



www.ijarcsse.com

## Security in Cloud Storage over Text Document using Hadamard Matrix and Compression Technique

Jalneesh Singh, Vinay Prakash Srivastava, Prof. Dileep Kumar Gupta

Computer Science & Engineering

B.B.D. University, Lucknow, India

---

**Abstract**— Cloud computing is a type of internet based computing, where different resources such as servers, data storage, and applications are delivered to the organizations through the internet as pay per use system. Cloud Computing reduces the cost. It is a location independent technology which leads access the services from anywhere anytime. In cloud computing major issue is security. In this paper, we have focused on security in cloud storage on text documents using Hadamard matrix and compression technique. Firstly the original text documents encrypted by Hadamard transform technique and then encrypted text document compress by compression technique. The encryption technique is chaining technique that uses Hadamard transforms for encryption and decryption which has increased the level of security. This method is so effective because the key size will be large and the most part in the key is different. But the chaining of the Hadamard transform can be as long as one pleases and the choices as far as the formation of blocks that are converted into non binary numbers is randomly large; the effectiveness of this strategy is potentially very high. So the cloud storage services become more effective when the level of security can be increased as per demand. The compression technique used for reduces the size of encrypted text document. This mechanism provides reliable, confidential text document security from leakage in cloud storage. Cloud computing is more popular model because it is not restricted to a particular location user can access the services from anywhere anytime by pay per use system.

**Keywords**— Cloud computing, Security, Hadamard transform, Compression technique.

---

### I. INTRODUCTION

The cloud computing is new vision of information technology enterprises. The user stored data in cloud may be frequently updated such as insertion, deletion, modification, appending etc. In cloud computing user many types of data stored and demands to safety of their data for long term. Cloud storage provides users to remotely store their data and use the on demand high quality cloud application without any burden of local hardware and software management, where computing infrastructure, software applications, business stages and information capability are delivered to organization through the Internet. Cloud computing is a pay per use model based service where we use network cloud storage space and computing resources. Cloud computing resources access anywhere at any time when would be needed. Cloud computing refers to application delivered as services over the internet as well as to the cloud infrastructure namely the hardware and software and data centers that provides this services. The cloud computing vendors such as Amazon Simple Storage Service (S3) and Amazon Elastic Compute Cloud (EC2) are internet based online services which provide large amounts of storage space and customizable computing resources. Cloud computing inevitably goes on new challenging security threats. The various types of data stored in cloud by each user and the demand of long terms continuous assurance of their data safety. Cloud computing is not just a third party data repository. Cloud computing is a type of computing, where IT related capabilities are provided “as a service” to end users. Clients can access technology without knowledge of, or even control over the infrastructure that supports them. Basically, cloud computing offers three layers of services such as Infrastructure as a Service, Platform as a Service and Software as a Service. It is shown in Fig. 1[13]. Organization of the paper is as follows: Types of cloud is discussed in the section 2. Related work has been discussed in the section 3. System model has been discussed in the section 4. Mathematical tools have been discussed in the section 5.

Algorithm has been discussed in the section 6. Proposed model has been discussed in the section 7. Conclusion has been discussed in the section 8.

**Software as a service (SaaS):** Software as a Service provider provides remunerator access on demands both applications and resources over the internet. The provider installs and operates application software on a cloud infrastructure. Example of SaaS is Google Apps, salesforce.com etc.

**Platform as a service (PaaS):** Platform as a Service is next level of software as service setup. Platform as a services provider provides platform to remunerator to develop application by use of programming language and tools over the internet. Example of PaaS is Google App Engine, Which is utilized for developing and hosting web application within Google managed data centre.

**Infrastructure as a service (IaaS):** Infrastructure as a service provider provides the computing resources such as processing power, network and storage to the client where client developed and run unbound application, which can include operating systems and applications. Example of IaaS is Amazons Elastic Compute Cloud (EC2) service. The user access to an EC2 for a period of time to be used as a resource for whatever purpose desires the user. And Amazons simple storage service (S3): The user is given access to low latency data storage that is reachable from any location through the internet.

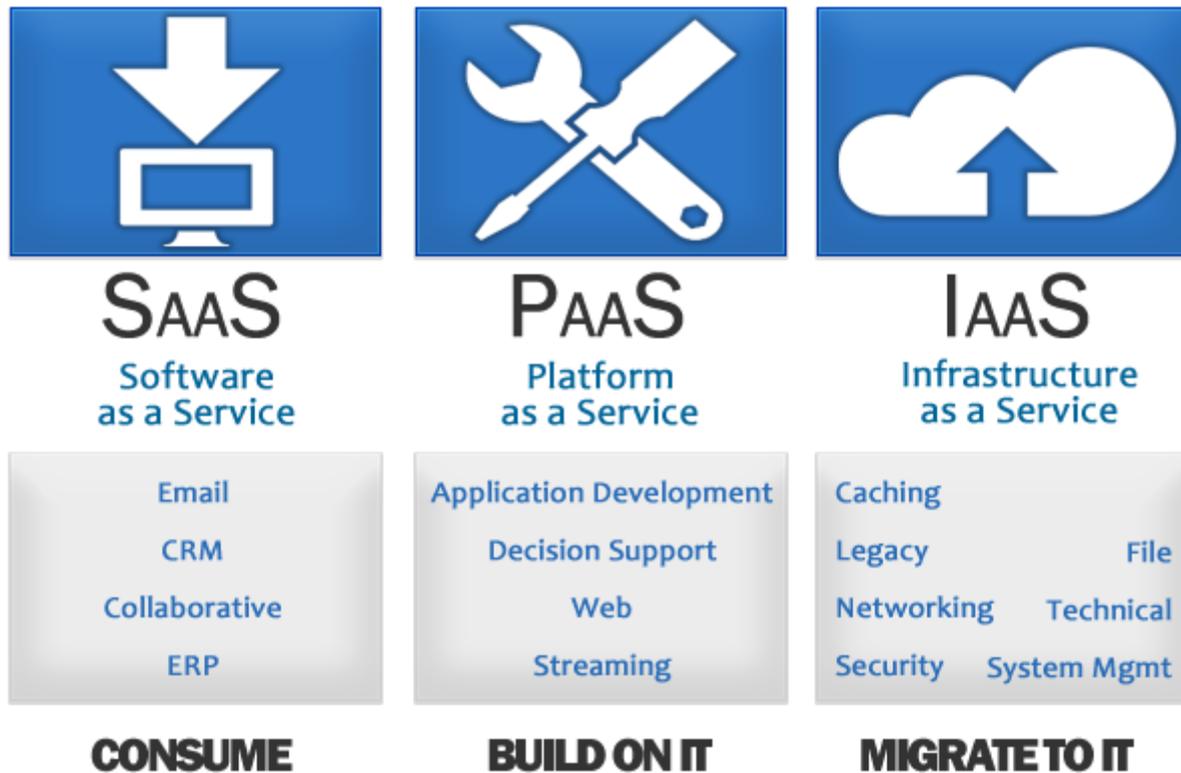


Fig.1.Types of Cloud Computing Services

## II. TYPES OF CLOUD

Following are the different types of clouds.

### A. Public Cloud

The public clouds are available for general public over the internet. It is owned, managed and operated by an organization selling cloud services.

### B. Private Cloud

The Private clouds are designed for exclusive use by a single organization. It is owned, managed and start up by the organization or third party and may exist on or off premises. Private clouds are highly control over performance, reliability and security by organization.

### C. Community cloud

The community clouds shared between several organizations and usually setup for their specific requirement. It is owned, managed and operated by the organizations or a third party and may exist on or off premises.

### D. Hybrid Cloud

Hybrid clouds are involvement of two or more cloud (public, private or community) that remain sole entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

## III. RELATED WORK

Resnick et al. [1] proposed how to trust between familiar and unfamiliar users in internet transactions information. The eBay is successful example of internet business. The internet transaction information of all familiar and unfamiliar is stored in official central server. Kamvar et al. [2][3][4] proposed the peer to peer trust. The peer to peer trust supporting a reputation based relation in electronic communities. In this paper peer to peer trust do not give full privacy just a part of it. This method directly not used in cloud environment. Cloud environment wants new cloud model to strongly operate the potential security problem. The new cloud model allows to no interruption between all user data in cloud and privacy of user data must be maintained [6] [7]. Song Xu et al. [8] proposed the new security model which secure the document for cloud computing and document separate into a different segment to handle and storing for guarantee the privacy of text document. Cong Wang et al. [9] described the problem of data security in cloud storage and get on the loyalty of cloud data integrity and availability and introduce the quality of dependable cloud services for authorized user. It gives the dynamic operations on the outsourced data, which include the block modification, append and deletion. Mukherjee

and Sahoo [10] described the algorithm for encryption of important document for C governance society and decryption to empower user by use of Hadamard matrix. Rohith Singi Reddy [12] described a new chaining technique for the use of Hadamard transforms for encryption and decryption of both binary and non binary data sequences. Abhishek Mishra et al. [13] proposed an algorithm which is very useful to increase the level of security as the types of data and we can also achieve cloud security purpose. So it is very useful to provide the storage as a service.

#### IV. SYSTEM MODEL

The representative network architecture for cloud data storage is illustrated in figure2 [8]. The different network entities can be defined as follows

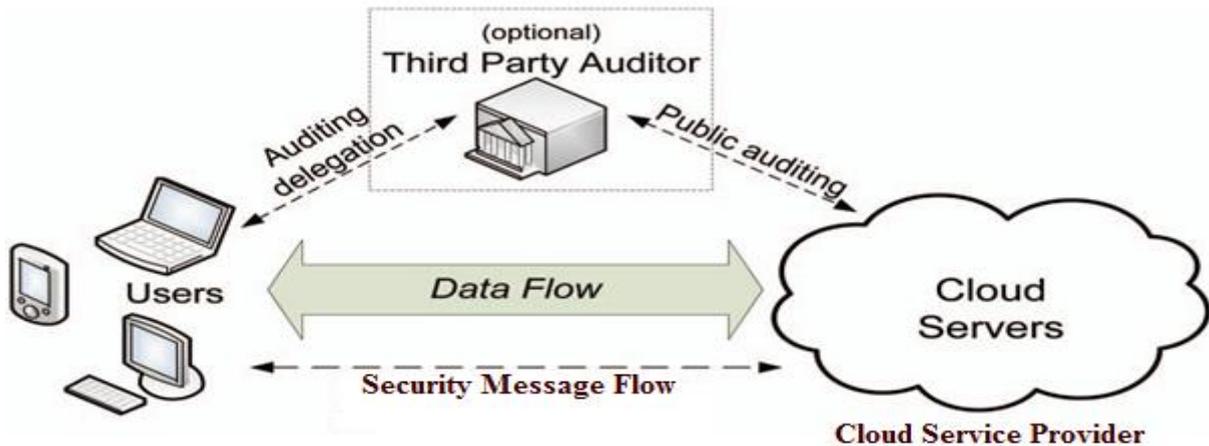


Fig.2 Cloud Data Storage Architecture

**Client:** an entity, which has large data files to be stored in the cloud and relies on the cloud for data maintenance and computation, can be either individual clients or organizations;

**Cloud Service Provider:** an entity, which is managed by Cloud Service Provider, has significant storage space and computation resources.

**Third Party Auditor:** an entity, which has attainment and capabilities that clients do not have, is trusted to assess and expose risk of cloud storage services on behalf of the clients upon request.

In cloud data storage, a client stores our data through a CSP into a set of cloud servers, which are governing in simultaneous, cooperated, and distributed fashion. Data redundancy can be employed within a technique of erasure correcting code to further tolerate faults or server crash as user's data grow in size and the importance. Thereafter, for application motive, the user interacts with the cloud servers via CSP to retrieve his data. In some cases, the user may require to perform block level operations on his data such as block updating, deletion, insertion, and append etc.

#### V. MATHEMATICAL TOOL

##### A. Hadamard matrix of order n

It is an M into M matrix having elements +1 and -1 such that any distinct row or column vectors are mutually orthogonal. A (normalized) Sylvester-Hadamard matrix [13] of size  $2^n$ ,  $n > 0$ , is a squared  $2^n$  into  $2^n$  matrix that is defined recursively by

$$H_{2^n} = \frac{1}{\sqrt{2}} \begin{pmatrix} H_{2^{n-1}} & H_{2^{n-1}} \\ H_{2^{n-1}} & -H_{2^{n-1}} \end{pmatrix}$$

Where recursion is initiated by  $H_1 = (1)$

##### B. Hadamard transform

It may be generated either recursively, or by using of binary representation. It is a generalize class of Fourier transform[16] Its symmetric form lends itself to applications ranging across many technical fields such as data encryption , signal processing , data compression algorithms, randomness measures[9][10] [11] and so on. It is given by mapping  $T: \mathbb{R}^{2^n}$  to  $\mathbb{R}^{2^n}$  defined by  $T(x) = H_{2^n} \cdot x$

##### C. Inverse Hadamard transform

The inverse of Sylvester-Hadamard matrix is equal to its transpose so inverse Hadamard transform [13] is performed by applying  $H_{2^n}^T$

So x is given as

$$x = H_{2^n}^T \cdot T(x) = H_{2^n}^T \cdot H_{2^n} \cdot x$$

#### D. Compression Technique

Compression is the reduction in size of data in order to save space or transmission time. Compression schemes are divided into two main groups:

- **Lossy compression:** It reduces the bits by identifying unnecessary information and removing it.
- **Lossless compression:** It reduces bits by identifying and eliminating statistical redundancy, no information is lost in lossless compression.

#### E. Run length Encoding

The run length encoding is a very simple compression technique that replaces runs of two or more of the same character with a number which represents the length of the run, followed by the original character; single characters are coded as runs of 1.

E.g. I/P: AABCCCB BBBBAAAAA

O/P: 2A1B3C4B6A

### VI. ALGORITHM

The encryption and decryption algorithms [11] used in the proposed model are given as follows.

#### A. Encryption Algorithm

1. In the encryption algorithm, first given binary input sequence and the key have considered. The given key can have numbers such that each number says  $s$  in  $n$  is a prime then  $2^s - 1$  is also a prime. Also, consider a two dimensional integer array. Here, the number of rows is equal to the number of elements in the array and the number of columns is equal to the number of input values at each row.
2. Next, first element in the key and group has considered the bits in the input sequence based on the number.
3. Now conversion of each group into corresponding decimal number has performed.
4. Divide the decimated sequence to equal length of sub sequences such that each sub sequence of length should be expressed as power of 2. This is reliant on the length of input sequence
5. Append 0s at last to make the length equal to other sub sequences length. If the sub-sequence length cannot be expressed as power of 2
6. Corresponding index in the array is notable with 1 if a number in the sub-sequence is equal to  $2^s - 1$  and remaining all indexes are notable with 0s which applicable for every sub-sequence
7. Represent each sub-sequence as a column matrix. Then multiply the each sub sequence matrix with the modified Hadamard matrix, and the matrix of the form modulo  $2^s - 1$  must be used to perform multiplication.
8. Perform modulo  $2^s - 1$  operation on the resultant values obtained after multiplication.
9. Each decimal number in the sequence is converted into to corresponding binary values.
10. Now, use next element in the key and group the bits of the sequence obtained from the previous step based on the number.
11. The sequence obtained after processing the last element in the key is the final encrypted message.
12. Repeat 4 to 9. The final encrypted message is compress by Run Length Encoding compression technique.

#### B. Decryption Algorithm

1. In this algorithm, first, the compressed message is decompress and form cipher message.
2. The ciphered message, sequence of key which is reversed and the array used in encryption.
3. Then, consider the first element in the key and group the bits in the encrypted sequence based on the number.
4. Now convert each group to corresponding decimal number.
5. Depending on the length of input sequence is dividing into the decimated sequence to equal length sub-sequences such that each sub-sequence length should be expressed as power of 2.
6. Represent each sub-sequence as a column matrix and multiply each sub-sequence matrix with the modified Hadamard matrix and the matrix of modulo  $2^s - 1$  must be used to perform multiplication.
7. Now, multiply two modulo matrices and find the divisor such that the resultant matrix obtained is thus represented as an Identity matrix.
8. Calculate modulo multiplicative inverse for the divisor that is,  $a*y \text{ mod } 2^s - 1 = 1$  [10] where  $y$  is the divisor and  $a$  is modulo multiplicative inverse.
9. Multiply the resultant matrix obtained in 7 with  $a$ .
10. Apply modulo  $2^s - 1$  on the resultant values obtained after multiplication.
11. For every 0 in the decimal sequence check for the corresponding array index, if the index has an element 1 then replace 0 with  $2^s - 1$ .
12. Convert each decimal number in the sequence to corresponding binary values.
13. Eliminate all successive 0s at end of the sequence in such a manner; the resultant sequence has a length equal to power of 2.
14. Now, use next element of the key and group the bits of the sequence obtained from the previous step based on the number.
15. Repeat 5 to 13. For the sequence obtained after processing the last element in the key and obtain original message.

#### C. Illustration

##### 1. Encryption

Let the input sequence (original message) is 11110010101111011000110 and Key is {3}.

To encrypt this input sequence, find the resultant matrix  $P = H8 \text{ mod } 7$ , where H8 is 8 X 8 Hadamard matrix.

At this level every 3 bits in the input sequence are grouped and corresponding decimal sequence is given as follows {3, 6, 2, 5, 7, 3, 0, 6}.

The maximum value corresponding to the input sequence is stored in the array  $A[0][0] = \{0, 0, 0, 0, 1, 0, 0, 0\}$

Multiply the resultant matrix P and decimal sequence and perform modulo 7 operations which give the sequence as {4, 6, 6, 3, 0, 3, 5, 4}.

After converting the above decimal sequence into binary, we get the encrypted sequence as {100110110011000011101100}.

Now, the encrypted sequence is compress, we get finally compress encrypted sequence as {112021102120214031102120}

## 2. Decryption

The compress encrypt message is decompress and form the encrypted message is given as {100110110011000011101100}.

Now the decimal sequence is given as {4, 6, 6, 3, 0, 3, 5, 4}.

Multiplying with the resultant matrix P and the decimal sequence, we find the sequence as follows

{31, 111, 121, 131, 91, 101, 91, 111}

Then, we find a multiplicative inverse using the formula  $a * y = 1 \text{ mod } 7$ , here  $a = 1$ .

After multiplying in the above sequence by 'a' and taking modulo 7. We get the sequence {3, 6, 2, 5, 0, 3, 0, 6}.

After comparing with stored array 5th element replaces with 7. We get the sequence {3, 6, 2, 5, 7, 3, 0, 6}.

Now, after converting it into binary, we get {110010001110111110} which is the original message.

## VII. PROPOSED MODEL

The proposed model is shown in the Fig. 3. [13]

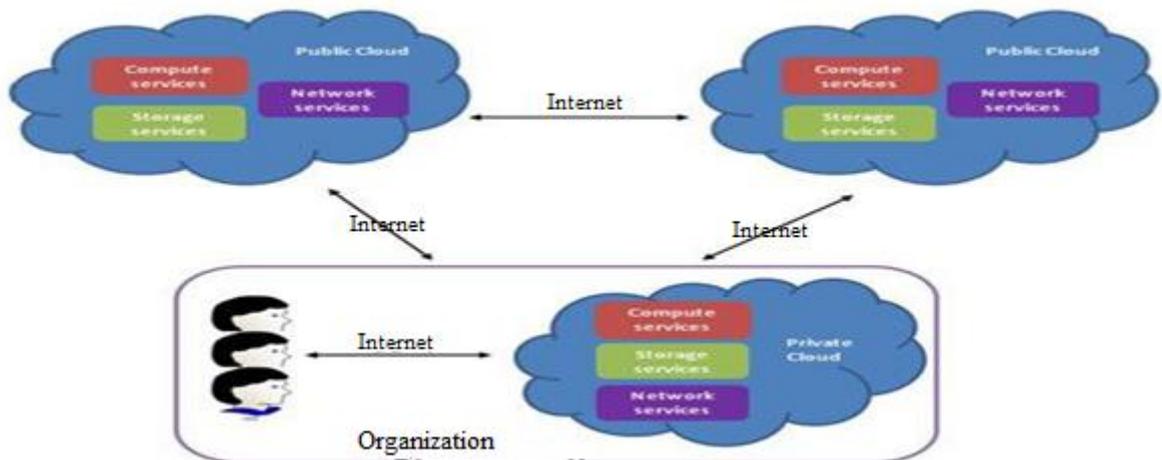


Fig.3. Proposed Model

1. In the given model, using above encryption algorithm in private cloud to encrypt the text document.
2. The text document is stored in the public cloud.
3. Only authorize user can access the original document after decryption of the stored document from the private cloud.

So in such a way cloud storage as a service is provide to the user and document is stored into the public cloud in encrypted form so unauthorized person cannot access the document.

## VIII. CONCLUSION

In this paper, we investigated the problem of text document security in cloud storage, which is essentially a distributed storage system. To ensure the correctness of users' document in cloud storage, we proposed an effective and flexible distributed strategy. This paper presents an approach to encryption of text document using a chain of Hadamard transforms. We have presented implementation of this system for binary and non-binary message sequence. An algorithm is given for both encryption and decryption. In this mechanism, the encryption and compression of text document are advantages to keep their privacy of data in cloud database. In this mechanism we improved all issues of security and reducing storing space of text document which detect the privacy of cloud data and take large storage. We want to give better approach to ensure for security of cloud data storage. The proposed model is very useful to increase the security of all documents. The databases deployed to the cloud contain critical and private information. The databases are uploaded

to the storage facility provided by the cloud service provider, who has higher priority to access the document. Since document is exposed to a third party, several security threats may occur. For the positive confidentiality of our document they are encrypted and compress before storing in to cloud. By using the proposed encryption scheme the confidentiality of our document is achieved and in the same time we solve the problem of key storage to make key more secure.

#### REFERENCES

- [1]. P. Resnick and R. Zeckhauser, "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System". *Advances in Applied Microeconomics: A Research Annual* 11, 127–157 (2002)
- [2]. D.Kamvar, T. Schlosser and Garcia-Molin, "The Eigen Trust Algorithm for Reputation Management in P2P Networks". *Proceedings of the 12th international conference on World Wide Web*, pp. 640–651. ACM, New York (2003)
- [3]. Li Xiong and Ling Liu, "Peertrust : Supporting Reputation Based Trust for Peer-to-Peer Electronic Communities". *IEEE transactions on Knowledge and Data Engineering* 16(7), 843–857 (2004)
- [4]. Runfang Zhou and Kai Hwang, "Power Trust: A Robust and Scalable Reputation System for Trusted Peer-to-Peer Computing". *IEEE Transactions on Parallel and Distributed Systems* 18(4), 460–473 (2007)
- [5]. F. John Krauthem, "Private virtual infrastructure for cloud computing", In: *Hot Cloud, USNIX* (2009)
- [6]. A. Juels and B.S. Kaliski Jr., "PORs: Proofs of Retrievability for Large Files", *Proc. 14th ACM Conf. Computer and Comm. Security (CCS '07)*, pp. 584-597, Oct. 2007.
- [7]. Jin Song Xu, Ru Cheng Huang, Wan Ming Haung and Geng Yang, "Secure Document Service for Cloud Computing", *CloudCom. Springer-Verlag Berlin Heidelberg, LNCS 5931*, pp. 541–546, 2009
- [8]. Cong Wang, Qian Wang, Kui Ren, Ning Cao, and Wenjing Lou, "Toward Secure and Dependable Storage Services in Cloud Computing", *IEEE transactions on services computing*, vol. 5, no. 2, april-june 2012
- [9]. Mukherjee and Sahoo, "Security Mechanism for C-Governance using Hadamard Matrices", *Computer and Communication Technology (ICCCT), 2011 2nd International Conference on* 15-17 Sept. 2011.
- [10]. T. Koshy, "Elementary Number Theory with Applications", 2nd Ed, Elsevier Inc., Academic Press Publications, Burlington, MA, 2007, pp. 346
- [11]. Rohith Singi Reddy, "Encryption of Binary and Non-Binary Data Using Chained Hadamard Transforms", *OK* 74078
- [12]. Katerina Tepla, "Hadamard transform" Charles University in Prague, 24th March 2012.
- [13]. <http://stack.nil.si/ipcorner/CoreCloud/#chapter2>.
- [14]. S. Kak, "Classification of random binary sequences using Walsh-Fourier analysis". *IEEE Trans. On Electromagnetic Compatibility (EMC)* 13, pp. 74-77, 1971.
- [15]. L. J. Yan and J. S. Pan, Generalized discrete fractional Hadamard transformation and its application on the image encryption, *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, IEEE Computer Press, pp. 457-460, 2007.
- [16]. K. J. Horadam, "A generalized Hadamard transform", in A. Grant (ed.) *Proceedings of the 2005 IEEE International Symposium on Information Theory*, Adelaide, Australia, 4-9 September 2005, pp. 1006-1008.