ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

Improving Data Encryption using Fine Grained Access Control and Semantic Keyword Search over cloud Storage

[1] M.Parthiban, [2]B.Ajay, [3]K.Kalaiyarasan, [4]A.Dhinesh Pandi
[1][2][3][4] Department of Computer Science of Engineering,
VSB Engineering College, Karur, Tamil Nadu

***Abstract* – In today's data intensive world, cloud computing is new type of computing paradigm which enables sharing of computing resources over the internet. The cloud characteristics are on-demand self-service, location independent network access, ubiquitous network access and usage based pay. Due to this charming features private and public organization are outsourcing their large amount of data on cloud storage. Organizations are motivated to migrate their data from local site to central commercial public cloud server. By outsourcing data on cloud users gets relief from storage maintenance. Although there are many benefits to migrate data on cloud storage it brings many security problems. Therefore, the data owners hesitate to migrate the sensitive data. In this case the control of data is going towards cloud service provider. This security problem induces data owners to encrypt data at client side and outsource the data. By encrypting data improves the data security but the data efficiency is decreased because searching on encrypted data is difficult. The search techniques which are used on plain text cannot be used over encrypted data. The existing solutions supports only identical keyword search; semantic search is not supported. In the proposed work, semantic multi-keyword ranked search system with verifiable outsourced decryption. To improve search efficiency this system includes semantic search by using fuzzy search.**

INTRODUCTION

Cloud computing is a type of Internet-based computing that provides shared computer processing resources and data to computers and other devices on demand. It is a model for enabling ubiquitous, on-demand access to a shared pool of configurable computing resources (e.g., computer networks, servers, storage, applications and services), which can be rapidly provisioned and released with minimal management effort. Cloud computing and storage solutions provide users and enterprises with various capabilities to store and process their data in either privately owned, or third-party data centers that may be located far from the user–ranging in distance from across a city to across the world. Cloud computing relies on sharing of resources to achieve coherence and economy of scale, similar to a utility (like the electricity grid) over an electricity network.

Cloud computing is the result of the evolution and adoption of existing technologies and paradigms. The goal of cloud computing is to allow users to take benefit from all of these technologies, without the need for deep knowledge about or expertise with each one of them. The cloud aims to cut costs, and helps the users focus on their core business instead of being impeded by IT obstacles. The main enabling technology for cloud computing is virtualization. Virtualization software separates a physical computing device into one or more "virtual" devices, each of which can be easily used and managed to perform computing tasks. With operating system–level virtualization essentially creating a scalable system of multiple independent computing devices, idle computing resources can be allocated and used more efficiently. Virtualization provides the agility required to speed up IT operations, and reduces cost by increasing infrastructure utilization. Autonomic computing automates the process through which the user can provision resources on-demand. By minimizing user involvement, automation speeds up the process, reduces labor costs and reduces the possibility of human errors. [37] Users routinely face difficult business problems. Cloud computing adopts concepts from Service-oriented Architecture (SOA) that can help the user break these problems into services that can be integrated to provide a solution. Cloud computing provides all of its resources as services, and makes use of the well-established standards and best practices gained in the domain of SOA to allow global and easy access to cloud services in a standardized way.

Cloud computing also leverages concepts from utility computing to provide metrics for the services used. Such metrics are at the core of the public cloud pay-per-use models. In addition, measured services are an essential part of the feedback loops in autonomic computing, allowing services to scale on-demand and to perform automatic failure recovery. Cloud computing is a kind of grid computing; it has evolved by addressing the QoS (quality of service) and reliability problems. Cloud computing provides the tools and technologies to build

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

data/compute intensive parallel applications with much more affordable prices compared to traditional parallel computing techniques. In the proposed work, semantic multi-keyword ranked search system with verifiable outsourced decryption is implemented which is used to improve search efficiency this system includes semantic search by using fuzzy search.

II. LITERATURE SURVEY

This literature survey explains how the data storage is secured over cloud computing environment. J. Tang at el, proposed a novel cryptographic primitives and various security protection proposals for cloud data services have been presented recently. They can be broadly classified into four categories: confidentiality-assured cloud data service, owner-controlled cloud data sharing, integrity guaranteed cloud data storage, and privacy-preserving cloud data access. More specifically, searchable encryption and homomorphic encryption techniques are proposed to enforce secure data search and data computation, respectively; selective encryption and attribute-based encryption techniques are introduced to achieve authorized access and secure data sharing; provable data possession and proof-of-retrievability techniques are presented to ensure data intactness and retrievability; and privacy preservation is enabled to protect multiple dimensions of private information. R. Curtmola at el., proposed for provisioning symmetric encryption with search capabilities; the resulting construct is typically called searchable encryption. The area of searchable encryption has been identified by DARPA as one of the technical advances that can be used to balance the need for both privacy and national security in information aggregation systems. In addition, it can allow services such as Google Desktop to offer valuable features (e.g., the ability of searching a client's data across several computers) without sacrificing the client's privacy.

W. Sun at el., Despites the tremendous business and technical advantages, privacy concern is one of the primary hurdles that prevent the widespread adoption of the cloud by potential users, especially if their sensitive data are to be outsourced to and computed in the cloud. Examples may include financial and medical records, and social network profiles. Cloud service providers (CSPs) usually enforce users' data security through mechanisms like firewalls and virtualization.

N. Cao et al., defined and solve the problem of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system-wise privacy in the cloud computing paradigm. Among various multi-keyword semantics, we choose the efficient similarity measure of "coordinate matching", i.e., as many matches as possible, to capture the relevance of data documents to the search query. Specifically, we use "inner product similarity", i.e., the number of query keywords appearing in a document, to quantitatively evaluate such similarity measure of that document to the search query. During the index construction, each document is associated with a binary vector as a sub-index where each bit represents whether corresponding keyword is contained in the document. The search query is also described as a binary vector where each bit means whether corresponding keyword appears in this search request, so the similarity could be exactly measured by the inner product of the query vector with the data vector.

Z. Xia et al., proposed a secure tree-based search scheme over the encrypted cloud data, which supports multi-keyword ranked search and dynamic operation on the document collection. Specifically, the vector space model and the widely-used "term frequency (TF) × inverse document frequency (IDF)" model are combined in the index construction and query generation to provide multikeyword ranked search. In order to obtain high search efficiency, we construct a tree-based index structure and propose a "Greedy Depth-first Search" algorithm based on this index tree. L. Wang et al., first presented a model of the PPTP issue based on the mapping to PPDP, which formally characterizes the interaction between users and Web applications, the observation made by eavesdroppers, the privacy requirement, and the overhead of padding. Based on the model, formulated several PPTP problems under different assumptions, and discuss the complexity. They show that minimizing padding cost under a given privacy requirement is generally intractable. Next, L. Wang et al., design several heuristic algorithms for solving the PPTP problems in polynomial time with acceptable overhead. Finally, they demonstrated the effectiveness and efficiency of algorithms by both analytical and experimental evaluations. The contribution of this research work is threefold. J. Li et al., focused on enabling effective yet privacy preserving fuzzy keyword search in Cloud Computing. To the best of knowledge, formalize for the first time the problem of effective fuzzy keyword search over encrypted cloud data while maintaining keyword privacy. Fuzzy keyword search

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

greatly enhances system usability by returning the matching files when users' searching inputs exactly match the predefined keywords or the closest possible matching files based on keyword similarity semantics, when exact match fails.

C. Wang et al., works among the first few ones to explore ranked search over encrypted data in Cloud Computing. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency), thus making one step closer towards practical deployment of privacy-preserving data hosting services in the context of Cloud Computing. To achieve our design goals on both system security and usability, C. Wang et al., proposed to bring together the advance of both crypto and IR communities to design the ranked searchable symmetric encryption scheme, in the spirit of "as-strong-as-possible" security guarantee. M. Chuah et al., proposed a privacy-aware bed tree based solution that supports fuzzy multi-keyword search and incremental updates. Given a collection of files, our solution first constructs a list of useful keywords (including multikeywords). In the proposed work, semantic multi-keyword ranked search system with verifiable outsourced decryption is done over cloud computing environment. To improve search efficiency the proposed system includes semantic search by using fuzzy search. The advantages of using our proposed system are retrieve large amount of relevant documents based on user searchable keyword, Anonymous access can be blocked, Mapping can be done in encrypted cloud storage, Easy to analyze relationship between multi keywords.

III. PROPOSED WORK

In cloud computing, scalable and elastic storage and computation resources are provisioned as measured services through the Internet. Outsourcing data services to the cloud allows organizations to enjoy not only monetary savings, but also simplified local IT management since cloud infrastructures are physically hosted and maintained by the cloud providers. To minimize the risk of data leakage to the cloud service providers, data owners opt to encrypt their sensitive data, e.g., health records, financial transactions, before outsourcing to the cloud, while retaining the decryption keys to themselves and other authorized users. This in turn renders data utilization a challenging problem. For example, in order to search some relevant documents amongst an encrypted data set stored in the cloud, one may have to download and decrypt the entire data set. This is apparently impractical when the data volume is large. Thus, mechanisms that allow users to search directly on the encrypted data are of great interest in the cloud computing era., efficient multi-keyword fuzzy search over encrypted data remains a challenging problem. We want to point out that the efforts on search over encrypted data involve not only information retrieval techniques such as advanced data structures used to represent the searchable index, and efficient search algorithms that run over the corresponding data structure, but also the proper design of cryptographic protocols to ensure the security and privacy of the overall system. Although multi-keyword search and fuzzy search have been implemented separately, a combination of the two does not lead to a secure and efficient multi-keyword fuzzy search scheme. In this paper, we propose a brand new idea for achieving multi-keyword (conjunctive keywords) fuzzy search. Different from existing multi-keyword search schemes, our scheme eliminates the requirement of a predefined keyword dictionary. The fuzziness of the keyword is captured by an innovative data structure and algorithmic design without expanding the keyword index, and hence exhibits a high efficiency in terms of computation and storage. Besides the search result, the cloud server should not deduce any keyword information of the file set from secure indexes and trapdoors. Keyword privacy requires indexes and queries be properly represented and securely encrypted. The need of a pre-defined dictionary is a limiting factor that makes dynamic data operations, such as dataset/index update, very difficult. In our design, we would like to eliminate this requirement. In order to improve the computation performance and reduce communication overhead, we propose a new verifiable outsourcing scheme with constant cipher text length. To be specific, our scheme achieves the following goals. (1) Our scheme is verifiable which ensures that the user efficiently checks whether the transformation is done correctly by the CSP. (2) The size of cipher text and the number of expensive pairing operations are constant, which do not grow with the complexity of the access structure. (3) The access structure in our scheme is AND gates on multivalued attributes and we prove our scheme is verifiable and it is secure against selectively chosen-plaintext attack in the standard model. (4) We give some performance analysis which indicates that our scheme is adaptable for various limited bandwidth and computation-constrained devices. With the cloud service being more and more popular in modern society, ECC technology has become a promising orientation. It allows users to use flexible access control to access files stored in the cloud

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

server with encrypted form. Though its advantages make it a powerful tool for cloud, one of its main performance challenges is that the complexity of decryption computation is linearly correlated with the access structure. Given a cipher text and a transformation key, CSP transforms a cipher text into a simple cipher text. The user only needs to spend less computational overhead to recover the plaintext from simple cipher text. However, the correctness of the transformation cipher text which the CSP gives to the user cannot be guaranteed because the latter does not have the original cipher text. It is a security threat that malicious cloud service provider (CSP) may replace the original cipher text and give the user a transformed cipher text from another cipher text which CSP wants the user to decrypt. The computational overhead for the decryption and transformation operations in our scheme is constant, which does not rely on the amount of attributes. In addition, we outsource the expensive operation to the cloud service provider and leave the slight operations to be done on user's device. Therefore, our scheme is very efficient. The advantages of using our proposed system are retrieve large amount of relevant documents based on user searchable keyword, Anonymous access can be blocked, Mapping can be done in encrypted cloud storage, Easy to analyze relationship between multi keywords.

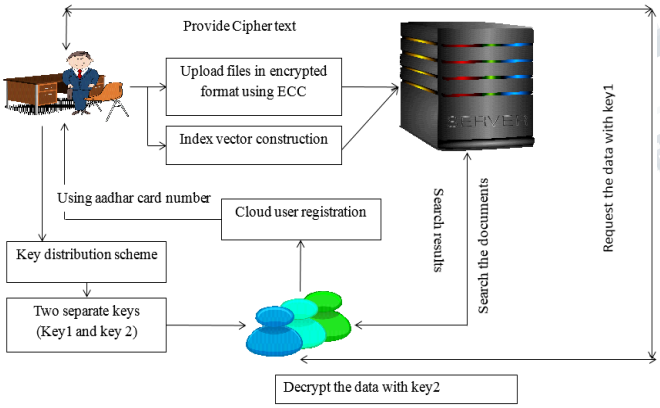

Fig.1 Data Encryption over Cloud Computing Environment

Fig.1 shows Data Encryption over Cloud Computing Environment, in which data is encrypted using fine grained access control and semantic keyword search. The data owner registers into the cloud server. The cloud owner encrypts the files using ECC algorithm. And stored in encrypted database in cloud system. The data owner trains the keywords into tables. Keywords are trained with term frequency values. Train the keywords with ranking concept. The data user searches the keywords and search in index table. The server provides data owner details. After that, user request data to owner and send secret key to user. User download the document with secure manner.

Semantic fuzzy search algorithm:

Fuzzy search using symmetric encryption has been a challenge as it was being carried out using single exact keyword only and there had been use of inverted indexes as well which were not so proficient. In order to preserve the privacy of the query keyword cosine similarity measurement has been used for the multi keyword search but it did not support fuzzy search and also required the use of predefined dictionary which lacked scalability and flexibility for modification and updating of the data. These drawbacks create the necessity of the new technique of multi keyword fuzzy search. Thus from a new perception semantic expansion based multi keyword fuzzy search reinforces the system usability by returning the exactly matched files and the files including the terms semantically related to the query keyword, which boosts the search flexibility and usability. Fuzzy searching will find a word even if it is misspelled. Fuzzy searching can be beneficial for searching text that may contain typographical errors. Multi-keyword search scheme abolishes the requirement of a predefined keyword dictionary and accomplishes this by several novel designs based on locality-sensitive hashing which is secure, efficient and accurate. A keyword is first transmuted to a bigram set, which contains all the contiguous 2 letters appeared in the keyword. The ultimate secure index for each file is a Bloom filter that comprises all the keywords in the file. The search can then be done by qualifying the relevance of the query to each file, which is done through a simple inner product of the index vector and the query vector. If a document consists of the keyword(s) in the query, the corresponding bits in both vectors will be 1 hence the inner product will return a high value. The data owner constructs the secure indexes by using the secret key and then uploads the indexes alongside with the data files to the cloud server. The cloud server will store the encrypted data and hence the data and be restructured efficiently, due to the fact that it doesn't need a pre-defined global dictionary a search and every document is individually indexed. Consequently, dataset updates, such as file adding, file deleting and file modifying, can be done easily carried out, concerning only the indexes of the files to be modified, without upsetting any other files.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

Verifiable outsourced decryption scheme:

For the decryption purpose the Access Policy plays a vital role. The access policy is the sentence formation by using all the attributes of the user that he uses to encrypt the data. Once if the access policy is matched with the attributes only then the user is an authorized user and will be able to decrypt the data that will be sent by the server. ABE also introduces the intermediate server known as the proxy server that plays a vital role in reducing the work load of the main server. The user once he encrypts the file, the file will be stored in the server. User once in need of file will request to the intermediate server that is proxy server and the proxy server will in turn send the user details to the server and will ask the server to check whether the user is the authorized user or not. Based on the reply of the server, proxy server will take the further action. If the user is an authorized user the server sends the reply as an authorized user, once getting the reply as such, the proxy server will request the user to send the transformation key. By using the transformation key, the proxy server will partially decrypts the file that will be known as transformed cipher-text. The transformed cipher-text is than sent to the user for the complete decryption where the plain text will be generated. As the proxy server is the transparent server there are chances of proxy server taking the wrong file or taking the correct file with the wrong information in it on the request of the authorized user. To eliminate this, we are using the checksum, which verifies whether the file that was encrypted and stored to the server and the file that is been received through the proxy server for the partial decryption that is the transformed cipher-text is the same. Attribute Based Encryption with Verifiable Outsourced Decryption guarantees the security property that no malicious cloud will be able to learn anything about the encrypted data.

IV. EXPRIMENTAL ANALYSIS

Data Encryption over Cloud Computing Environment, in which data is encrypted using fine grained access control and semantic keyword search. Hypertext Preprocessor (the name is a recursive acronym) is a widely used, general-purpose scripting language that was originally designed for web development to produce dynamic web pages. For this purpose, PHP code is embedded into the HTML source document and interpreted by a web server with a PHP processor module, which generates the web page document. As a general-purpose programming language, PHP code is processed by an interpreter application in command-line mode performing desired operating system operations and producing program output on its standard output channel. It may also function as a graphical application. PHP is available as a processor for most modern web servers and as standalone interpreter on most operating systems and computing platforms. PHP was originally created by Rasmus Lerdorf in 1995 and has been in continuous development ever since. The main implementation of PHP is now produced by The PHP Group and serves as the de facto standard for PHP as there is no formal specification. PHP is free software released under the PHP License, which is incompatible with the GNU General Public License (GPL) because restrictions exist regarding the use of the term PHP.

Hypertext refers to files linked together using hyperlinks, such as HTML (Hypertext Markup Language) files. Preprocessing is executing instructions that modify the output. Below is a demonstration of the difference between HTML and PHP files.

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

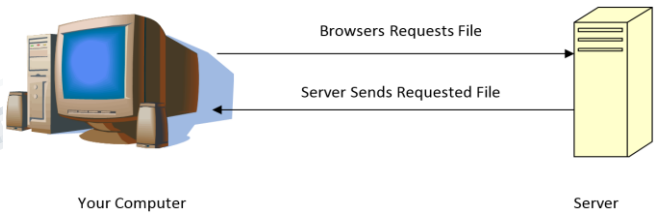

Fig.2 Accessing an HTML Page

Your browser sends a request to that web page's server (computer) for the file (HTML or image) you wish to view shown in figure 2. The web server (computer) sends the file requested back to your computer. Your browser displays the file appropriately. If you request a PHP file (ends with ". php"), the server handles it differently.

Fig 3 Accessing a PHP Page

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

When accessing a PHP Page, the browser sends a request to that web page's server for the PHP file you wish to view shown in figure 3. The web server calls PHP to interpret and perform the operations called for in the PHP script. The web server sends the output of the PHP program back to your computer. Your browser displays the output appropriately.

Benefit of PHP are the server does processing, the output of PHP files changes when its input changes. For example, most of the pages on the Horticulture site have only two (2) PHP commands: Include the header file that defines the links on the left, the banner, and the quick links at the top and include the footer file that displays the mission statement and Horticulture contact information. Because including the files is performed every time the PHP file is accessed, when the header/footer files change, the new content will be immediately updated. In other words, if you add a new link, every page that includes the header will immediately display the new link.

Security: About 30% of all vulnerabilities listed on the National Vulnerability Database are linked to PHP. These vulnerabilities are caused mostly by not following best practice programming rules: technical security flaws of the language itself or of its core libraries are not frequent (23 in 2008, about 1% of the total). Recognizing that programmers make mistakes, some languages include taint checking to detect automatically the lack of input validation which induces many issues. Such a feature is being developed for PHP, but its inclusion in a release has been rejected several times in the past. There are advanced protection patches such as Suhosin and Hardening-Patch, especially designed for Web hosting environments.

PHPIDS adds security to any PHP application to defend against intrusions. PHPIDS detects attacks based on cross-site scripting (XSS), SQL injection, header injection, directory traversal, remote file execution, remote file inclusion, and denial-of-service (DoS)

V. CONCLUSION

Cloud computing enables sharing of computing resources over the internet in which software industries are motivated to migrate their data from local site to central commercial public cloud server. The existing work which supports only identical keyword search was the semantic search is not supported. The proposed semantic multi-keyword ranked search system with verifiable outsourced decryption which supported and improved search efficiency were the system includes semantic search by using fuzzy search.

REFERENCES

[1] G. Eason, B. Noble, and I.N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[3] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[4] Annu.N “Classification of Glaucoma Images using Wavelet based Energy Features and PCA”. International Journal of Scientific &

[5] Engineering Research, Volume 4, Issue 5, May-2013 ISSN 2229-5518.

[6] Cheng-Hsuan Li, Bor-Chen Kuo, Member, IEEE, “A Spatial-contextual Support vector machine for remotely sensed image classification”,

[7] IEEE Transanction on Geoscinece and remote sensing, Vol.50,No.3, March 2012, 0196-2892.

[8] Fereidoun A and Mianji, Member, IEEE, “SVM-Based Unmixing-to-Classification Conversion for Hyperspectral Abundance

[9] Quantification”, IEEE Transanction on Geoscinece and remote sensing, Vol.49,No.11, Novenber 2011, 0196-2892.

[10] Gwenole quellec, Stephen R. Russell and Michael D. Abramoff, senior Member, IEEE, “Optimal filter framework for automated,

[11] instantaneous detection of lesions in retinal images”, IEEE Transanction on medical imaging, Vol.30, N0.2, February 2011, 0278-0062.

[12] Linlin Shen and Sen Jia, "Three-Dimensional gobar wavelets for pixel based hyperspectra imahery classification", IEEE Transanction on

[13] Geoscinece and remote sensing, Vol.49, N0.12, December 2011, 0196-2892.

[14] Rajendra Acharya.U Sunmeet Due, Xian Du and Chua Kuang, "Automated diagnosis of glaucoma using texture and higher spectra

[15] features", IEEE Transanction on information technology in biomedicine, Vol. 15, No.3, May 2011, 1089-7771.

[16] Shen.L and L. Bai, "Mutualboost learning for selecting Gabor features for face recognition," Pattern Recognit. Lett., vol. 27, no. 15, pp.

[17] 1758–1767, Nov.2006.

[18] Sumeet Dua, Senior Member, IEEE, "Wavelet-based energy features for glaucomatous image classification", IEEE Transanction on

[19] information technology in biomedicine, Vol. 16, No.1, January 2012, 1089-7771.

[20] Tarabalka.Y M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM- and MRF-based method for accurate classification of hyperspectral

[21] images," IEEE Geosci. Remote Sens. Lett., vol. 7, no. 4, pp. 736–740, Oct. 2010.

[22] Yorozu .Y M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface,"

[23] IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

Zhao.Y.-Q L. Zhang, and S. G. Kong, "Band-subset-based clustering and fusion for hyperspectral imagery classification," IEEE Trans.Geosci. Remote Sens., vol. 49, no. 2, pp. 747–756, Feb. 2011