A Proposed Desktop Grid/Cloud computing network design for HSI Target Detection applications

José O. Nogueras Colón
Universidad del Turabo, Gurabo, Puerto Rico, jnogueras6@email.suagm.edu
Yahya M. Masalmah
Universidad del Turabo, Gurabo, Puerto Rico, yamasalmah@suagm.edu

Christian Martinez Nieves

Universidad del Turabo, Gurabo, Puerto Rico, clmartinez151@gmail.com

Rafael Rivera Soto
Universidad del Turabo, Gurabo, Puerto Rico, rafaelriverasoto@gmail.com

ABSTRACT

Hyperspectral Imagery (HSI) applications involve large data storage and computational power on real time. Grid Computing has become a powerful solution for the current computational challenges. On the other hand, Cloud Computing offers high storage scalability into a virtualization environment. This paper outlines a review of the best practice existing for both infrastructures. We propose a secure network design for HSI using the best of grid and cloud computing. Justifying this proposal, we also are including a brief comparative study between three known Hyperspectral Target Detection algorithms: Automatic Target Generation Process (ATGP), Spectral Angle Mapper (SAM) and RX Detector with our Constrained Positive Matrix Factorization (cPMF) algorithm, evaluating speed and accuracy.

Keywords: hyperspectral imagery, target detection, grid computing, cloud computing, network design.

1. INTRODUCTION

The human desire to observe and analyze remotely Earth has been creating and developing high-tech instruments with high spatial and spectral resolution. The study of spectral images has evolved from a highly complex research tool to a commodity available to us. This remote sensing technique has reached boom in many areas related to the remote study of the Earth. The use of HSI has allowed us to have greater knowledge and understanding of the processes occurring on the planet. Some important advances and discoveries have helped us to improve and maintain our quality of life (Shippert, 2004). The study of soil, vegetation, water, ozone, forest fires, erosion, planets like Mars (Girouard, et al, 2004) the detection of nuclear weapons (Grupta, 2008), bacterial identification, among others, has been achieved by the hyperspectral image processing. HSI is based on quantitative analysis of the spectral properties of different materials in the earth's surface, contiguous spectral bands recorded at different wavelengths of the electromagnetic spectrum. For each pixel is possible to obtain a full reflectance spectrum. This spectrum is a result of reflection, absorption and emission of electromagnetic energy with each material, which responds to the presence of sunlight.

The concept of hyperspectral imaging was implemented when the NASA Jet Propulsion Laboratory developed the AVIRIS (Airborne Visible-Infrared Imaging Spectrometer), which currently covers the wavelengths between 0.4 and 2.5 μm using 224 spectral channels at a resolution of 10nm. Hyperspectral sensors capture digital images with a large number of spectral channels are very close together, giving each pixel portion or a particular spectral signature of each material. This allows a very precise characterization of the surface. There many numbers of bands available, that helps to better the characterization of materials present in the scene.Most applications of this technology require rapid responses, so that critical decisions can be made in near real time. The efficiency of hyperspectral image processing algorithms in these applications has added great relevance and importance to the development and optimization of the techniques used in its implementation. The main problem is the phenomenon hyperspectral analysis of the mixture. Many pixels in the scene contain several sub-pixel level, thus resulting spectral signature in these pixels is not "pure" but is given by a composition or mixture of different substances that coexist at the sub-pixel (Paz, et al, 2008), (Plaza, et al, 2009b). Special features of hyperspectral data processing, pose different problems that must be necessarily addressed with mathematical formalisms, such as classification and segmentation or spectral mixture analysis (Plaza, et al 2009a). Other problem of hyperspectral analysis techniques is the excessive computational cost, since in certain applications it is essential to develop analytical techniques computationally efficient and capable of providing a response in real time (Masalmah and Vélez, 2006), (Plaza, et al, 2009a).

The emergence of these technologies and applications in science and engineering, results in the need for greater power infrastructures that provide computational resources to solve their problems more efficiently in terms of accuracy, time, security and cost. In the mid 1990s, the term Grid was used to describe technologies that would allow consumers to obtain computing power on demand (Foster, et al, 1998). According to (Rakib, 2013) the field of Grid computing was born to fill the gap between available technology and increasing demand for computational power. Grids integrate networking, communication, computation and information to provide a virtual platform for computation and data management in the same way that the Internet integrates resources to form a virtual platform for information. Grid computing enables the use of existing IT infrastructure to optimize compute resources and manage data and computing workloads. Its performance is based on building computational power arrays from a network of many small and pervasive computers and used to perform large calculations and operations that can be decomposed into independent work units (Rakib, 2013). This allows the creation of massive computational projects supporting scientific collaboration between research institutions, besides bringing great benefits to industries. Unlike grid computing, cloud computing leads to a future in which computing will not use only local machines, but on centralized facilities operated by utilities equipment and third-party storage. According to (Foster, et al, 2008) Cloud Computing has three characteristics that distinguish this type of distributed computing other: 1) it is massively scalable, 2) can be encapsulated as an abstract entity that offers different levels of services to customers outside the Cloud, 3) is driven by economies of scale, and 4) the services can be configured dynamically through virtualization and other methods; and delivered according to demand. As a solution to the computing and storage problems, many academic institutions, government and industries have opted for Cloud Computing. Motivated by the great advances and benefits of grid and cloud computing; and the need of more computational power for improve our HSI target detection algorithm cPMF speed, we present in this paper a review of both infrastructures with a special focus on its use for HSI applications. Then we propose a network design based on the integration and fusion of grid and cloud computing frameworks.

The rest of the paper is organized as follows; a brief comparative study between three known Target Detection algorithms and our cPMF algorithm are showing at section 2. Section 3 provides a review of HSI on grid computing environment. Section 4 gives an overview of grid computing, desktop grid and cloud computing. Section 5 presents our converged network design. Finally section 6 draws conclusions and potential future work.

2. A COMPARATIVE STUDY OF HSI TARGET DETECTION ALGORITHMS

We reviewed and implement four HSI target detection algorithms to compare their speed and accuracy: ATGP (Reed and Yu, 1990), (Ren and Chang, 2003), (Paz, et al, 2008), cPMF (Masalmah and Vélez, 2006), SAM (Yuhas, et al, 1992), (De Carvalho and Meneses, 2000) and RX (Reed and Yu, 1990), (Paz, et al, 2008). We used a real data set to validate the algorithms speed and accuracy. We used the Enrique Reef Hyperion image. The HYPERION image is 30m resolution. The low resolution motivates the spectral unmixing. The image contains areas of deep water, sea grass, carbonate sand, and reef crest. Figure 1 displays the testing image with the physical distribution of the main materials (Endmembers) and Figure 2displays the endmembers presents in the real image used in this experiment.

Figure 1: Enrique Reef Image

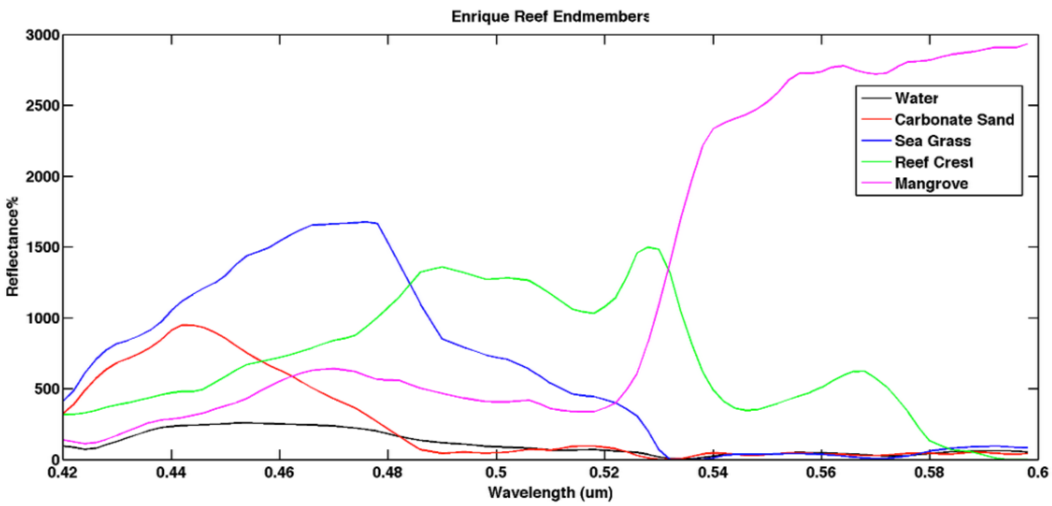

Figure 2: Endmembers

2.1 Experiment Results

An experiment was conducted by running the selected algorithms mentioned above using the real image with endmembers shown in Figure 2. The aim of this experiment was to measure two parameters: execution time and recovered endmembers accuracy. We selected these two parameters as a base for comparison between algorithms.

The execution time obtained from running the target detection algorithms on the input hyperspectral image. The obtained results are tabulated in Table 1.

Table 1: Execution Time

Algorithm	RX	SAM	CPMF	ATGP
Execution Time (Sec)	0.5000	0.03130	1225.0000	0.0625

To measure the detection accuracy of the target detection algorithms we plotted the obtained endmembers detected by each a;gorithm together with the original endmember. The obtained results are shown in Figure 3 through Figure 7.

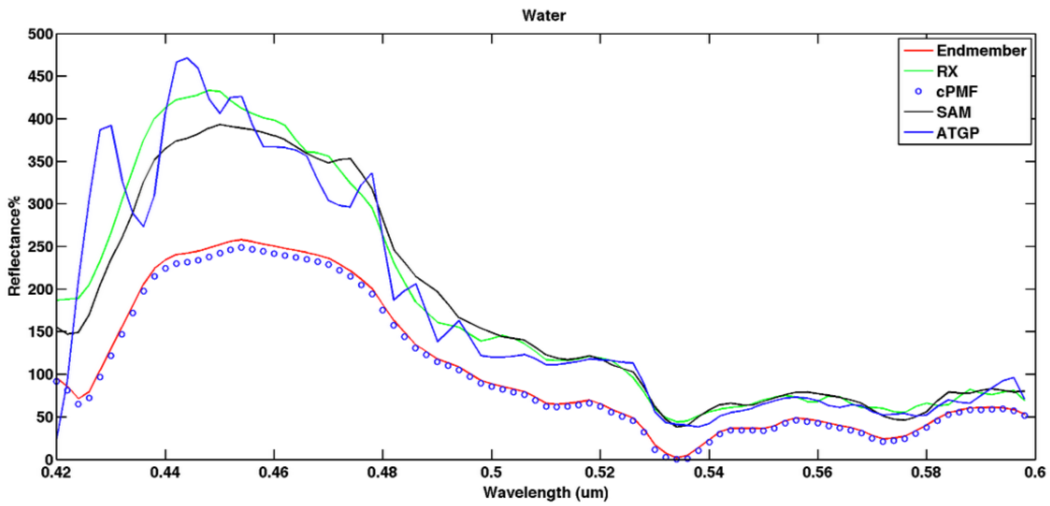

Figure 3: Water

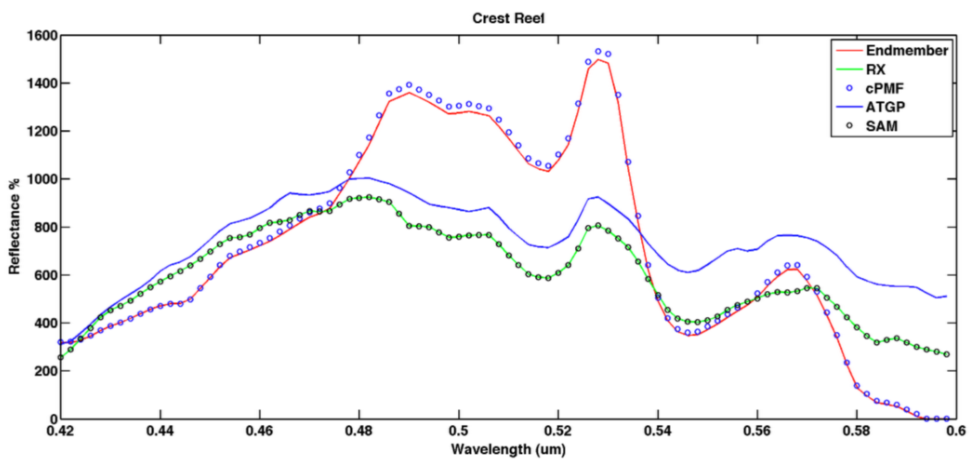

Figure 4: Crest Rest

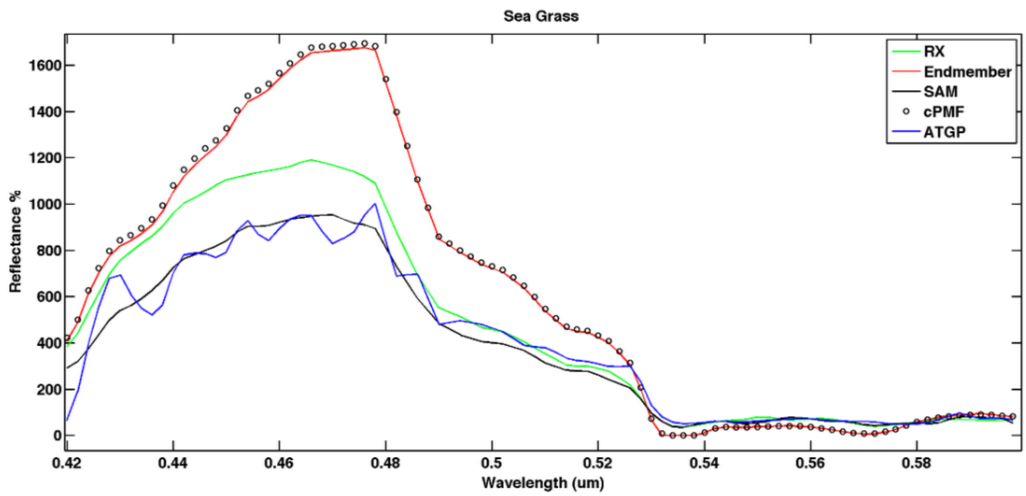

Figure 5: Sea Grass

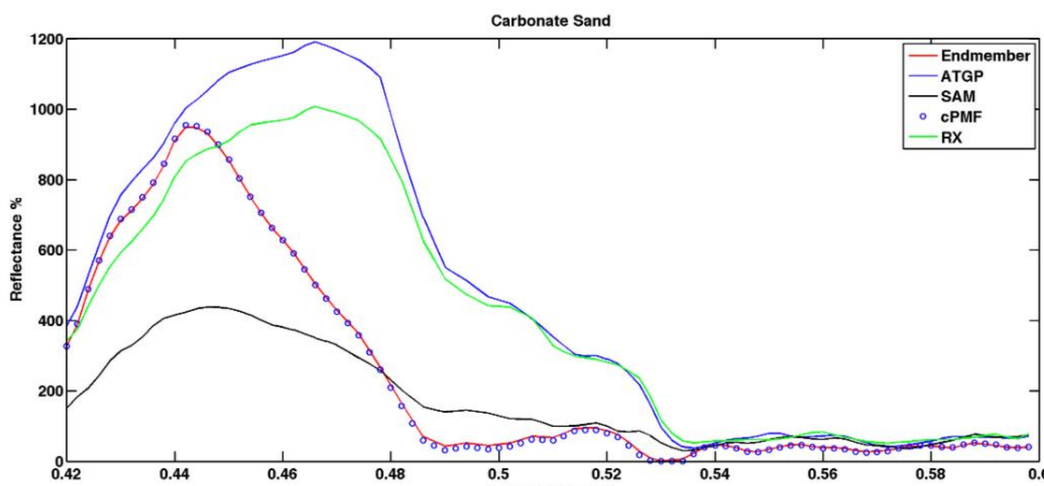

Figure 6: Carbonate Sand

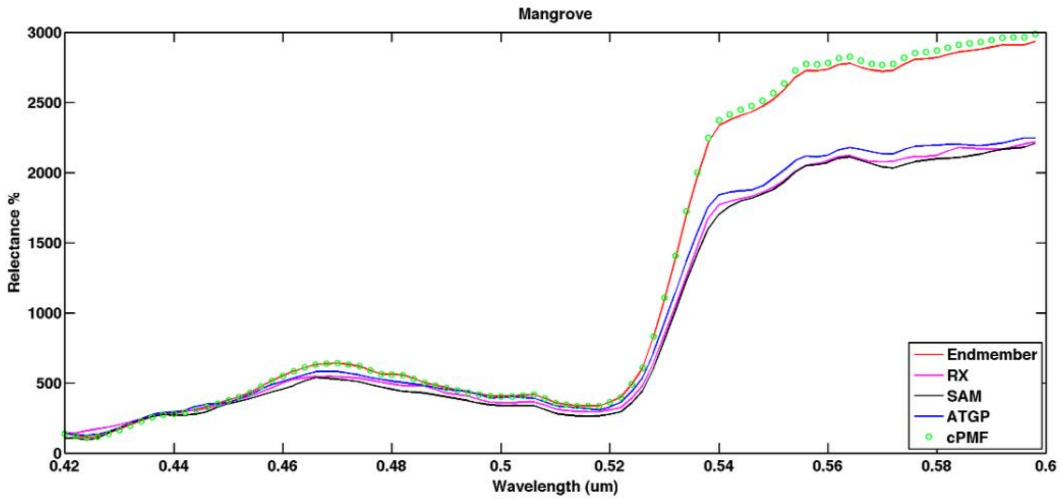

Figure 7: Mangrove

The obtained results reveals the facts eventhough the CPMF has much longer execution time, it has the higher accuracy. This in turns, motivated the use of different computing techniques to make use of this high accuracy of CPMF in a reasonable execution time. The next sections presents a proposed computing architecture to achieve the mentioned goals.

3. GRID COMPUTING ON HSI DATA PROCESSING

(Robila and Senedzak , 2007) presents the grid computing technology as an efficient alternative for hyperspectral data processing. They discuss some general considerations for grid architectures and present the results of its implementation demonstrating a significant computational speed increase. (Carvajal, et al, 2004) proposed the design and implementation of Grid -HSI architecture based on service-oriented grid applications for HIS remote analysis and visualization.

4. GRID COMPUTING

The Grid Computing concept dates back to the mid-90s. (Foster and Kesselman ,1998) define it as a hardware and software infrastructure that provides dependable, consistent, pervasive and inexpensive access to large computing capabilities. Grid computing has evolved as a high-demand technology in the scientific, instructional and business community. Allows the union and consolidation of heterogeneous systems and resources locally or distributed across multiple administrative domains for easy access and use. These resources can be distributed procedure , bandwidth and storage capacity in a virtual context. According to (Ferreira, et al, 2005) grid computing is divided into three types, according to its function: 1) computational grid , 2) data grid and 3) network grid. The computational grid is responsible for managing and leveraging processing resources providing more computational power. The data grid uses and manages data storage resources. Furthermore, the network grid handles network resources such as bandwidth, to improve the network capacity. (Poonam and Anoopta, 2013) added two more types: collaboration grids and utility grids . On the other hand, (Constantinescu, 2008) classifies grid computing into two general classes based on its infrastructure, 1) heavy weight and large scale feature- rish systems, like clusters or multiprocessors, and 2) desktop grids. Desktop grid is a very popular solution based on the use of non-dedicated PCs . Takes advantage of idle CPU cycles, storage space and other computers resources on the network to work together in an application. In these systems, the execution of an application is run by a central scheduler node (master) that is responsible for distributing the tasks between the working nodes to expect the results of these nodes. Desktop Grid offers great advantages and benefits to the scientific community, especially academic and research institutions . Its main advantages are: low cost (Kondo, et al, 2009), Green Information Technology (Schott and Emmen , 2011), flexible scalability (Foster , et al, 2008), allows the use of heterogeneous systems, easy maintenance and support and great computational power.

4.1 Desktop Grid and Cloud Computing

(Kondo, et al, 2009) conducted a comparative study that exposes the performance and cost benefits of cloud computing in conjunction with the desktop grid. Cloud computing has had great impact in recent years. (Foster, et al, 2008) cites a definition of what is cloud computing "A large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically-scalable, managed computing power, storage, platforms, and services are delivered on demand to external customers over the Internet". They also contrast two technologies, grid and cloud, for its architecture. Both technologies have the same goal or purpose: reduce computing costs and increase reliability and flexibility. However, cloud computing offers two major advantages over grid computing: virtualization and easy decentralization or centralization of all resources. Virtualization is one of the most important and indispensable elements for cloud computing, allowing abstraction and encapsulation of all its services. Cloud computing provides three levels of service: Infrastructure as a service (IaaS), Platform as a service (PaaS) and Software as a service (SaaS) to give greater freedom to the user to choose what they need and pay for it (Foster, et al, 2008). Desktop grid computing works in a master-worker paradigm. The application is split up into many small subtaks that can be processed independently (Marosi, et al, 2008b). It can be implemented locally (smaller-scale) as a large scale hierarchical, which constitutes the known volunteer computer. Another of the great advantages is its accessibility. There are a variety of middleware platforms and open source software such as Boinc, Condor and XtreamWeb. Boinc (Berkeley Open Infrastructure for The Network Computing) is a software platform computing distributed created in 2002 for volunteer computing. Boinc allows users to donate computing power of their computers to be used for science in various projects throughout the world. Their most famous project is SETI @ home, which uses all its computing power for extraterrestrial intelligence searching (Korpela, 2012). XtremeWeb is a software used to build lightweight desktop grid. Condor is a software developed by the University of Wisconsin- Madison as part of their research project named Condor Research Project. In this project they developed and implemented policies and mechanisms for High Throughput Computing (HTC) (Robila and Snedzuk, 2007).

5. Network Design

As part of our strategy to congregate the two technologies, grid and cloud computing, different tools will be used in the implementation. The major tools are: OpenNebula (Sempolinski and Thain, 2010), SZTAKI Local Desktop Grid (SZTAKI LDG) (Marosi, et al, 2008a) and (Marosi, et al, 2008b).

OpenNebula is an open-source cloud framework that lets you create a private group of virtual machines creating your own cloud (Sempolinski and Thain, 2010). SZTAKI LDG is a desktop grid system based on Boinc created by the Computer and Automation Research Institute of the Hungarian Academy of Sciences (MTA) (Marosi, et al, 2008a). Its main purpose is the easy implementation of applications in a local environment, such as companies or university departments. SZTAKI LDG also offers the ability to grow by connecting with other hierarchical Desktop Grids. (Marosi, et al, 2008b)

As a platform for virtualization we will use Virtual box for our SZTAKI LDG server and Xen (Barham , 2003) for virtual machines customers. For communications and data transfers we are going to use the SSH protocol. We are going to set two clusters with internet access connected toguether, through VPN (Virtual Private Network). The two clusters will consist of three nodes: a master node with the OpenNebula server, which is the only one with direct internet access, and two nodes that contains virtual machines including SZTAKI LDG virtual server (cluster 2). Figure 8 below shown our physical network design.

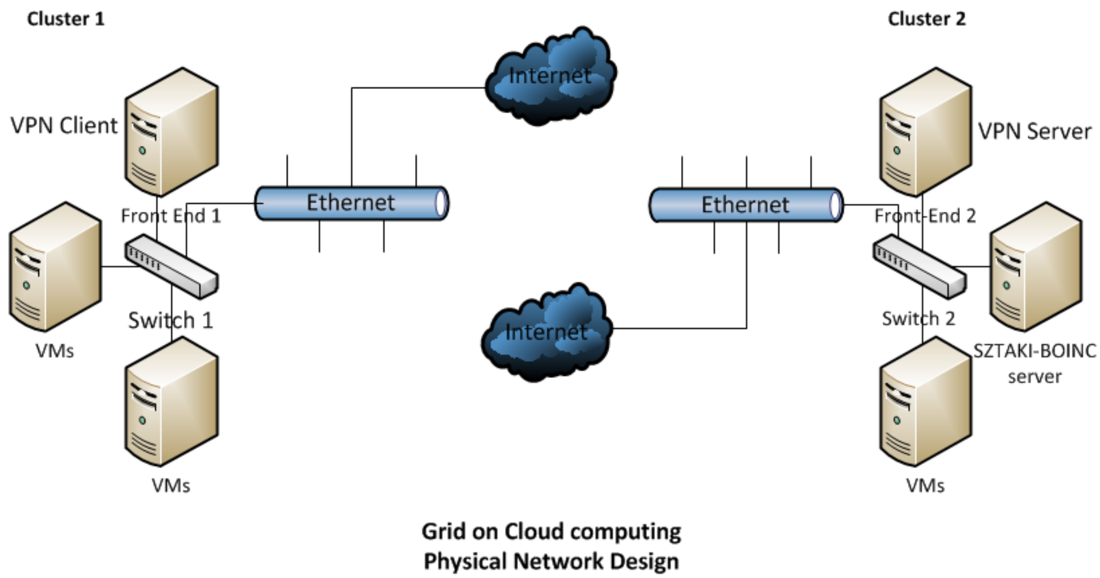

Figure 8: Desktop Grid/Cloud computing Physical Network design.

6. CONCLUSIONS AND FUTURE WORK

HSI Target Detection Applications require large computing capabilities in real time. According to the literature reviewed the Grid computing is a powerful solution for such applications. However, other technologies such as cloud computing may complement and provide greater scalability and flexibility at lower costs and high efficiency. We have highlighted the features and benefits of both solutions. Our comparative study showed that although cPMF algorithm is slower than the other three algorithms, is more accurate Motivated by these results, we have proposed a network design that combines, in a convergent form, desktop grid and cloud computing in an academic environment. Future work includes the implementation, deployment, and validation of our network, and the performance evaluation and assesment of our cPMF algorithm on a Desktop Grid / Cloud computing environment.

7. REFERENCES

Barham, P., Dragovic, B., Fraser, K., Hand, S., Harris, T., Ho, A., . . . Warfield, A. (2003). Xen and the art of virtualization. ACM SIGOPS Operating Systems Review, 37(5), 164-177.

Constantinescu-Fülöp, N. (2008). A Desktop Grid Computing Approach for Scientific Computing and Visualization,

De Carvalho, O., & Meneses, P. R. (2000). Spectral correlation mapper (SCM): An improvement on the spectral angle mapper (SAM). Paper presented at the Summaries of the 9th JPL Airborne Earth Science Workshop, JPL Publication 00-18, , 9

Ferreira, L., Lucchese, F., Yasuda, T., Lee, C. Y., Queiroz, C. A., Minetto, E., . . . Mungioli, A.Grid computing in research and education

Foster, I., & Kesselman, C. (1998) Computational grids.

Foster, I., Zhao, Y., Raicu, I., & Lu, S. (2008). Cloud computing and grid computing 360-degree compared. Paper presented at the Grid Computing Environments Workshop, 2008. GCE'08, 1-10.

Girouard, G., Bannari, A., El Harti, A., & Desrochers, A.Validated spectral angle mapper algorithm for geological mapping: Comparative study between QuickBird and landsat-TM. Paper presented at the

Grupta, N. (2008). Hyperspectral imager development at army research laboratory. Paper presented at the SPIE Defense and Security Symposium, 69401P-69401P-10.

Kondo, D., Javadi, B., Malecot, P., Cappello, F., & Anderson, D. P. (2009). Cost-benefit analysis of cloud computing versus desktop grids. Paper presented at the Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on, 1-12.

Korpela, E. J. (2012). SETI@ home, BOINC, and volunteer distributed computing. Annual Review of Earth and Planetary Sciences, 40, 69-87.

Marosi, A., Balaton, Z., Kacsuk, P., & Drotos, D. (2008). SZTAKI desktop grid: Adapting clusters for desktop grids. Remote Instrumentation and Virtual Laboratories, , 133-144.

Marosi, A., Gombas, G., Balaton, Z., Kacsuk, P., & Kiss, T. (2008). Sztaki desktop grid: Building a scalable, secure platform for desktop grid computing. Making grids work (pp. 365-376) Springer.

Masalmah, Y. M., & Vélez-Reyes, M. (2006). Unsupervised unmixing of hyperspectral imagery using the constrained positive matrix factorization. Paper presented at the Defense and Security Symposium, 62470B-62470B-9.

Paz, A., Plaza, A., & Blázquez, S. (2008). Parallel implementation of target and anomaly detection algorithms for hyperspectral imagery. Paper presented at the Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International, , 2 II-589-II-592.

Plaza, A., Benediktsson, J. A., Boardman, J. W., Brazile, J., Bruzzone, L., Camps-Valls, G., . . . Gualtieri, A. (2009). Recent advances in techniques for hyperspectral image processing. Remote Sensing of Environment, 113, S110-S122.

Plaza, A., Plaza, J., Sánchez, S., & Paz, A. (2009). Lossy hyperspectral image compression tuned for spectral mixture analysis applications on NVidia graphics processing units. Paper presented at the SPIE Optical Engineering Applications, 74550F-74550F-12.

Poonam Dabas, A. A. (2013). Grid computing: An introduction. International Journal of Advanced Research in Computer Science and Software Engineering, 3(3), 466.

Rakib, A. (2013). TCP/IP security protocols for high performance grid computing architecture International Indexed & Refereed Research Journal, IV(40), 40.

Reed, I. S., & Yu, X. (1990). Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution. Acoustics, Speech and Signal Processing, IEEE Transactions on, 38(10), 1760-1770.

Ren, H., & Chang, C. (2003). Automatic spectral target recognition in hyperspectral imagery. Aerospace and Electronic Systems, IEEE Transactions on, 39(4), 1232-1249.

Schott, B., & Emmen, A. (2011). Green desktop-grids: Scientific impact, carbon footprint, power usage efficiency. Scalable Computing: Practice and Experience, 12(2)

Sempolinski, P., & Thain, D. (2010). A comparison and critique of eucalyptus, opennebula and nimbus. Paper presented at the Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on, 417-426.

Shippert, P. (2004). Why use hyperspectral imagery? Photogrammetric Engineering and Remote Sensing, 70(4), 377-396.

Yuhas, R., Goetz, A., & Boardman, J. (1992). Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. Paper presented at the JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop. , 1

Authorization and Disclaimer