American-Eurasian Journal of Scientific Research 10 (3): 175-181, 2015
ISSN 1818-6785

DOI: 10.5829/idosi.aejsr.2015.10.3.22284

Private Information Retrieval in Hybrid Cloud

C. Pabitha and K. Devi

Department of CSE, Valliammai Engineering College,
SRM Nagar, Kattankulathur, India

Abstract: A multitude of systems are now available through which the vision of access to distributed data for personalised medicine or e-Health is now very much a reality, but retrieving these private data's are very much in question. Private Information Retrieval (PIR) allows a user to retrieve bits from a database while hiding the user's access pattern. The goal of Private Information Retrieval (PIR) is for a client to query a database in such a way that no one, including the database operator, can determine any information about the desired database record. This can be applied for retrieving useful data's of patients affected by cancer across the world., while the information is stored in cloud database for each region connected through network. This paper discusses more about hybrid cloud in trust building knowledge discovery for Private informational retrieval. Existing processes for patients' vital data collection require a great deal of labour work to collect, input and analyze the information. Privacy is also a major issue. Creating and updating these cloud databases also proves to be a difficult task, with major improvement in cancer treatments day by day, offering a solution to the needful users with private dealing is almost welcoming.

Key words: PIR · Hybrid Cloud · e – health · Cloud Computing · IR

INTRODUCTION

eHealth (also written **e-health**) is a relatively recent term for healthcare practice supported by electronic processes and communication, dating back to at least 1999. eHealthis defined as "the cost-effective and secure use of information and communications technologies in support of health and health-related fields, including healthcare services, health surveillance, health education, knowledge and research [1]. The goal of eHealth is to improve the cooperation and coordination of health care, in order to improve the quality of care and reduce the cost of care at the same time. In the medical domain, recent progress in the research and development along with advancement in Patient healthcare technologies have resulted in generation of enormous amount of data in various forms. Basically Private Information Retrieval (PIR) allows a user to retrieve a record from a database while hiding the identity of the record from a database server. For several years, cloud computing has been the focus of Health sectors and corporate bean counters, but the extremely security-conscious have been hesitant to move their data and workloads into the cloud. Now, with the underlying technology behind cloud services available for deployment inside organizations, a new model of cloud computing is gaining a foothold in business: the hybrid cloud.

Literature Review: In the recent years, E – health has enjoyed the uninterrupted development of new information technologies solutions. This has led to several benefits in terms of new hardware infrastructure, new specific and more sophisticated medical applications, increased speed in data processing, etc., that has allowed to improve and speed up the health services. The growth of data produced by the medical and clinical community requires the introduction of advanced techniques and resources in terms of computational and storage capabilities. In order to meet the needs of medical departments, hospital must continuously improve the level of modern system by using advanced science and technology innovation. The new technologies in E-Healthcare are an important part of medical treatment and follow up procedures. The needs in this community ranging from patient registration, appointment with physician, medical prescription and record of clinical

Corresponding Author: C. Pabitha, Department of CSE, Valliammai Engineering College, SRM Nagar, Kattankulathur, India.

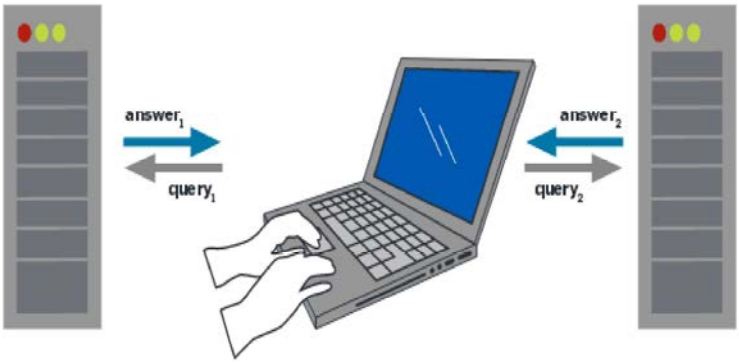
diagnostics, to laboratory tests or surgical procedures that are recorded in Electronic Patient Record (EPR) [2] and specific applications. In particular the applications are related to many different kind and producing several types of data and information: specific and experimental applications [3], knowledge discovery [3,4], orthodontics applications [5,6], SPECT images analysis [7] and other similar applications that involves new device to capture the patient feedback like eye-tracker [8]. In order to manage and process this heterogeneous amount of data and applications, is important to consider powerful distributed computational and storage technologies. Grid computing models have been successfully used for high performance computing, data library services and Medical grids, data storage etc. [9]. However grid computing could not become a widely acceptable business model primarily due to lack of marketability in non-trustful domains. Today, Cloud computing represents an essential opportunity to develop applications that ensure high performance data processing and easy management of the different tools in medical environment ensuring a consistency storage capabilities, overcoming the Grid lack. Therefore the use of new technologies and new infrastructure involves several issues ranging from costs, appropriate security of data, to the development of specific applications. This trend is highlighted by the several existent applications and architecture based on Cloud in health. Cloud computing offers several advantages by allowing users to use infrastructure, platforms and software provided by the cloud providers.

With the advancement in Cloud Computing, usage of Hybrid Cloud along with the Personalised Information Retrieval will be a boon to all health care facilities. In this paper we present a comprehensive discussion on the need to integrate PIR in e-Health applications. This paper discusses about the implications of PIR in Hybrid Cloud and proposes a new design for eHealth application. Also further extensions to the proposed are discussed and given the suggestions for the same [10].

Computational PIR: A private information retrieval (PIR) allows a user to retrieve an item from a server in possession of a database without revealing which item is retrieved. PIR is a weaker version of 1-out-of-n oblivious transfer, where it is also required that the user should not get information about other database items. One trivial, but very inefficient way to achieve PIR is for the server to send an entire copy of the database to the user. In fact, this is the only possible protocol (in the classical or the quantum setting[l]) that gives the user information theoretic privacy for their query in a single-server setting. There are two ways to address this problem: one is to make the server computationally bounded and the other is to assume that there are multiple non-cooperating servers, each having a copy of the database.

The first single-database computational PIR scheme to achieve communication complexity less than n was created in 1997 by Kushilevitz and Ostrovsky and achieved communication complexity of n^{ϵ} for any ϵ, where n is the number of bits in the database. The security of their scheme was based on the well-studied Quadratic residuosity problem. In 1999, Christian Cachin, Silvio Micali and Markus Stadler achieved poly-logarithmic communication complexity. The security of their system is based on the Phi-hiding assumption. In 2004, Helger Lipmaa achieved log-squared communication complexity $O(l\log n + k\log^2 n)$, where l is the length of the strings and k is the security parameter. The security of his system reduces to the semantic security of a length-flexible additively homomorphic cryptosystem like the Damgård–Jurik cryptosystem. In 2005 Craig Gentry and Zulfikar Ramzan achieved log-squared communication complexity which retrieves log-square (consecutive) bits of the database. The security of their scheme is also based on a variant of the Phi-hiding assumption. All previous sublinear-communication computational PIR protocol required linear computational complexity of $\Omega(n)$ public-key operations. In 2009, Helger Lipmaa [6] designed a computational PIR protocol with communication complexity $O(l\log n + k\log^2 n)$ and worst-case computation of $O(\text{n } /\log n)$ public-key operations. Amortization techniques that retrieve non-consecutive bits have been considered by Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky and Amit Sahai. As shown by Ostrovsky and Skeith, the schemes by Kushilevitz and Ostrovsky and Lipmaa use similar ideas based on homomorphic encryption. The Kushilevitz and Ostrovsky protocol is based on the Goldwasser–Micali cryptosystem while the protocol by Lipmaa is based on the Damgård–Jurik cryptosystem [11].

Achieving information theoretic security requires the assumption that there are multiple non-cooperating servers, each having a copy of the database. Without this assumption, any information-theoretically secure PIR protocol requires an amount of communication that is at least the size of the database n. Multi-server PIR protocols tolerant of non-responsive or malicious/colluding servers are called *robust* or *Byzantine robust* respectively. These issues were first considered by Beimel and Stahl (2002). An l-server system that can



Two-server information-theoretic PIR scheme. To retrieve a database record, the user queries two servers, each of which stores a copy of the database, but individual queries carry no information about what the user is after. The desired record is obtained from the servers' combined responses.

Fig. 1.1: Two servers PIR Scheme

operate where only *k* of the servers respond, ν of the servers respond incorrectly and which can withstand up to *t* colluding servers without revealing the client's query is called "*t*-private ν-Byzantine robust *k*-out-of-l PIR" [DGH 2012]. In 2012, C. Devet, I. Goldberg and N. Heninger (DGH 2012) proposed an optimally robust scheme that is Byzantine-robust to $v < k - t - 1$ which is the theoretical maximum value. It is based on an earlier protocol of Goldberg that uses Shamir's Secret Sharing to hide the query. Goldberg has released a C++ implementation on Sourceforge [12].

PIR Scheme: Let d be a small integer. Our goal here is to demonstrate how nontrivial PIR is possible by presenting a simple (d+1)-server scheme with O(n1/d) communication to access an n-bit database. The key idea behind this scheme is polynomial interpolation in a finite-field setting. We begin with some technical background. Let p> d be a prime. It is well known that addition and multiplication of numbers {0,…, p–1} modulo p satisfy the standard identities that one is used to over the real numbers. That is, numbers {0,…, p–1} form a finite field with respect to these operations. This field is denoted by Fp. In what follows we deal with polynomials defined over finite fields. Such polynomials have all algebraic properties that one is used to with polynomials over the real numbers. Specifically, a univariate polynomial over Fp of degree dis uniquely determined by its values at any d+ 1 points. Let m be a large integer. Let E1,…,E n be a certain collection of n vectors over F p of dimension m. The collection is fixed and independent of the n-bit database x. We assume the collection is known to both the servers and the user. On the preprocessing stage of the PIR protocol, each of (d+ 1) servers represents the database x by the same degree d polynomial f in m variables. The key property o f such[13]a polynomial is that for every i in [n] : f(E i) = x i. In order to ensure that such a polynomial f exists we choose m to be reasonably large compared to n. Setting[13]m = O (n1/d) suffices. Now suppose the user wants to retrieve the i-th bit of the database and knows the collection of vectors E 1,…, E n. The user's goal is thus to recover the value of the polynomial f(held by the servers) at E i. Obviously, the user cannot explicitly request the value of f at Ei from any of the servers, since such a request would ruin the privacy of the protocol; that is, some server will get to know which database bit the user is after. Instead, the user obtains the value of f(Ei) indirectly, relying on the rich structure of local dependencies between the evaluations of a low-degree polynomial f at multiple points. Specifically, the user generates a randomized collection of m –dimensional vectors P1,…,P d+1 over Fp such that:

Each of the Vectors:

- Pl is individually uniformly random and thus provides no information about Ei; and The values of any degree
- d polynomial (including the polynomial f) at P1,…,Pd+1 determine the value of the polynomial at Ei.

The user sends each server one of the vectors P1,…,Pd+1. The servers then evaluate the polynomial f at the vectors they receive and return the values they obtain back to the user. The user combines the values f(P1),…,f(Pd+1) to get the desired value f(Ei). The protocol is perfectly private and the com munication amounts to sending (d+ 1) vectors of dimension m to the servers and

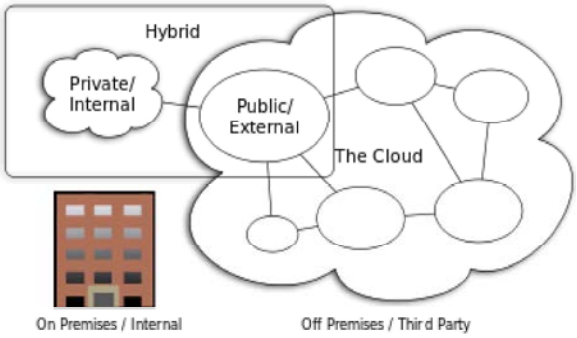


Fig. 4.1: Hybrid Cloud representation

a single value back to the user. Here, 178laborate on how vectors P1,…,Pd+1are chosen. The user picks an m-dimensional vector V uniformly at random and for every l between 1 and d+1 sets Pl= Ei+l V. Clearly, every individual vector Pl is uniformly random. To see that values of f(P1),…,f (Pd+1) determine the value of f(Ei) consider a univariate polynomial g(l) = f(Ei+ l V). Note that the degree of g is at most d. Therefore the values of g at d+ 1 points determine the polynomial g uniquely. It remains to notice that g(l) = f(Pl) and g(0) = f(Ei) = xi [14].

Hybrid Cloud: The hybrid cloud is the combination of a public cloud provider with a private cloud platform — one that's designed for use by a single organization. The public and private cloud infrastructures, which operate independently of each other, communicate over an encrypted connection, using technology that allows for the portability of data and applications.

The precision of this definition is quite important: The public and private clouds in a hybrid cloud arrangement are distinct and independent elements. This allows organizations to store protected or privileged data on a private cloud, while retaining the ability to leverage computational resources from the public cloud to run applications that rely on this data. This keeps data exposure to a bare minimum because they're not storing sensitive data long-term on the public cloud component. It's important to understand that the concept of a hybrid cloud is not simply connecting any arbitrary server to a public cloud provider and calling it hybrid. The private infrastructure must run some type of cloud services, such as NemakiWare, an open-source enterprise content management (ECM) software stack based on the interoperable CMIS standard, or Joyent Smart Datacenter, a cloud management platform for private and hybrid cloud deployments [15].

The Benefits of Going Hybrid: One clear benefit of a hybrid cloud model is having on-premises, private infrastructure that's directly accessible — in other words, not being pushed through the public internet. This greatly reduces access time and latency in comparison to public cloud services. With the looming risk of the consolidation of ISPs at the consumer/business level in the United States, the current halting of Net Neutrality and the volleying of threats between ISPs and service providers, reliance on the proper functioning of the internet — a single point of failure that can bring down the normal operations of an entire company — is an unacceptably high risk. Another benefit of a hybrid cloud model is the ability to have on-premises computational infrastructure that can support the average workload for your business, while retaining the ability to leverage the public cloud for failover circumstances in which the workload exceeds the computational power of the private cloud component. This provides the added benefit of paying for the extra compute time only when these resources are needed. Accordingly, for businesses that have milestones throughout the year where a much higher than normal amount of compute time is needed (tax season, perhaps), extending to the public cloud is a cheaper proposition than building out a private infrastructure that sits idle for most of the year. Building out the private end of a hybrid cloud also allows for flexibility in server designs. This gives companies the flexibility to provision rapid and archival storage at a likely lower cost. Combined with the announcement of new 19nm server-grade SSDs and the Helium-filled 6TB drives from HGST, data storage — fast or slow — can be achieved without the use of backup tapes [16].

PIR in Hybrid: We implement a combined version of PIR scheme for use in a Hybrid cloud computing environment that improves security issues in cloud database. Private Information Retrieval (PIR) schemes allow users to retrieve information from a database while keeping their query private

The requirements in the e-health application are as follows:

- **Privacy for the Receiver: The Receiver wants to retrieve records** from the medical database, without the Sender (DB) learning the index of those records and thus the identity of his patient.
- **Privacy for the Sender: The Sender (DB) wants to be sure that, for each** query, the Receiver learns information only on one record (defined in the query) and nothing about the other records.

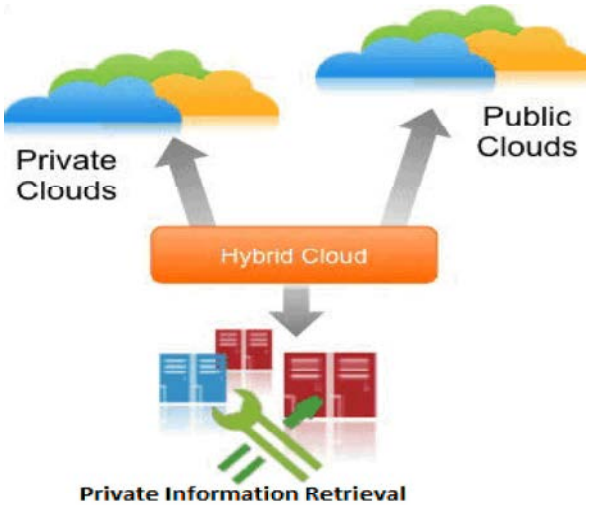


Fig. 5.1: Representation of PIR in Hybrid Cloud

- **Privacy for the data subject: In order to comply with privacy legislation,** the Sender wants to be sure that the Receiver has an valid reading authorization from the owner of the targeted record, *an Authorizer.*

In the PIR setting we have three players: Sender, a Receiver and an Authorizer.

The Receiver submits a query *Q to* the Sender, who replies with a response *R. The Receiver recovers the answer* to his query from *R. The main contribution of PIR, is to assure that* the Sender, before processing the query *Q, that the Receiver has obtained an* explicit consent from the owner of the record defined in *Q, without revealing the* identity of this owner (i.e., the Authorizer) [17].

The concept of privacy implemented in hybrid cloud databases enables the usage to be more personalized and secured. A common concern about cloud technology among enterprises is security and multi-tenancy. A hybrid portfolio eases these concerns by allowing you to choose dedicated servers and network devices that can isolate or restrict access. Healthcare organizations have several critical issues to address, including cleaning up their data infrastructure, embracing the mobile mind shift, utilizing emerging technology and building cloud-based business networks. The hybrid cloud strategy able to support existing data center and private cloud environments with multiple service delivery models, a hybrid cloud strategy can offer reduced costs, improved performance and increased scalability and flexibility to address changing needs on an ongoing basis. Healthcare organizations stand to inherit tremendous benefits.

Reevaluating data infrastructure. As more confidential records and critical communications – such as patient records and interoffice communications – are stored digitally, there is a driving need for additional storage space. A hybrid cloud strategy offers healthcare providers the means to scale resources to meet long- and short-term needs. By acquiring additional storage capacity as they need it, whether for long-term data storage or short-term spikes in usage, healthcare enterprises are able to meet their needs without the high capital costs of building additional data centers or paying for underutilized capacity. By employing a hybrid cloud strategy, IT dollars are maximized and efficiency and elasticity realized.

Compliance and Security. Patient privacy and record confidentiality have always been key concerns of the healthcare industry. With HIPAA and HITECH regulations, strong information security is more critical than ever. Operating under HIPAA- and HITECH guidelines, healthcare providers have a lot to lose by deploying insecure infrastructure that jeopardizes critical or confidential data. As the healthcare industry continues to expand its interaction with third parties, such as pharmacies, specialists and payment brokers, the potential for an information security breach increases. ealthcare organizations need to do their due diligence to ensure their HIPAA/HITECH compliance and working with a cloud service provider (CSP) – like Peak 10 which is audited for HIPAA compliance - can be a starting point.

Hybrid cloud capabilities can help healthcare organizations improve their efficiency and cost-effectiveness. By supplementing existing capability, a hybrid cloud solution can provide a scalability and flexibility that legacy systems cannot. As business needs continue to evolve and data usage continues to grow, a hybrid cloud solution may be just what the doctor ordered.

CONCLUSION

Cloud computing poses privacy concerns because the service provider can access the data that is on the cloud at any time. It could accidentally or deliberately alter or even delete information. Many cloud providers can share information with third parties if necessary for purposes of law and order even without a warrant. That is permitted in their privacy policies which users have to agree to before they start using cloud services. Solutions

to privacy include policy and legislation as well as end users' choices for how data is stored. Users can encrypt data that is processed or stored within the cloud to prevent unauthorized access. According to the Cloud Security Alliance, the top three threats in the cloud are "Insecure Interfaces and API's", Data Loss & Leakage" and "Hardware Failure" which accounted for 29%, 25% and 10% of all cloud security outages respectively - together these form shared technology vulnerabilities. In a cloud provider platform being shared by different users there may be a possibility that information belonging to different customers resides on same data server. Therefore Information leakage may arise by mistake when information for one customer is given to other. The use of PIR in Hybrid cloud covers the privacy issues in Cloud, this enables the Hybrid Cloud to be a greatest source in Health Care facilities. Hybrid clouds are frequently deployed in the financial sector, particularly when proximity is important and physical space is at a premium — such as on or adjacent to a trading floor. Pushing trade orders through the private cloud infrastructure and running analytics on trades from the public cloud infrastructure greatly decreases the amount of physical space needed for the latency-sensitive task of making trade orders. This is crucial for data security, as well. Threshold-defined trading algorithms are the entire business of many investment firms. Trusting this data to a public cloud provider is, to most firms, an unnecessary risk that could expose the entire underpinnings of their business. Hybrid cloud technology is also widely used in the healthcare industry, as the need to relay data between healthcare providers and insurance companies for hundreds of thousands of patients is a daunting task. Compliance with HIPAA (the Health Insurance Portability and Accountability Act) in this regard is a regulatory hurdle, since compartmentalizing information to comply with HIPAA over not disclosing protected health information requires extensive permissions settings. The Combined process of PIR in Hybrid can still be improved upon focusing on the efficiency in retrieving information from hybrid cloud database.

REFERENCES

1. Chia-Chi Teng, J. Mitchell, C. Walker, A. Swan, C. Davila, D. Howard and T. Needham, 2010. A medical image archive solution in the cloud, Software Engineering and Service Sciences (ICSESS), 2010 IEEE International Conference on, 431-434: 16-18.
2. Chao-Tung Yang, Lung-Teng Chen, Wei-Li Chou and Kuan-Chieh Wang, 2010. Implementation of a Medical Image File Accessing System on Cloud Computing, Computational Science and Engineering (CSE), 2010 IEEE 13th International Conference on, 321-326: 11-13.
3. Shini, S.G., Tony Thomas and K. Chithraranjan, 2012. Cloud Based Medical Image Exchange-Security Challenges, Procedia Engineering, 38: 3454-3461.
4. Bastiao Silva, L.A., C. Costa, A. Silva and J.L. Oliveira, 2011. A PACS Gateway to the Cloud, Information Systems and Technologies (CISTI), 2011 6th Iberian Conference on, 1-6: 15-18.
5. Rolim, C.O., F.L. Koch, C.B. Westphall, J. Werner, A. Fracalossi and G.S. Salvador, 2010. A Cloud Computing Solution for Patient's Data Collection in Health Care Institutions, eHealth, Telemedicine and Social Medicine, 2010. ETELEMED '10. Second International Conference on, 95-99: 10-16.
6. Poulymenopoulou, M., F. Malamateniou and G. Vassilacopoulos, 2011. E-EPR: a cloud-based architecture of an electronic emergency patient record. In Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '11). ACM, New York, NY, USA, Article, 35: 7.
7. Koufi, V., F. Malamateniou and G. Vassilacopoulos, 2010. Ubiquitous access to cloud emergency medical services, Information Technology and Applications in Biomedicine (ITAB), 2010 10th IEEE International Conference on, 1-4: 3-5.
8. Yu, D. Weider and D.R. Bhagwat, 2011. Modeling Emergency and Telemedicine Heath Support System: A Service Oriented Architecture Approach Using Cloud Computing, IJEHMC, 2(3): 63-88.
9. Mohapatra, S. and K. Smruti Rekha, 2012. Sensor-Cloud: A Hybrid framework for remote patient monitoring. International Journal of Computer Applications, (0975-8887): 55-2.
10. Faro, A., D. Giordano and C. Spampinato, 2012. Combining literature text mining with microarray data: Advances for system biology modeling, Briefings in Bioinformatics, (1): 61-82.
11. Faro, A., D. Giordano, F. Maiorana and C. Spampinato, 2009. Discovering genes-diseases associations from specialized literature using the grid, IEEE Transactions on Information Technology in Biomedicine, (4): 554-560.

12. Faro, A., D. Giordano, C. Pino and C. Spampinato, 2010. Visual attention for implicit relevance feedback in a content based image retrieval, Eye Tracking Research and Applications Symposium (ETRA), pp: 73-76.
13. Faro, A. and D. Giordano, 2003. Design memories as evolutionary systems socio-technical architecture and genetics, Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, pp: 4334-4339.
14. Giordano, D., R. Leonardi, F. Maiorana, G. Cristaldi and M.L. Distefano, 2005. Automatic landmarking of cephalograms by cellular neural networks, Lecture Notes in Computer Science, 3581: LNAI, 333-342.
15. Leonardi, R., D. Giordano and F. Maiorana, 2009. An evaluation of cellular neural networks for the automatic identification of cephalometric landmarks on digital images, Journal of Biomedicine and Biotechnology, art. no. pp: 717102.
16. Faro, A., D. Giordano, C. Spampinato, S. Ullo and A. Di Stefano, 2011. Basal ganglia activity measurement by automatic 3-D striatum segmentation in SPECT images, IEEE Transactions on Instrumentation and Measurement, (10): 3269-3280.
17. Giordano, D., C. Pino, C. Spampinato, M. Fargetta and A. Di Stefano, 2011. Nuclear Medicine Image Management System for Storage and Sharing by using Grid Services and Semantic Web. HEALTHINF, pp: 80-86.