ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 5, Issue 5, May 2016

Big Data in Healthcare Using Cloud Database with Enhanced Privacy

P.M.Lavanya[1], P. Valarmathie[2]

[1]PG Scholar,
Department of Computer Science and Engineering,
Saveetha Engineering College,
Tamilnadu, India

[2]Professor,
Department of Computer Science and Engineering,
Saveetha Engineering College,
Tamilnadu, India

***Abstract*: The world's population keeps on increasing and the treatment models also gets on changing rapidly. The medical records of the patient increases and difficult to maintain these records in the local databases. In order to access the medical records of the patients from anywhere it is stored in the cloud. The information regarding the patient is extracted by the doctor from the large number of records and produces the details of the patient. The decision support system in healthcare is used to diagnose the illness of the patient based upon the symptoms. Hadoop helps in predicting the disease using the classification and clustering techniques. The records which are stored at the local database using Big data is then transferred to the cloud. The Personal Health Records (PHR) of the patients are stored in an encrypted manner by providing the unique id for the individual patient. In reference to the id, information regarding the patient is retrieved within an elapsed tolerable time from the cloud. Since the records are stored at a third party, cloud providers there exists privacy issues which may lead to unauthorized access. The security is provided to PHR's, by using Homomorphic-Based Encryption (HBE) technique which helps to encrypt all the medical records.This helps to keep the patient records to be kept confidential.**

***Keywords:* Personal Health Records, Big Data, Hadoop, Disease, Decision Support System, Security, Homomorphic Based encryption, Privacy**

I. INTRODUCTION

Big Data in Healthcare makes a lot of difference in the Healthcare industry. It is hugely a challenging experience in healthcare for diagnosing the diseasesbut at the same time it provides more meaningful information by helping clinicians for providing treatment to the patients. It helps to store the huge volume of medical records and process it within a specific time. Big Data with Cloud Computing makes the healthcare organisations to move in an effective way. Personal health records (PHR) are distributed over the cloud as they are need to be accessed from anywhere. It also contains the privacy issueswhen personal health information is exposed to the third party servers and to unauthorized parties. The patient controls the access to their own PHRs from third parties, by encryptingthe PHRs using some algorithms. The issues include the risk of exposing their records, flexibility in accessing the records, scalability and efficiencyarethe most important challenges for achieving the fine-grained, encrypted data access control. The challenge involves how to keep the data confidential and integrity when storing PHR's in cloud though it makes data available, by extracting the specific information by the medical clinicians, and share it with the collaborators, while preserving the privacy of the patients and giving them the privilege to access their data at all times.

Cloud allows the users to access their records based upon Pay-as-you-go basis technique where the users can make use of the information from the centralised database. It provides the retrieval of medical records services over the internet. By this it allows both the patients and doctors to use its services that are managed by the third parties. Though Cloud Computing offers many benefits it is vulnerable to attacks. The medical records contain the sensitive information and it is distributed all over the network. Since the data ispresent in many machines the hacker may steal the data. In order to avoid this, security is provided by encrypting the data using different encryption keys and stored centrally behind strong firewalls. By this way if the hacker gets the data it cannotbe extracted in a meaningful information and it will not be misused by the hacker. This results by keeping the patient data in an encrypted manner and provides confidentiality from unauthorized access. The records which are encrypted cannot be hacked by the hacker at any time.

II. LITERATURE REVIEWANDHYPOTHESIS

A. Literature Review:

Some years ago, data mining is used to extract the data and data warehousing for storing the data. These two technologies help the organizations for processing the larger data. The medical records of the patients are stored at local databases which may even lead to loss of data in case of failure in servers. Even some hospitals now-a-days follow the same procedure of storing the patient records in format of paper records in a separate room. This may lead to misplacement of medical records, staff making errors in replacing the files. When the records are stored in the traditional database there occurs a problem of data backup and the security issues. These kind of local database does not contain the unstructured data as they contain only the structured data. The medical records contain all types of format like structured, unstructured and semi structured data where it cannot be stored in the local database using the data mining.

B. Hypothesis:

The correct diagnosis of any illness is done by decision support system and also helps to predict the disease by the doctor based upon the symptoms. The personal details of the patient are collected for registration. Based on the registration a unique id is generated. The generated unique id is then provided to the patient. The patient approaches the doctor with the id. The Personal Health Record (PHR) of the patients are stored in the public cloud. Since the PHR contains the sensitive information each and every patient records are encryptedusing the Homomorphic based encryption. When the PHR of the patient is needed,the doctor retrieves it from the cloud by decrypting it with the key. By the retrieved results the doctor prescribes to take specific test based on symptoms. The test reports taken by the patients are provided to the doctor and updates the database. Even though the doctor is not available the patient can be referred by the other doctor using the ID as it can be retrieved from the cloud. The records are maintained securely using Homomorphic Based Encryption (HBE) technique. This results in providing the confidentiality to the data. By this way the records are kept in an encrypted manner.

III. ARCHITECTURE DIAGRAM

Fig.1 shows the disease prediction diagram in Big data using Cloud which involves the process of act between the patient, adminand a doctor through the use of Big Data and Cloud Computing. The patient visits the hospital for undergoing the necessary treatment in order to get cured quickly.First, the patient moves up to the hospital admin for the basic procedures likegetting appointments, treatment, etc. The admin gathers the patient's personal details and symptoms. After the registration, a unique id is generated. The ID is provided to the patient. With reference to the ID the patient consults the doctor.

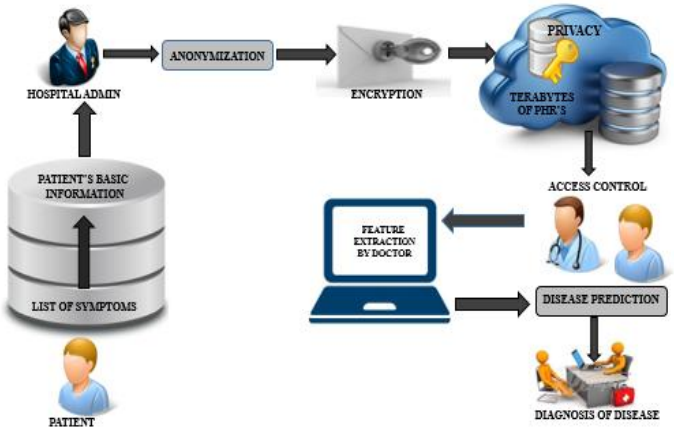


Fig.1 Disease Prediction Design in Big Data Using Cloud

By having the ID, the doctor retrieves the PHR of a particular patient and check for the details of the patient. The doctor can even update the medical record of the patient. The doctor diagnoses the disease by predicting it based on symptoms with the help of Hadoop. These all are stored in a local database using Hive. Since the details must be accessed from anywhere it is then stored in the cloud. Even though the doctor is not available the patient can be referred by the other doctor using the ID as it can be retrieved from the cloud. Since the PHR contains the sensitive information, the security is provided to keep the data confidential.The security is provided by encrypting the details of PHR by using Homomorphic Based Encryption (HBE) technique. While retrieving the medical records of patients are decrypted using some keys by the doctor. This results the records to be stored in an encrypted manner.

IV. IMPLEMENTATIONAND METHODOLOGY

A. Patient Enrolment

When the patient visiting the hospital asks the admin for further details for taking the necessary treatment. Hospitals admins are responsible for guiding the patients in order to get diagnose of disease by the doctor. In order to maintain all the departments these must perform specific actions and must contain set of skills/knowledge. The patient must register first in the hospital for their further contact with the hospital. The admin gathers the required details from the concerned patients for registration and maintains it in a local database. The admin keeps on registering the new patient's. The registration form contains the personal details of the patient like name, address, male/female, DOB, Age, mobile, height, weight, Drugs,Symptoms- High Fever/normal, Considerable weight loss, cough, yellow correlation of eyes, head ache, sweats, anaemia, vomiting, etc., The patient provide his/her personal details and the symptoms. Once the admin gathers the details of the patient, submits the form. Once registration is

completed,a unique ID is generated. After registration the PHR of the patient gets stored in the hive database. The hospitals in the world performs the same job for registration and stores it in database.The admin performs the task until the working of hospital.

B. Retrieval of Patient Record

The doctors in hospitals use todiagnose, examine and treat the patients who have been suffering from severe symptoms andalso referred to the hospital by other health professionals. They use their medical skills and knowledge in order to diagnose and manage to cure the disease. The doctor will be provided with the login details where he/she can use their account in private for providing the treatment to the patients. The doctor logs in to the screen when his/her consultation time begins. Once logged in, the doctor can able to view the new and the existing patient's information. The new patient's record contains the personal information and the symptoms. The existing patient record contains the test report values that are taken before and the disease that they have diagnosed by other doctor. The particular patient information can also be retrieved form the cloud. It displays the filled details of the patient and also the symptoms. The doctor is an authorized person to make any updates in the patient's medical record. The purpose of updating the PHR is that when the patient visits the regular doctor where he/she knows well about the patient condition and when that doctor is not available at that time the patient who consults the another doctor can provide the ID so that the temporary doctor can gain more knowledge regarding the patient's details within a few seconds. This helps the doctor to better perform with more patients in a lesser time. When the consultation time gets over, the doctor logs out the account when he /she leaves the hospital.

C. Privacy Control in Cloud Database

The PHR's which are stored in a local database are based on the window based application, as now- a-days there is a need of the information being accessed by anyone at anywhere, the PHR's are then transferred to the cloud for accessibility through online. This helps to access the PHR from anywhere and at any time. The data present in cloud are kept exposed to the third parties and to unauthorized access.Since, the PHR contain the sensitive information security must be provided to prevent access from unauthorized users. The PHR in the cloud are encrypted using Homomorphic Based Encryption (HBE). The patient records are encrypted using the different keys. When the PHR is needed by the patient or doctor it is decrypted by the key. The appointments are even scheduled by the admin or the doctor for consultation. These are all done in the cloud by creating the application and storing it. This makes the application to be used on pay-as-you-basis. The datasets are distributed centrally over the cloud in an encrypted manner.This provides the privacy to all the patient's records.

D. Disease Prediction

The doctor asks the new patients to take the prescribed test based upon symptoms. When the patient consultsthe doctor again he/she provides their unique ID by which the doctor gets the information about the patient. The test report values are taken as input values. The doctor enters the test reports values in the medical recordand updates it. From this, the doctor predicts the disease based upon the patient's unique ID and symptoms by updating the patient record. Previously the trained data is used to predict the disease based upon the symptoms. In this dataset, the symptoms are collected and categorized under some predictable diseases. Hadoop uses the classification and clustering techniques for predicting the disease. The clustering involves the k-means technique where the symptoms are grouped together for some predictable diseases. Now the patients undergo the medical test given by doctor and provide the test result to trained dataset file. Here the doctor analyses the symptoms and compare the trained dataset with it and finally predict the kind of disease. By this the doctor diagnose the patient and updates the PHR of the patient. Here are some of the symptoms and diseases that are mentioned High Fever/normal, Considerable weight loss, Cough, yellow correlation of eyes, head ache, sweats, anaemia, vomiting, swelling of joints, muscle pain and the diseases are Chikungunya, Dengue, Jaundice, Malaria, Typhoid. Once the disease is predicted the doctor diagnose the disease and provide necessary treatment to the patient.

The following process shows that how to predict the disease.
Step1: The patient registers their personal details with some basic criteria.

Condition1: Name, Date of Birth and symptoms are necessary for registration process.

Step2: An unique id is generated based upon on the continuous patient registration.

Step3: Based on the id the doctor retrieves the medical record of the patient.

Step4: The diseases which are similar are grouped together by the clustering technique.

Process:
1. Randomly pick k symptoms from the database D and consider these symptoms are cluster centers for initial clusters;
2. **Repeat**
2.1 Consider the remaining symptoms and (re) assign these symptoms to clusters based on the similarity.
2.2 Next step is the cluster mean updation process, here new mean value is calculated based on the newly entered symptoms in the clusters.
3. **Until** there is no change;

Step5: Based on the test report values the doctor predict the disease. (E.g.) Dengue disease

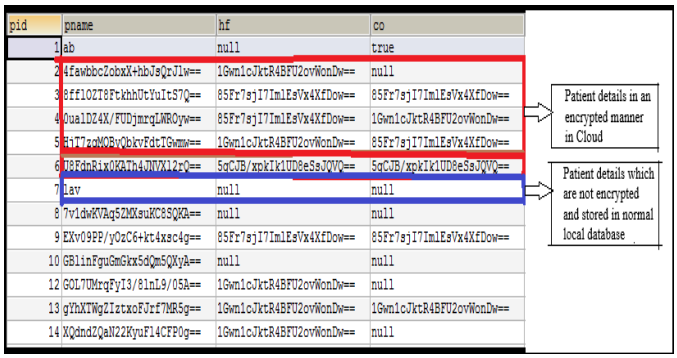
If platelets-count<100000
Then
Dengue positive
Else
Normal

V. ADVANTAGES

The proposed system overcomes the problem of delay in consulting a doctor. Terabytes of medical records can be stored using Big data and it can be accessed from anywhere at any time.It makes the information to be available up-to-date. It contains the all dimensions of Bigdata. It provides the ease of gathering/processing of particular information (for example, patient management and data entry). Since the medical records are stored in different nodes it is said to be fault tolerant.The medical records are stored in an encrypted manner and it provides the confidentiality to each and every record.

VI. DATA ANALYSISAND RESULTS

Table.1 shows the medical records of the patients which are analysed and comparedbetween the existing and the proposed system. Fig.2 shows the number of patients consulted by a doctor is less and takes more time for diagnosing them whereas in the Fig.3 the number of patients are consulted by the doctor are more and takes less time by the doctor for diagnosing the disease. The unstructured data is not supported in the data mining whereas the unstructured data is supported by using Big data in cloud technique. The processing speed takes more time in the data mining whereas it takes less time to retrieve the records of the patient in a less time. The storage capacity is less using data mining whereas it is high using Big Data in cloud. It contains less security using data mining but it is highly secured using Bid data in cloud by the means of encrypting the medical records in the cloud. With the help of Big Data, the datasets of the patients are easily retrieved from the large number of records. The medical records are easily accessible from anywhere as it contains in the cloud for the proposed system. By this the performance is said to be high using Big data in cloud than without using the Big Data in cloud. This makes to predict the disease in a more effective manner by the doctor.



pid	pname	hf	co
1	ab	null	true
2	4fawbbcZobxX+hbJsQrJlw==	1GwnlcJktR4BFU2ovWonDw==	null
3	8ff1OZT8FtkhhUtYuIt57Q==	85Fr7sjI7ImlEsVx4XfDow==	85Fr7sjI7ImlEsVx4XfDow==
4	0ualDZ4X/FUDjmrqLNROyw==	85Fr7sjI7ImlEsVx4XfDow==	1GwnlcJktR4BFU2ovWonDw==
5	H1T2aMOBvQbkvFdtIGwmw==	1GwnlcJktR4BFU2ovWonDw==	85Fr7sjI7ImlEsVx4XfDow==
6	I8FdnRix0KATh4JNVX12rQ==	5qCJB/xpkIk1UD8eSaJQVQ==	5qCJB/xpkIk1UD8eSaJQVQ==
7	lav	null	null
8	7vldwKVAq5ZMXsuKC8SQKA==	null	null
9	EXv09PP/yOzC6+kt4xsc4g==	85Fr7sjI7ImlEsVx4XfDow==	85Fr7sjI7ImlEsVx4XfDow==
10	GBlinFguGmGkx5dQm5QXyA==	null	null
12	GOL7UMrqFyI3/8lnL9/05A==	1GwnlcJktR4BFU2ovWonDw==	null
13	gYhXTWgZIztxoFJrf7MR5g==	1GwnlcJktR4BFU2ovWonDw==	1GwnlcJktR4BFU2ovWonDw==
14	XQdndZQaN22KyuF14CFP0g==	1GwnlcJktR4BFU2ovWonDw==	null

Fig.2 Encrypted Data Stored in Database

Fig.2 shows the medical information are kept encrypted in the database using the Homomorphic Based Encryption (HBE). The information when registered by the admin are encrypted using AES technique. This provides the confidentiality to the data by which the data is not accessible by the unauthorized people. When the information is needed it is decrypted by the secret key.

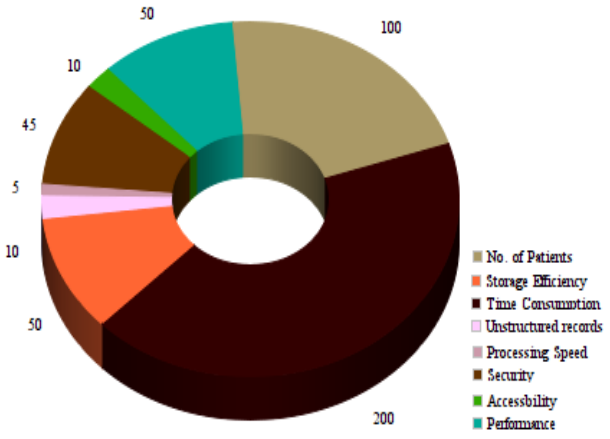


Fig.3 Comparison Diagram Based on Some Medical Constraints Without using Big Data and Cloud

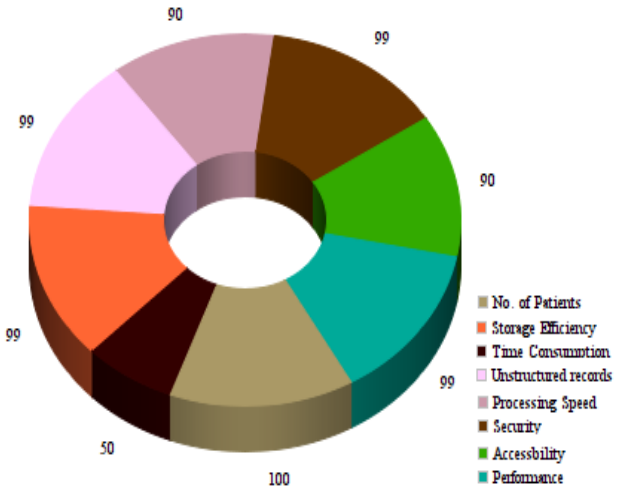


Fig.4Comparison Diagram Based on Some Medical Constraints using Big Data and Cloud

The following table 1 provides an example by considering 100 patients and the time consumed by the doctor for providing the treatment to the patient. The records which are accessible by the authorized people. It also discusses on the storage efficiency by comparing with existing and the proposed system. It clearly gives a comparative analysis between using data mining and the Big data in cloud technique.

CONSTRAINTS	USAGE WITH DATAMING	USAGE WITH BIGDATA IN CLOUD
Storage Efficiency	50	99
Unstructured records	10	99
Processing Speed	5	90
Security	45	99
Accessibility	10	90
Performance	50	99
No. of Patients	100	100
Time Consumption	200	50

Table.1 Comparative Analysis

VII. CONCLUSION AND FUTURE WORK

From the above discussion, an approach is proposed for predicting the disease based upon the symptoms with the use of Big Data. The personal health records of the patients which are retrieved from the cloud helps the doctors by easily diagnosing the disease with in a less time.This is achieved by using the Big Data with cloud.In case if the records are needed by any doctor or patient it can be accessed from anywhere as it is stored in the cloud by encrypting the medical records from unauthorized access.

As a future work it can be implemented by using the sensors in the patient body,and getting the necessary details about their healthand storing that information in the cloud database using Big data.

REFERENCES

[1]Ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.
[2]Abdur Rahim Mohammad Forkan, Member IEEE, Ayman Ibaida, Member, IEEE(2015), "*BDCaM: Big Data for Context-aware Monitoring* - A Personalized Knowledge Discovery Framework for Assisted Healthcare".
[3]Chen and Hui Yang*, Member (2014)," Heterogeneous*Postsurgical Data Analytics for Predictive Modelling of Mortality Risks in Intensive Care Units".*
[4]Marco Viceconti, Peter Hunter and Rod Hose (2015), "*Big data, big knowledge: Big Data for Personalized Healthcare*".
[5] Xiaoyong Xu, Fen Yu, Jiannong Shi, Institute of Psychology, "*PersonalityAnd Abusive Supervision: A Study on Leadership in the Healthcare Industry in China"-2010 3rd International Conference on Biomedical*Engineering and Informatics (BMEI 2010).
[6] Keith Feldman University of Notre Dame, USA Nitesh V. Chawla University of NotreDame, USA (2014)," *Scaling Personalized Healthcare with Big data*".
[7] V. Rocha1, 3, P. Borza 4, J. Correia1, G. Gonçalves3, A.Puscas4 R. Seromenho 5, A. Mascioletti2, A. Picano2, S. Cocorada 4 M.Carp 4 INOVAMAIS 4 Transilvania University of Brasov, Electronics and Computers department," Wearable*computing for patients with coronary diseases: Gathering efforts bycomparing methods*".
[8] Venkata Narasimha Inukollu1, Sailaja Arsi 1 and Srinivasa Rao Ravuri 31 Department of Computer Engineering, Texas Tech University, USAdepartment Of Banking and Financial Services, Cognizant Technology (2013), "*Security Issues associated with big data in cloud computing*" - International Journal Network Security & Its Applications (IJNSA), Vol.6, No.3.
[9] P. Groves, B. Kayyali, D. Knott, and S. Van Kuiken, (2013), "*The big data Revolution in Healthcare*," McKinsey & Company.
[10] GramHealth : *A bottom - up approach to provide preventive healthcare services for unreached community* Ashir Ahmed, LutfeKabir, Eiko Kai and Sozo Inoune.
[11] Abdur Rahim Mohammad Forkan, Member IEEE, Ayman Ibaida, Member, IEEE (2015), "*BDCaM: Big Data for Context-aware Monitoring - A Personalized Knowledge Discovery Framework for Assisted Healthcare".*
[12] A. Pantelopoulos and N. Bourbakis, (2010), "*A survey on wearable sensor-based for health monitoring and prognosis*" IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 40, no. 1, pp. 1–12.
[13] S. Keehan, A. Sisko, C. Truffer, S. Smith, C. Cowan, J. Poisal, M. K. Clemens National Health Expenditure Accounts Projections Team, (2008),"*Health and spending projections through 2017: the baby Boom generation is coming to Medicare*" Health Affairs, vol. 27, pp. w145155.
[14] Security-Enhanced Linux (2013)," *Security-Enhanced Linux*". N.p. Web. 13.
[15] Securing Big Data: " Security Recommendations for Hadoop and NoSQL Environment ", Securosis blog, version 1.0 (2012).
[16] B. Wixom, T. Ariyachandra, D. Douglas, M. Goul, B. Gupta, L. Iyer, U. Kulkarni, J. G. Mooney, G. Phillips-Wren, and O. Turetken, "The Current State of Business Intelligence in Academia: The Arrival of Big Data," Communications of the Association for Information Systems, vol. 34, no. 1, pp. 1, 2014.
[17] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh,and A. H. Byers, "Big data : The next frontier for innovation, competition, and productivity," 2011.

[18] Austin, P. C., Tu, J. V., Ho, J. E., Levy, D., & Lee, D. S. (2013).Using Methods from the data- mining and machine - learning literaturefor disease classification and prediction : a case study examining classification of heart failure subtypes. Journal ofclinicalepidemiology.
[19] Chawla, N. V., & Davis, D. A. (2013). Bringing big data personalized Healthcare : A patient - centered framework. Journal of general internal medicine, 28 (3), 660–665.
[20] C. Wang, Q. Wang, K. Ren, N. Cao, W. Lou, "Toward Secure and Dependable Storage Services in Cloud Computing,"IEEE Transactions on Services Computing, 5(2), pp. 220-232, 2012.
[21] Y. Zhu, G.J. Ahn, H. Hu, S.S. Yau, H.G. An, S. Chen, "Dynamic Audit Services for Outsourced Storages in Clouds," IEEE Transactions on Services Computing, 2011
http://doi.ieeecomputersociety.org/10.1109/TSC.2011.51
[22] O. Goldreich, "Foundations of Cryptography: Basic Tools",Publishing House of Electronics Industry, Beijing, 2003, pp. 194-195.
[23] D. Boneh, M. Franklin, "Identity-based Encryption from the Weil Pairing", CRYPTO 2001, LNCS 2139, 2001, 213-229.
[24] A. Miyaji, M. Nakabayashi, S. Takano "New Explicit Conditions of Elliptic Curve Traces for FR-reduction", IEICE Transactions Fundamentals, 5, pp. 1234-1243, 2001.
[25] D. Boneh, B. Lynn, H. Shacham, "Short Signatures from the Weil Pairing", ASIACRYPT 2001, LNCS 2248, pp. 514-532, 2001.