Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 113 (2017) 73–80

Procedia Computer Science

www.elsevier.com/locate/procedia

The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017)

Big data security and privacy in healthcare: A Review

Karim ABOUELMEHDI[a] [1], Abderrahim BENI-HSSANE[a], Hayat KHALOUFI[a], Mostafa SAADI[b]

[a]*Department of computer science Laboratory LAMAPI and LAROSERI*
Chouaib doukali University El Jadida Morocco
[b]*Ecole Nationale des Sciences Appliquée(ENSA) de Khouribga laboratry IPOSI Université Hassan 1er - Settat, Morocco*

Abstract

The ever-increasing integration of highly diverse enabled data generating technologies in medical, biomedical and healthcare fields and the growing availability of data at the central location that can be used in need of any organization from pharmaceutical manufacturers to health insurance companies to hospitals have primarily make healthcare organizations and all its sub-sectors in face of a flood of big data as never before experienced. While this data is being hailed as the key to improving health outcomes, gain valuable insights and lowering costs, the security and privacy issues are so overwhelming that healthcare industry is unable to take full advantage of it with its current resources. Managing and harnessing the analytical power of big data, however, is vital to the success of all healthcare organizations. It is in this context that this paper aims to present the state-of-the-art security and privacy issues in big data as applied to healthcare industry and discuss some available data privacy, data security, users' accessing mechanisms and strategies.

Peer-review under responsibility of the Conference Program Chairs.

Keywords: Big Data, Privacy and security , Big data in healthcare , privacy preserving

[1]Corresponding author karim.abouelmehdi1@gmail.com

1877-0509
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2017.08.292

1. Introduction

The current trend toward digitizing healthcare workflows and moving to electronic patient records has seen a paradigm shift in the healthcare industry. The quantity of clinical data that are available electronically will be then dramatically increased in terms of complexity, diversity and timeliness, resulting what is known as big data. Driven by mandatory requirements and the potential to improve care, save lives and lower costs, big data hold the promise of supporting a wide range of unprecedented opportunities and use cases, including these key examples: clinical decision support, health insurance, disease surveillance, population health management, adverse events monitoring, and treatment optimization for diseases affecting multiple organ systems[1,2]

Even though the adoption of big data technologies in healthcare sector carries many benefits and promises, it raises also some barriers and challenges. Indeed, the concerns over the sensitive information security and privacy are increased year by year because of several growing trends in healthcare, such as clinician mobility and wireless networking, health information exchange, cloud computing and so on. Moreover, healthcare organizations found that a reactive, bottom-up, technology-centric approach to determining security and privacy requirements is not adequate to protect the organization and its patients [3].

To prevent breaches of sensitive information and other types of security incidents, a proactive, preventive approach and measures must be taken by every healthcare organization with attention to future security and privacy needs.

In this paper, we will discuss some successful and interesting related works. We will also present risks to the security of health data and discuss some newer technologies and redressal of these risks using new techniques. Then, we will focus on the privacy issue in healthcare, and mention various laws and regulations established by different regulatory bodies as well as some feasible methods and techniques used to ensure the patient's privacy. Finally, we will present some limitations of the proposed techniques and approaches to deal with security and privacy risks before concluding the paper and highlighting the future work that we plan to carry on.

2. Related works

Seamless integration of highly diverse big data technologies in healthcare can facilitate faster and safer throughput of patients, enable to gain deeper insights into the clinical and organizational processes, create greater efficiencies and help improve patient flow, safety, quality of care and the overall patient experience meanwhile contain costs.

Such was the case with the UNC Health Care (UNCHC), which is a not-for-profit integrated healthcare system in North Carolina that has implemented a new system allowing clinicians to rapidly access and analyze unstructured patient data using natural-language processing. In fact, UNCHC has accessed and analyzed huge quantities of unstructured content contained in patient medical records to extract insights and predictors of readmission risk for timely intervention and providing safer care for high-risk patients and reduce re-admissions. [4]

An other example in United States is the Indiana Health Information Exchange, which is a non-profit organization, provides a secure and robust technology network of health information linking more than 90 hospitals, community health clinics, rehabilitation centers and other healthcare providers in Indiana. It allows medical information to follow the patient hosted in one doctor office or only in a hospital system. [5]

One more example, is Kaiser Permanente medical network based in California, which has more than 9 million members, estimated to manage large volumes of data ranging from 26,5 Petabytes to 44 Petabytes. [6]

Big data analytics is used also in Canada, e.g. the infant hospital of Toronto. This hospital succeeded to improve the outcomes for newborns prone to serious hospital infections.

In Europe this time, exactly in Italy, the Italian Medicines agency collects and analyzes a large amount of clinical data concerning expensive new medicines as part of a national profitability program. Based on the results, it may reassess the medicines prices and market access terms. [7]

After Europe, Canada, Australia, Russia, and Latin America, Sophia Genetics [8], global leader in Data-Driven Medicine,announced at the recent 2017 Annual Meeting of the American College of Medical Genetics and Genomics (ACMG) that its artificial intelligence has been adopted by African hospitals to advance patient care across the continent.

In Morocco for instance, PharmaProcess in Casablanca, ImmCell, The Al Azhar Oncology Center and The Riad Biology Center in Rabat are some medical institutions at the forefront of innovation that have started integrating Sophia to speed and analyze genomic data to identify disease-causing mutations in patients' genomic profiles, and decide on the most effective care. As new users of SOPHiA, they become part of a larger network of 260 hospitals in 46 countries

that share clinical insights across patient cases and patient populations, which feeds a knowledge-base of biomedical findings to accelerate diagnostics and care [9].

While the automations have lead to improved patient care workflow and reduced costs, it is also rising healthcare data to increased probability of security and privacy breaches. In 2016, CynergisTek has released the Redspin's 7th annual Breach Report: Protected Health Information (PHI)[10] in which it has reported that hacking attacks on healthcare providers was increased 320% in 2016, and that 81% of records breached in 2016 resulted from hacking attacks specifically. Additionally, ransomware, defined as a type of malware that encrypts data and holds it hostage until a ransom demand is met, has identified as the most prominent Threat to Hospitals.

These findings point to a pressing need for providers to take a much more proactive and comprehensive approach to protecting their information assets and combating the growing threat that cyber attacks present to healthcare.

3. Big data security in healthcare

Healthcare organizations store, maintain and transmit huge amounts of data to support the delivery of efficient and proper care. Nevertheless, securing these data has been a daunting requirement for decades. Complicating matters, the healthcare industry continues to be one of the most susceptible to publicly disclosed data breaches. In fact, attackers can use data mining methods and procedures to find out sensitive data and release it to public and thus data breach happens. While implementing security measures remains a complex process, the stakes are continually raised as the ways to defeat security controls become more sophisticated.

As a result, it is crucial that organizations implement healthcare data security solutions that will protect important assets while also satisfying healthcare compliance mandates.

Technologies in use

Various technologies are in use for protecting the security and privacy of healthcare data. Most widely used technologies are:

1) **Authentication**: Authentication is the act of establishing or confirming claims made by or about the subject are true and authentic. It serves a vital function within any organization: securing access to corporate networks, protecting the identities of users, and ensuring that a user is who he claims to be.

Most cryptographic protocols include some form of endpoint authentication specifically to prevent man-in-the-middle (MITM) attacks. For instance,[11] Transport Layer Security (TLS) and its predecessor, Secure Sockets Layer (SSL), are cryptographic protocols that provide security for communications over networks such as the Internet. TLS and SSL encrypt the segments of network connections at the Transport Layer end-to-end. Several versions of the protocols are in widespread use in applications like web browsing, electronic mail, Internet faxing, instant messaging and voice-over-IP (VoIP). One can use SSL or TLS to authenticate the server using a mutually trusted certification authority. Additionally, Bull Eye algorithm can be used for monitoring all sensitive information in 360° . This algorithm has been used to make sure data security and manage relations between original data and replicated data. It is also allowed only authorized person to read or write critical data. Paper [24] proposes a novel and a simple authentication model using one time pad algorithm. It provides removing the communication of passwords between the servers. In a healthcare system, both healthcare information offered by providers and identities of consumers should be verified at the entry of every access.

2) **Encryption**: Data encryption is an efficient means of preventing unauthorized access of sensitive data. Its solutions protect and maintain ownership of data throughout its lifecycle — from the data center to the endpoint (including mobile devices used by physicians, clinicians, and administrators) and into the cloud. Encryption is useful to avoid exposure to breaches such as packet sniffing and theft of storage devices.

Healthcare organizations or providers must ensure that encryption scheme is efficient, easy to use by both patients and healthcare professionals, and easily extensible to include new electronic health records. Furthermore, the number of keys hold by each party should be minimized.

Although various encryption algorithms have been developed and deployed relatively well (RSA, Rijndael, AES and RC6 [20, 22, 23] , DES, 3DES,RC4 [21], IDEA, Blowfish …), the proper selection of suitable encryption algorithms to enforce secure storage remains a difficult problem.

3) **Data Masking**: Masking replaces sensitive data elements with an unidentifiable value, but is not truly an encryption technique so the original value cannot be returned from the masked value. It uses a strategy of de-identifying the data sets or masking personal identifiers such as name, social security number and suppressing or generalizing quasi-

identifiers like data-of-birth and zip-codes. Thus, data masking is one of the most popular approach to live data anonymization. k-anonymity first proposed by Swaney and Samrati [12,13] protects against identity disclosure but failed to protect against attribute disclosure. Truta et al.[14] have presented p-sensitive anonymity that protects against both identity and attribute disclosure. Other anonymization methods fall into the classes of adding noise to the data, swapping cells within columns and replacing groups of k records with k copies of a single representative. These methods have a common problem of difficulty in anonimyzing high dimensional data sets [15, 16].
A significant benefit of this technique is that the cost of securing a big data deployment is reduced. As secure data is migrated from a secure source into the platform, masking reduces the need for applying additional security controls on that data while it resides in the platform.
4) **Access Control**: Once authenticated, the users can enter an information system but their access will still be governed by an access control policy which is typically based on the privilege and right of each practitioner authorized by patient or a trusted third party. It is then, a powerful and flexible mechanism to grant permissions for users. It provide sophisticated authorization controls to ensure that users can perform only the activities for which they have permissions, such as data access, job submission, cluster administration, etc.
A number of solutions have been proposed to address the security and access control concerns. Role-Based Access Control (RBAC) [17] and Attribute-Based Access Control (ABAC)[18,19] are the most popular models for EHR. RBAC and ABAC have shown some limitations when they are used alone in medical system. Paper [25] proposes also a cloud-oriented storage efficient dynamic access control scheme cipher text based on the CP-ABE and symmetric encryption algorithm (such as AES). To satisfy requirements of fine-grained access control yet security and privacy preserving, we suggest adopting technologies conjunction of other security techniques, e.g. encryption, and access control method.

4. Big data privacy in healthcare

Recent years have seen the emergence of advanced persistent threats, targeted attacks against information systems, whose main purpose is to smuggle recoverable data by the attacker. Therefore, invasion of patient privacy is considered as a growing concern in the domain of big data analytics, which make organizations in challenge to address these different complementary and critical problems. In fact, data security governs access to data throughout the data lifecycle while data privacy sets this access based on privacy policies and laws which determine, for example, who can view personal data, financial, medical or confidential information. An incident reported in the Forbes magazine raises an alarm over patient privacy [26]. In the report, it mentioned that Target Corporation sent baby care coupons to a teen-age girl unbeknown to her parents. This incident impels big data to consider privacy for analytics and developers should be able to verify that their applications conform to privacy agreements and that sensitive information is kept private regardless of changes in the applications and/or privacy regulations.
Privacy of medical data is then an important factor which must be seriously considered.

Data protection laws

More than ever it is crucial that healthcare organizations manage and safeguard personal information and address their risks and legal responsibilities in relation to processing personal data, to address the growing thicket of applicable data protection legislation. Different countries have different policies and laws for data privacy. Data protection regulations and laws in some of the countries along with salient features are listed in the Table below.

Table 1: Data protection laws in some of the countries

Country	Law	Salient Features
U.S.A	HIPAA Act Patient Safety and Quality Improvement Act (PSQIA) HITECH Act	Requires the establishment of national standards for electronic health care transactions. Gives the right to privacy to individuals from age 12 through 18. Signed disclosure from the affected before giving out any information on provided health care to anyone, including parents. Patient Safety Work Product must not be disclosed [27]. Individual violating the confidentiality provisions is subject to a civil penalty. Protect security and privacy of electronic health information.

EU	Data Protection Directive	Protect people's fundamental rights and freedoms and in particular their right to privacy with respect to the processing of personal data. [29]
Canada	Personal Information Protection and Electronic Documents Act ('PIPEDA')	Individual is given the right to know the reasons for collection or use of personal information, so that organizations are required to protect this information in a reasonable and secure way.[28]
UK	Data Protection Act (DPA)	Provides a way for individuals to control information about themselves. Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects.
Morocco	The 09-08 act, dated on 18 February 2009	Protects the one's privacy through the establishment of the CNDP authority by limiting the use of personal and sensitive data using the data controllers in any data processing operation. [30]
Russia	Russian Federal Law on Personal Data	Requires data operators to take "all the necessary organizational and technical measures required for protecting personal data against unlawful or accidental access".
India	IT Act and IT (Amendment) Act	Implement reasonable security practices for sensitive personal data or information. Provides for compensation to person affected by wrongful loss or wrongful gain. Provides for imprisonment and/or fine for a person who causes wrongful loss or wrongful gain by disclosing personal information of another person while providing services under the terms of lawful contract.
Brazil	Constitution	The intimacy, private life, honour and image of the people are inviolable, with assured right to indenization by material or moral damage resulting from its violation.
Angola	Data Protection Law (Law no. 22/11of 17 June)	With respect to sensitive data processing, collection and processing is only allowed where there is a legal provision allowing such processing and prior authorization from the APD is obtained

5. Privacy preserving methods in big data

Few traditional methods for privacy preserving in big data is described in brief here. Although These techniques are used traditionally to ensure the patient's privacy[31,32,33], their demerits led to the advent of newer methods.

A. De-identification:

a traditional method to prohibit the disclosure of confidential information by rejecting any information that can identify the patient, either by the first method that requires the removal of specific identifiers of the patient or by the second statistical method where the patient verifies himself that enough identifiers are deleted. Nonetheless, an attacker can possibly get more external information assistance for de-identification in the big data. As a result, de-identification is not sufficient for protecting big data privacy. It could be more feasible if develop efficient privacy-preserving algorithms to help mitigate the risk of re-identification. The concepts of k-anonymity [34, 36, 37], l-diversity [35, 36, 38] and t-closeness [34, 38] have been introduced to enhance this traditional technique.

- ***k-anonymity***: In this technique, higher the value of k, lower will be the probability of re-identification. However, it may lead to distortions of data and hence greater information loss due to k-anonymization. Furthermore, in k-anonymization, if the quasi-identifiers containing data are used to link with other publicly available data to identify individuals, then the sensitive attribute (like Disease) as one of the identifier will be revealed. Various measures

have been proposed to quantify information loss caused by anonymization, but they do not reflect the actual usefulness of data [39, 40].

- ***L-diversity:*** It is a form of group based anonymization that is utilized to safeguard privacy in data sets by diminishing the granularity of data representation. This model (Distinct, Entropy, Recursive) [34,36, 41] is an extension of the k-anonymity which utilize methods including generalization and suppression to reduces the granularity of data representation in a way that any given record maps onto at least k different records in the data. The l-diversity model handles a few of the weaknesses in the k-anonymity model in which protected identities to the level of k-individuals is not equal to protecting the corresponding sensitive values that were generalized or suppressed. The problem with this method is that it depends upon the range of sensitive attribute. If want to make data L-diverse though sensitive attribute has not as much as different values, fictitious data to be inserted. This fictitious data will improve the security but may result in problems amid analysis. As a result, L-diversity method is also a subject to skewness and similarity attack [41] and thus can't prevent attribute disclosure.
- ***T-closeness:*** is a further improvement of l-diversity group based anonymization. The t-closeness model(Equal/Hierarchical distance) [34, 38] extends the l-diversity model by treating the values of an attribute distinctly by taking into account the distribution of data values for that attribute. The main advantage of this technique is that it intercepts attribute disclosure, and its problem is that as size and variety of data increases, the odds of re-identification too increases.

B. HybrEx

Hybrid execution model [42] is a model for confidentiality and privacy in cloud computing. It utilizes public clouds only for an organization's non-sensitive data and computation classified as public, i.e., when the organization declares that there is no privacy and confidentiality risk in exporting the data and performing computation on it using public clouds, whereas for an organization's sensitive, private data and computation, the model executes their private cloud. Moreover, when an application requires access to both the private and public data, the application itself also gets partitioned and runs in both the private and public clouds. It considers data sensitivity before a job's execution and provides integration with safety.

The problem with HybridEx is that it does not deal with the key that is generated at public and private clouds in the map phase and that it deals with only cloud as an adversary[43].

C. Identity based anonymization

These techniques encountered issues when successfully combined anonymization, privacy protection, and big data techniques [44] to analyze usage data while protecting the identities of users.

To meet the significant benefits of Cloud storage [45], Intel created an open architecture for anonymization [44] that allowed a variety of tools to be utilized for both de-identifying and re-identifying web log records. In the implementing architecture process , enterprise data has properties different from the standard examples in anonymization literature [46]. Intel also found that in spite of masking obvious Personal Identification Information like usernames and IP addresses, the anonymized data was defenseless against correlation attacks. After exploring the trade-offs of correcting these vulnerabilities, they found that User Agent information strongly correlates to individual users. This is a case study of anonymization implementation in an enterprise, describing requirements, implementation, and experiences encountered when utilizing anonymization to protect privacy in enterprise data analyzed using big data techniques. This investigation of the quality of anonymization used k-anonymity based metrics. At the same time, learned that anonymization needs to be more than simply masking or generalizing certain fields—anonymized datasets need to be carefully analyzed to determine whether they are vulnerable to attack.

6. Discussion

In this paper, we have investigated the security and privacy challenges in big data, by discussing some existing approaches and techniques for achieving security and privacy in which healthcare organizations are likely to be highly beneficial. This is by no means an exhaustive list. In this section, we focused on citing some approaches and techniques presented in different papers with emphasis on their focus and limitations.

For instance, the approach suggested in [48], which introduced an efficient and privacy-preserving cosine similarity computing protocol, needs significant research efforts for addressing unique Privacy issues in some specific big data analytics. Another research paper [44] discusses experiences and issues encountered when successfully combined anonymization, privacy protection, and Big Data techniques to analyze usage data while protecting the identities of

users. But, it still uses K-anonymity technique which is vulnerable to correlation attack. Beside these papers, [49] proposed a scalable two-phase top-down specialization (TDS) approach to anonymize large-scale data sets using the Map Reduce framework on cloud. Nevertheless, it uses anonymization technique which is vulnerable to correlation attack. [50] That proposed various privacy issues dealing with big data applications, is also limited, because customer segmentation and profiling can easily lead to discrimination based on age gender, ethnic background, health condition, social, background, and so on. Additionally, [51] proposed an anonymization algorithm (FAST) to speed up anonymization of big data streams, where further research required, to design and implement and it must be implemented in a distributed cloud-based framework in order to gain cloud computation power and achieve high scalability. As the final discussed technique in this section, [52] proposed a methodology that provides data confidentiality, secure data sharing without Re-encryption, and access control for malicious insiders, and forward and backward access control that is limiting the trust level in the cryptographic server (CS).

These increased complexity and limits make the new models more difficult to interpret and their reliability less easy to assess, compared to previous models.

7. Conclusion

Limitless opportunities are offered for big data to drive health research, knowledge discovery, clinical care, and personal health management. However, there are a number of obstacles and challenges that impede its true potential in the healthcare field, including technical challenges, privacy and security issues and skilled talent. Big data security and privacy are considering as the huge barrier for researchers in this field.

In this paper, we have discussed some examples of successful related work across the world. Privacy and security issues in each phase of big data life cycle are also presented along with the advantages and disadvantages of existing privacy and security technologies in the context of big healthcare data.

In addition of the methods currently used to ensure patient's privacy in healthcare that we have presented, there are more various techniques include Hiding a needle in a haystack [47], Attribute based encryption Access control, Homomorphic encryption, Storage path encryption and so on. However, the problem is always imposed.

In this context, as our future direction, perspectives will focus more on achieving effective solutions to the scalability problem of big data privacy and security in the era of healthcare. And to go further, we will try to solve the problem of reconciling security and privacy models by simulating diverse approaches using exploiting the MapReduce framework. To, ultimately, support decision making and planning strategies.

References

1. Burghard C: Big Data and Analytics Key to Accountable Care Success. IDC Health Insights; 2012.
2. Fernandes L, O'Connor M, Weaver V: Big data, bigger outcomes. J AHIMA 2012:38–42.
3. David Houlding, MSc, CISSP: « Health Information at Risk: Successful Strategies for Healthcare Security andPrivacy » Healthcare IT Program Of ce Intel Corporation, white paper 2011.
4. "UNC Health Care relies on analytics to better manage medical data and improve patient care." IBM press release. October 11, 2013.
5. Indiana Health Information Exchange: http://www.ihie.org/ (Accessed Date: March 24, 2016).
6. Transforming Healthcare through Big Data, Strategies for leveraging big data in the health care industry. Institute for health- 2013
7. The Big Data revolution in healthcare, accelerating value and innovation – Peter Groves, Basel Kayyali, David Knott , Steve Van Kuiken –2013
8. Sophia Genetics: « Product & Technology Overview » 2014
9. Sophia Genetics: http://www.sophiagenetics.com/news/media-mix/details/news/african-hospitals-adopt-sophia-artificial-intelligence-to-trigger-continent-wide-healthcare-leapfrogging-movement.html (March 24, 2017)
10. CynergisTek, Redspin : « BREACH REPORT 2016: Protected Health Information (PHI)» February 2017
11. Rui Zhangand Ling Liu: '' Security Models and Requirements for Healthcare Application Clouds'' in IEEE 3rd International Conference on Cloud Computing, 2010
12. L. Sweeney, "Achieving k-anonymity privacy protection using gener- alization and suppression," in International journal on uncertainity, fuzziness and knowledge based systems, vol. 10, 2002, pp. 571 – 588.
13. P. Samrati, "Protecting respondents identities in microdata release," in IEEE transactions on knowledge and data engineering, vol. 13, 2001, pp. 1010 – 1027.
14. T. M. Truta and B. Vinay, "Privacy protection: p-sensitive k-anonymity property," in Proceedings of 22nd International Conference on Data Engineering Workshops, 2006, p. 94.
15. N. Spruill, "The confidentiality and analytic usefulness of masked business microdata," in Proceedings on survey research methods, 1983, pp. 602–607.
16. S. Chawala, C. Dwork, F. M. Sheny, A. Smith, and H. Wee, "Towards privacy in public databases," in Proceedings on second theory of cryptography conference, 2005.
17. Science Applications International Corporation (SAIC). Role-Based Access Control (RBAC) Role Engineering Process Version 3.0. 11 May 2004.

18. A. Mohan, D. M. Blough, An Attribute-Based Authorization Policy Framework with Dynamic Conflict Resolution, Proceedings of the 9th Symposium on Identity and Trust on the Internet, 2010.
19. M. Hagner. Security infrastructure and national patent summary. In Tromso Telemedicine and eHealth Conference, 2007.
20. Federal Information Processing Standards Publication 197,"Specification for the Advanced Encryption Standards (AES)", 2001.
21. S. Fluhrer, I. Mantin, and A. Shamir, "Weakness in the Key schedualing algorithm of RC4", 8th Annual International Workshop on Selected Areas in Cryptography, Springer-Verlag London, UK, 2001.
22. J. Shafer, S. Rixner, and A. L. Cox. The Hadoop Distributed File system: Balancing Portability and Performance. Proc. of 2010 IEEE Int. Symposium on Performance Analysis of Systems & Software (ISPASS), March 2010, White Plain, NY, pp. 122-133.
23. N. Somu, A. Gangaa, and V. S. Sriram, "Authentication Service in Hadoop Using one Time Pad," Indian Journal of Science and Technology, vol. 7, pp. 56-62, 2014.
24. C. Yang, W. Lin, and M. Liu, "A Novel Triple Encryption Scheme for Hadoop-Based Cloud Data Security," in Emerging Intelligent Data andWeb Technologies (EIDWT), 2013 Fourth International Conference on, 2013, pp. 437-442.
25. H. Zhou and Q. Wen, "Data Security Accessing for HDFS Based on Attribute-Group in Cloud Computing," in International Conference on Logistics Engineering, Management and Computer Science (LEMCS 2014), 2014.
26. K. Hill, "How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did," Forbes, Inc., 2012.
27. Data Protection Laws of the World. 2017 DLA Piper. [Online]. Available: http://www.dlapiperdataprotection.com
28. Challenges of privacy protection in Big Data Analytics –Meiko Jensen- 2013 IEEE International Congress on Big Data. 2013.
29. Privacy and Big Data – Terence Craig & Mary E.Ludloff
30 Data protection overview (Morocco) – Florence Chafiol-Chaumont and Anne-Laure Falkman – 2013.
31. Big Data security and privacy issues in healthcare – Harsh Kupwade Patil, Ravi Seshadri – 2014
32. Sectorial healthcare strategy 2012-2016- Moroccan healthcare ministry.
33. Big Data in Healthcare – Pranav Patil, Rohit Raul, Radhika Shroff, Mahesh Maurya – 2014
34. Li N, et al. t-Closeness: privacy beyond k-anonymity and L-diversity. In: Data engineering (ICDE) IEEE 23rd international conference; 2007.
35. Machanavajjhala A, Gehrke J, Kifer D, Venkitasubramaniam M. L-diversity: privacy beyond k-anonymity. In: Proc. 22nd international conference data engineering (ICDE); 2006. p. 24.
36. Ton A, Saravanan M. Ericsson research. [Online]. http://www.ericsson.com/research-blog/data-knowledge/big-data-privacy-preservation/2015.
37. Samarati P. Protecting respondent's privacy in microdata release. IEEE Trans Knowl Data Eng. 2001;13(6):1010–27
38. Samarati P, Sweeney L. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical Report SRI-CSL-98-04, SRI Computer Science Laboratory; 1998.
39. V. Iyenger, "Transforming data to satisfy privacy constraints," in Pro- ceedings of the ACM SIGKDD, 2002, pp. 279–288.
40. K. LeFevre, R. Ramakrishnan, and D. J. DeWitt, "Modorian multidi- mensional k-anonymity," in Proceedings of the ICDE, 2006, p. 25.
41. Sweeney L. K-anonymity: a model for protecting privacy. Int J Uncertain Fuzz. 2002;10(5):557–70.
42. Ko SY, Jeon K, Morales R. The HybrEx model for confidentiality and privacy in cloud computing. In: 3rd USENIX workshop on hot topics in cloud computing, HotCloud'11, Portland; 2011.
43. Priyank J., Manasi G. and Nilay K. Big data privacy: a technological perspective and review. In Journal of Big Data2016.
44. Sedayao J, Bhardwaj R. Making big data, privacy, and anonymization work together in the enterprise: experiences and issues. Big Data Congress; 2014.
45. Yong Yu, et al. Cloud data integrity checking with an identity-based auditing mechanism from RSA. Future Gener Comp Syst. 2016;62:85–91.
46. Oracle Big Data for the Enterprise, 2012. [online]. http://www.oracle.com/ca-en/technoloqies/biq-doto.
47. Jung K, Park S, Park S. Hiding a needle in a haystack: privacy preserving Apriori algorithm in MapReduce framework PSBD'14, Shanghai; 2014. p. 11–17.
48. Lu R, Zhu H, Liu X, Liu JK, Shao J. Toward efficient and privacy-preserving computing in big data era. IEEE Netw. 2014;28:46–50.
49. Zhang X, Yang T, Liu C, Chen J. A scalable two-phase top-down specialization approach for data anonymization using systems, in MapReduce on cloud. IEEE Trans Parallel Distrib. 2014;25(2):363–73.
50. Mehmood A, Natgunanathan I, Xiang Y, Hua G, Guo S. Protection of big data privacy. In: IEEE translations and content mining are permitted for academic research. 2016.
51. Mohammadian E, Noferesti M, Jalili R. FAST: fast anonymization of big data streams. In: ACM proceedings of the 2014 international conference on big data science and computing, article 1. 2014.
52. Wei L, Zhu H, Cao Z, Dong X, Jia W, Chen Y, Vasilakos AV. Security and privacy for storage and computation in cloud computing. Inf Sci. 2014;258:371–86.