by fingerprint biometrics also. Each passenger goes through a checkpoint and his/her fingerprint is scanned by an ultrasonic scanning device. The fingerprint(s) are sent with a tag (person's identity number and country code for fast look up) for verification at data centres which may be distributed around the world. There the tag is used to retrieve from database the person's fingerprint and verification made. Since a passenger at airport A in country A may be a citizen of another country, say B, and therefore normally his enrolled fingerprint would be enrolled in the database located in his/her country B. Every country has laws protecting the privacy and security of their citizens. The scanned fingerprint image has to be transmitted through the internet to the country to verify the identity of a passenger. A database of fingerprints maybe in hundreds of millions or billions depending on the population of the country. The fingerprint image size *NxM* varies depending on the device used. For a size of 768 x 768 at 500 dpi, assuming 8 bits grayscale image, the amount of storage would be 589,824 bytes [7]. Therefore, database size would require a total of image size x population size of storage bytes.

 ISBN: 978-1-61208-238-7

The problem is to authenticate the person based on his fingerprint scanned and sent for verification by cloud-computing services. One question is how much time would it take for the verification result to arrive back (i.e., the time to communicate fingerprint image plus person's data and time to process for verification). The time then depends on part where the databases are located. Ideally one location containing all countries fingerprints is preferred, but, in reality, due to various reasons, such as privacy issues and security concerns would be located at multiple locations. Another question is how reliable and accurate is the verification. Can we trust the result which comes from another country? Assuming each country would have their own database, therefore *J* scanned fingerprints would have to be transmitted via cloud-services points to *K* countries for verification of an identity of a person as in Figure 1. How accurate is the verification result? Fingerprints are considered sensitive information by any government and should not fall into the wrong hands. Another scenario is the person carries a biometric card which holds the person's fingerprint already enrolled. Then, both the scanned fingerprint image and the one on the biometric card is sent to a cloud service point for verification and result communicated back as "pass/fail". Reliable service is one of the key goals. This paper focuses on two key issues *Speed* and *Accuracy* of fingerprint verification. For *Speed*, communication bound and compute bound problems of fingerprint image(s) identification or verification at distributed databases are explored. The main contribution in this paper is presenting a distributed algorithm based on correlation method using 9 patches for high accuracy of fingerprint verification.

The paper is organized as follows: in Section II, some related work is presented. In Section III, fingerprint background is presented, and, in Section IV, methods are presented. Section V shows the results, Section VI presents the discussion, and in Section VII, conclusion and future direction are given.

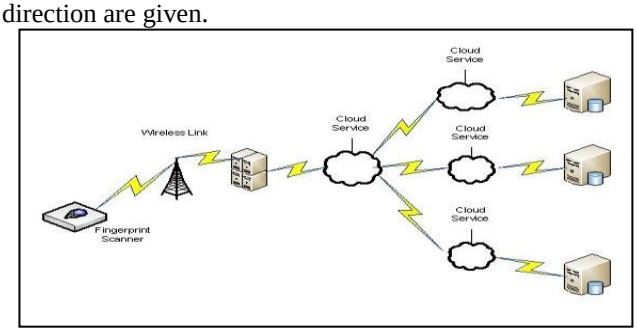

Figure 1. Fingerprint verification through peer Cloud Services to peer Cloud Services located around the world.

II. RELATED WORK

Work on fingerprint identification and verification is vast in the literature. There are many algorithms appearing in the literature mainly based on minutia properties [3]-[7]. In this work, we mainly use a correlation method for verification for several reasons. One reason is that we store the original fingerprint image (using Wavelet Scalar Quantization compression technique, a standard set by FBI) [7] and not only the extracted features as done by methods based on the minutias. Many algorithms in literature are fast and have high reliability, but are not 100% accurate [3] [4]. Researchers have also proposed combination of biometric methods, for example fingerprint plus voice recognition [6]. Government of South Africa is allowing banks to have access to fingerprint database [8]. Chang et al. [9] have implemented a real-time video/voice over IP (VVoIP) applicatons on a Hadoop cloud computing system [15]. For access control, to prevent illegal intrusions, they have used facial recognition and fingerprint identification via cloud computing. It takes about 2.2 seconds to exactly identify the subject [9]. In our work, we present a distributed fingerprint verification algorithm and the fingerprint database are not local to cloud-computing services but within the domain of each country.

III. FINGERPRINT BACKGROUND

Fingerprint-based identification is one of the oldest biometric technique [6]. A fingerprint consists of three-dimensional lines called ridges and the spaces between them are called valleys. Fingerprint identification is different from fingerprint verification. In identification, the question is to answer whose fingerprint is this. In verification, the question is, are you who you claim to be. In fingerprint identification, a large database has to be searched and match is of a form 1:N, whereas in verification the original image is to be matched with the live scan image and the match is of 1:1 form. Time required for identification is much larger than the time for verification.

As raw fingerprint image storage demands large amount of memory space; usually, images are not stored, but their properties are stored such as minutia type, etc. Here Wavelet Scalar Quantization (WSQ) [7], a compression technique developed by the FBI is used to compress images for transmittance or storage. WSQ has a better preservation of fine details over other compression techniques. A fingerprint image of 589284 bytes is compressed to WSQ image of size 45621 bytes, a compression ratio of 12.9 [7].

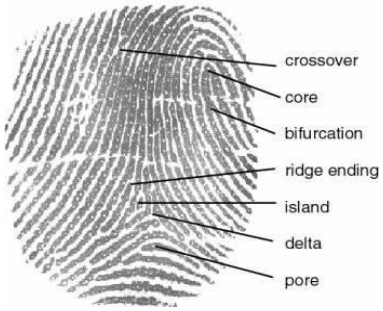

Figure 2. Fingerprint showing various features labeled [6].

 ISBN: 978-1-61208-238-7

IV. METHODS

A. Fingerprint Methods

Algorithms use various techniques at different stages for example, during preprocessing (using segmenatation, or enhancement), during alignment (before matching, during matching, displacement, scale, rotation), during feature extraction (e.g., minutia, singular points, ridges, counts, orientation field, texture measures, raw/enhanced image parts), and during comparisons (based on minutiae global, local, ridge pattern geometry or texture, and correlation) [3]. All categories fall under estimation and approximation theory and the main ones are summarized below:

Minutiae Based Method [3]

1. Image acquisition or capture by a device. Objective is to capture image of the center of the finger as this part contains unique features. There are different technologies available in the market. The three main are optical, silicon, and ultrasound. Ultrasound is better than the other two.

2. Extracting unique characteristics of the fingerprint and their locations. A fingerprint consists of various ridges and valleys and formation of loops, arches, and swirls. Minutiaes are extracted which are of mainly two type a. ridge endings and b. bifurcations.

3. Creation of minutiae template: Type, location, position, quality, direction of ridges, etc.

4. Template matching between enrollment template with the verification template.

Correlation Based Method [6]

The mean-square difference is defined as

$$E(p,q) = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1}[G(i,j) - H(i+p, j+q)]^2 \quad , \qquad (1)$$

where $G(i,j)$ is the feature image and $H(i,j)$ is the original image. For an ideal case, where there is no noise in the images, an exact match will make E zero for all possible points. In practice, E would not be zero due to noise in the original image and the feature image. Therefore, a decision is made on a comparison with a threshold level $T(p,q)$. If $E < T$, then the decision is said to be a match has occurred otherwise the decision is no match. In practice, the Normalized Cross-Correlation (NCC) given below [6],

$$R(p,q) = \frac{\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} G(i,j)H(i+p, j+q)}{\left[\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} G^2(i,j)\right]^{1/2}\left[\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} H^2(i+p, j+q)\right]^{.1/2}}, \quad (2)$$

is used. When evaluating algorithms the maximum limits on algorithms are imposed such as time limits of enrollment and comparison and size limits on template and memory. Typical values maximum limits used in fingerprint verification competition are shown in Table 1 [3].

TABLE I. FINGERPRINT VERIFICATION COMPETITION LIMITS.

Maximum Limits on Algorithms				
	Enrollment seconds	Comparison seconds	Template KBytes	Memory MBytes
Open Category	10	5	No limit	No limit
Light Category	.5	.3	2	4

The minimum and maximum values depend on type of databases; see [3], for further details.

B. Distributed Fingerprint Verification Algorithm

Most algorithms are of minutia type compared to correlation type. Here we chose to implement algorithm based on correlation method as it does not require much preprocessing as compared to minutia type. With preprocessing in minutia type algorithms there is a possibility of false registrations of minutias due to poor image quality of fingerprints.

A cloud-service point may use a PC cluster to speed up a fingerprint verification. The steps of an algorithm based on correlation method are outlined as follows:

1. WSQ compression is performed on raw fingerprint image, tagged, and sent to cloud services point.
2. The tag is used to find the original image from the database.
3. A procedure is used to extract 9 feature patches of size ***N x N*** pixels from the original image and their locations are marked and stored. The 9 patches are selected toward the center of fingerprint image and equal distance apart. Core or delta points would be preferable to be included in one of the patch.
4. Scatter ***P=9*** feature patches and their locations to each node in a PC cluster. Assume the original image is accessible by all nodes in a PC cluster.
5. Each node tries to detect the patch on the scanned fingerprint image by calculating the normalized cross-correlation equation (2) and sends the result back to master node.

 ISBN: 978-1-61208-238-7

6. If majority of patches match with those in the scanned image, the scanned image is super-imposed on the original image. If there is mis-alignment, further techniques, such as Phase Correlation Method combined with log-polar method (translation and rotation) are used to align the image. The normalized correlation value is compared with a threshold *T*, and if it is greater than or equal to 0.9, is considered a match. The neighboring patches may also be checked to achieve even greater accuracy.
7. If majority of patches do not match, declare a non-match.
8. Send the result "pass" or "fail" back to source.

There are two types of errors that may occur in matching fingerprints, one is named a False Match rate FMR (False Acceptance) and the other False Non-Match FNMR (False Rejection). A trade-off between the two errors exists, depending on the threshold *T*. The point where FMR equals FNMR is called the Equal Error Rate.

V. RESULTS

A. Beowulf PC cluster Specifications

Our test-bed consists of a PC cluster including 20 Lenovo machines with the following specifications: Intel Core™ 2 Duo CPU, E4400 2.00GHz, 1.00 GB of RAM. Network Card: Broadcom Netlink, Gigabit Ethernet, Driver date 8/28/2006 version 9.81.0.0. The PCs are connected to a Gigabit D-Link Ethernet switch. Each machine has a RedHat Enterprise AS Linux operating system installed, and uses LAM 7.0.6/MPI 2.

B. Model

A PC cluster of size 20 nodes is used to model 20 airports. Each airport is located in a different country. Also assume each node in a cluster which represents an airport is also a processing centre (cloud service point). The cloud service point is assumed to have access to fingerprint databases around the world. The link between any two nodes may contain a number of routers with different latencies and bandwidths.

C. Experimental result

Here, we present results of our algorithm using 10 nodes in a PC cluster. Given an original fingerprint of size *N x M*, partition the image into *B* blocks each of size *N/patch size*. For example, if *N*=512, then for patch size of 16 pixels there are *B*=32 blocks, choose 9 patches located approximately in the center, for example, blocks no. (12,12), (12,15), (12,18), (15,12), (15,15), (15,18), (18,12), (18,15), and (18,18). Other ways can be used to choose the patches, for example, choosing a patch at the core or delta. The 8 others chosen equidistant from the center patch enclosing the core or delta point. Also, the number of patches selected is arbitrary, however we choose 9 to achieve greater accuracy. The master node uses Message Passing Interface (MPI) to scatter the scanned fingerprint and the 9 patches to the 9 nodes in a cluster. Each node searches in the neighborhood of each block by sliding the patch pixel by pixel to find maximum normalized correlation. The database of fingerprints consisted of 20 people with 3 prints of the same finger for a total of 60 images. Figure 3 shows a sample of two prints of the same finger with the second print clipped on the top. Figure 4 shows the 9 patches selected of size 32x32 pixels and Figure 5 shows a sample of normalized correlation. We experimented with different patch sizes of 16x16, 32x32, and 64x64 pixels. Patch sizes of 64x64 pixels exhibited sharpest peak in comparison to peaks obtained by using patches of sizes 16x16 and 32x32 pixels. The FAR for threshold values of *T* =0.85 was 0.02% for the patch size of 32x32. For patch size of 16x16, FAR starts to occur when *T* was set to a value of 0.8. For patches of size of 32x32 and 64x64, FAR starts to occur when *T* was set to a value of 0.7. The FRR was 11.3%, which is due to images having rotation. It is observed 9 patches are more than sufficient to discriminate an impostor from the genuine fingerprint.

Figure 3. Two fingerprint impressions from the same finger.

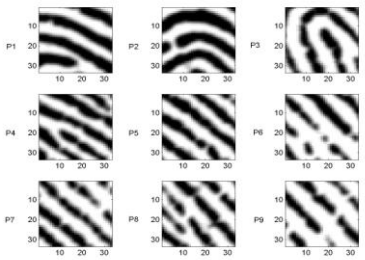

Figure 4. Nine patches of size 32 x 32 pixels.

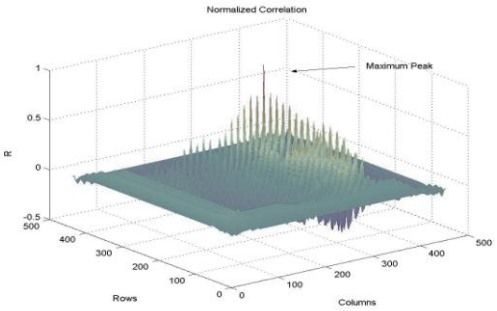

Figure 5. Sample maximum peak obtained by normalized cross-correlation.

D. Communication and Computation Times

The waiting time of the i^{th} passenger will be the time to scan his/her finger, the time to transmit the fingerprint image with a tag (data such as civil card id number and country code), the time to perform the verification through a cloud service point and the time to receive the result,

$$Time_i = t_{scanA} + t_{sd} + t_{verification} + t_{ds}, \quad (3)$$

where t_{scanA} is the time to scan the fingerprint by a device, t_{sd} is the time taken to transmit the image plus the tag from source A to destination B, $t_{verification}$ time taken by the peer (PC Cluster or Super Computer to perform the computation of verification and depends on the hardware and algorithms used), and t_{ds} the time to receive the result as "Pass" or "Fail". If all passengers are local then local database is accessed, if passengers are international then of course distributed database. Can the result be obtained in real-time? From table 2 the scan time dominates both the communication and compute time. It is important to know how much computation time is in comparison with communication time. There are numerous fingerprint verifications algorithms and increasing in numbers [3], each algorithm having its own speed and accuracy. Communication time would depend on the network links, routers in between, etc. Queue delays at cloud service points. Latency will vary with each path as paths change. Therefore, each path would have a different time. Computation time is function of both algorithm and hardware used. If T_s = time to execute on a single computer then using a PC cluster the time to execute would be approximately T_s / n , where n is the number of computers in a PC cluster. Table 2 shows time to process fingerprint verification by algorithms based on minutia method and correlation method. Correlation method is more compute intensive than minutia method.

TABLE II. TIME TO PROCESS FINGERPRINT VERIFICATION

Algorithms based on	*Approximate Time in seconds*				
	Scan	***Send X_s***	***Process C_s***	***Receive Y_r***	***Total***
Minutia	5 sec	.5	.3	.4	6.2
Correlation	5 sec	.5	1	.4	6.9

VI. DISCUSSION

Fingerprint recognition is one of the oldest and most popular biometric technologies being used in commercial applications. A consensus among countries has to be made on setting standards and how fingerprint databases are to be accessed by cloud-computing service points. Highly secure encryption methods for protection of fingerprint data is a must. Another point is where to store fingerprint databases of any country, so it is securely accessed. Should it be stored at cloud-service premises or in country of origin? This is to protect from any fraudulent use. If not on cloud service premises, the cloud-services points would have to have contracts with each country either to access their fingerprint database or with the countries cloud-computing service. In other words, peer cloud-computing to peer cloud-computing services [10] [11]. The services should be highly reliable, secure, and results with ideal accuracy of 100%.

Compute bound problems: Each country may be using different algorithms of extracting fingerprint features and storing fingerprint data. Therefore, different algorithms with various execution times depending on complexity and type of hardware. Here, the goal would be reduction of search time and low computational complexity with high accuracy of verification.

Communication bound problems: There are many, such as encryption and compression of raw images, and secure transmittance over the internet. A universal tag has to be agreed among countries for easy retrieval of fingerprint data from a large database. In networks, data is not sent in a continuous stream, but in packets. Fingerprint image sizes *N x M* vary according to various devices used. Usually, *N>M* with values anywhere from 256 to 768. Therefore, transmission can be anywhere from 64KB and up. Line speeds range from slowest to fastest as shown in Table 3. For example, a 10 MB transfer of data at 9600 bps would take around 3 hours.

Since communication time maybe defined as,

*T = Latency + Bandwidth*message length,* (4)

Latency becomes significant when small size packets are sent often and less significant for large packet size [12]-[14]. Latency would be *d/C,* where *d* is distance and *C* is speed of light $300 \ x \ 10^6$ *m/s*. Actual latency over the networks would be larger. Bandwidth indicates the maximum rate at which message can be sent.

 ISBN: 978-1-61208-238-7

Some problems of scanned images: The devices to scan are different in home country and foreign country. Therefore, the fingerprint images of the same person are differently captured. A solution is for all countries to use devices which meet universal standards and specifications.

Each country will have different computing power, for example, different PC cluster with different hardware /software. Each country will have a different algorithm for verification purpose. Therefore, different accuracy and computation times will be observed.

Verification to be done at source or destination ? In this case, image data is received from country of origin and processed with image at source. The problem is that, usually, the image is not stored, but only its properties are stored such as minutia type, location, orientation, etc. Therefore, the method used would most probably be not known.

Verification may be done at destination and more appropriate. In this case, image is sent with a tag to a cloud-service point, which, in turn, would send image to country of origin and processed with image data at the destination using the same method used by that country.

There are country security issues on how to deal with unfriendly countries or uncooperative countries. Images are sent via a third country, compressed and encrypted. In other words, cloud computing is done via a third country. Another question arises on whether can we trust the country's results. Also, on how to deal with non-participant countries.

TABLE III. DIFFERENT LINES AVAILABLE FOR DATA TRANSMISSION.

	Speed of Lines		
	Kbps	Mbps	Gbps
1	9.6	1.024	0.622 (OC12)
2	14.4	1.544 (DS1, T1)	1 (1000Base-T)
3	28.8	2.048 (E1, ISDN-32)	2.4 (OC48)
4	33.6	25.6 (ATM 25)	9.6 (OC192)
5	56	34 (E3)	
6	64 (ISDN)	45 (DS3, T3)	
7	128 (ISDN-2)	51 (OC1)	
8	256	100(100Base-T)	
9	512	155 (OC3)	

VII. CONCLUSION AND FUTURE WORK

In this paper we have looked at various issues concerning fingerprint verification via cloud-computing services. As numerous different devices capture fingerprint images differently, standards are required to ensure image quality is good. Countless fingerprint verification algorithms are appearing and have to ensure that they meet very near 100% accuracy. Biometric methods are probabilistic methods with decision based on estimations. Accuracy of current algorithms in the literature is not 100%. In our experiments, the threshold values depends on the patches selected. We presented a distributed fingerprint verification algorithm based on normalized correlation. We looked at communication issues such as secure exchange of fingerprint data through the internet. Computation issues such as high accuracy and reliable methods of verification. As communication time is higher than computation time, it would be more appropriate to use multi-core computers rather than a PC cluster. Peer cloud services to peer cloud services would be required to have secure contracts among each other. WSQ would be used for transmitting fingerprint images through the Internet and storage.

ACKNOWLEDGMENT

The authors would like to thank CSSE Research Centre at University of Hail for carrying out experiments on the PC cluster. This work was supported by University of Hail Research Centre grant no. 1433/6/15 and in partially supported by ARTEMIS project DEMANES (Design, Monitoring and Operation of Adaptive Networked Embedded Systems, contract 295372).

REFERENCES

[1] K. R. Jackson, et. al., "Performance Analysis of High Performance Computing Applications on the Amazon Web Services Cloud", Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference, Nov. 30 2010-Dec. 3 2010, pp. 159-168.

[2] S. Hazelhurst, "Scientific computing using virtual high-performance computing: a case study using Amazon Elastic Computing Cloud", ACM 2008 The Proceedings of the South African Institute of Computer Scientists and Information Technologists (SAICSIT) Conference, pp. 94-103, 2008.

[3] R. Cappelli, et al., "Performance Evaluation of Fingerprint Verification Systems", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 1, pp. 3-18, January 2006.

[4] D. Maltoni, "A Tutorial on Fingerprint Recognition", Advance Studies in Biometrics, pp. 43-68, 2003.

[5] R. Cappelli, et al, Fingerprint Verification Competition at International Joint Conference on Biometrics (IJCB2011), Washington DC, 2011.

[6] L. O'Gorman, Chapter 2 Fingerprint Verification, Biometrics : Personal Identification in Networked Society by Anil K. Jain, Ruud Bolle, Sharath Ankanti, pp. 1-20, Springer 1999.

[7] http://www.c3.lanl.gov/~brislawn/FBI/FBI.html, [retrieved: 6, 2012]

[8] http://saitnews.co.za/e-government/supports-banks-access-fingerprint/, [retrieved: 6, 2012]

[9] B. Chang, et. al., Adaptive Performance for VVOIP Implementation in Cloud Computing Environment, LNCS, 2012, vol. 7198/2012, pp 256-365, 2012.

[10] M. Li, W. Lee, A. Sivasubramaniam, "Efficient Peer-to-Peer Information Sharing over Mobile Ad Hoc Networks", In MobEA, pp. 1-6, 2004.

[11] X. M. Huang, C.Y. Chang, M.S. Chen, "PeerCluster: A Cluster Based Peer-to-Peer Sytem", IEEE Transactions on Parallel and Distributed Systems", vol. 17, No. 10, Oct. 2006, pp. 1110-1123.

 ISBN: 978-1-61208-238-7

[12] M. Matsuda, T. Kudoh, Y. Ishikawa, "Evaluation of MPI Implementation on Grid-connected Clusters using an Emulated WAN Environment", Proceedings of the 3rd IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID'03), pp. 10-17, 2003.

[13] G. Huston, "Measuring IP Network Performance", The Internet Protocol Journal, vol. 6, no. 1, March 2003, http://www.cisco.com/ipj , [retrieved: 6, 2012]

[14] F. Noor, M. Alhaisoni, and A. Liotta, "An Empircal Study of MPI over PC Clusters", The Third International Conference on Advances in P2P Systems, AP2PS 2011, November 20-25, Lisbon, Portugal, pp. 65-70, 2011.

[15] http://hadoop.apache.org, [retrieved: 6, 2012]

 ISBN: 978-1-61208-238-7