



# PRIVACY PRESERVING IN MULTIMEDIA BIG DATA USING SECURE USER DATA REPOSITORY SYSTEM (SUDRS)

Sibi.V<sup>1</sup>, Ramkumar.S<sup>2</sup>

1. P.G. Student, Dept. of MCA, VSB Engineering College, Karur, Tamilnadu, India
2. Asst.Professor, Dept. of MCA, VSB Engineering College, Karur, Tamilnadu, India

**Abstract**—The proliferation of multimedia big data for dissemination and sharing of massive amounts of information raises important security and privacy concerns. One such concern is the composition and enforcement of privacy policies in order to securely manage access of multimedia big data. Several researchers have pointed out that for proper enforcement of privacy policies, the privacy requirements should be captured in access control systems. In this paper, we propose a hybrid approach where privacy requirements are captured in an access control system and present a framework for composition and enforcement of privacy policies. The focus is to allow a user, not a system or security administrator to compose conflict free policies for their online multimedia data. An additional requirement is that such a policy be context-aware. We also present a methodology for verifying the privacy policy in order to ensure correctness and logical consistency. The verification process is also used to ensure that sensitive security requirements are not violated when privacy rules are enforced. A prototype, named Intelligent Privacy Manager (iPM), has been implemented for sharing of multimedia big data in a secure and private manner. **Index Terms**—Access control, context, data privacy, formal verification, multimedia databases.

## I. INTRODUCTION

Today, an increasing number of users use the Internet to manage their multimedia data regarding health-care, e-business, social networking, intelligent transportation systems, etc. [1]–[7]. This trend is further being fueled by an ever-growing number of companies and government agencies such as banks, hospitals and employers, managing users personal data in some form of online

applications and databases. The aim is to save time and money, by streamlining and facilitating access to and manipulation of online data using the Internet both in a static and mobile environment. However, theft of private information is a significant problem for online applications [8]. Hence, the overriding concern for using any internet-based service dealing with users personal data, especially multimedia data due to its sheer volume and rich semantics, is ensuring security and privacy of their personal information. An important security and privacy concern of online multimedia systems is the composition and enforcement of privacy policies in order to securely manage access of multimedia big data. The relation between multimedia big data and composition of user-centered privacy policies is two fold. First, the large volume of multimedia data and large numbers of users desiring enforce/control distributed access to multimedia data assets (i.e. images, videos etc.) in applications such as healthcare and social networking introduces the problem of who should compose the policy. That is, should there be a unified/centralized policy for all users OR multimedia data owners (a user) should compose their own policy in order to grant or deny access. Second, it is more conducive and intuitive for a normal user with very little or no domain knowledge or expertise to compose a policy to manage privileges of multimedia data by simple



drag and drop mechanism in comparison to structured data based on formal schema such as relational data or XML. Hence, this relation introduces the need for a comprehensive framework to allow for composition and enforcement of policies in order to have seamless and context-driven secure access to multimedia big data that traditional multimedia systems cannot handle effectively.

## II. LITERATURE SURVEY

### A. Cloud-Based Multimedia Content Protection System (2015)

The system is used to protect different multimedia content types under the cloud environment 3-D Video Signatures Scheme and Distributed Matching Engine are used to provide multimedia data access with security. It supports creating amalgamated signatures that consist of one or more of the following elements: 1. Visual mark: Created based on the optical parts in multimedia objects and how they change with time; 2. Audio mark: Created based on the audio signals in multimedia object s; 3. Depth mark: If multimedia objects are 3-D videos, signatures from their depth signals are created; 4. Meta data: Created from information associated with multimedia items such as their names, tags, descriptions, layout types, and IP addresses of their uploaders. The disadvantages of this paper are Computational complexity is high in online redistribution verification process.

### B. Privacy and Quality Preserving Multimedia Data Aggregation for Participatory Sensing Systems (2015)

The system supports privacy and quality preserving participatory sensing with multimedia data SLICER scheme integrates a data coding technique and message transfer strategies to achieve strong

protection of participants" privacy. The domestic attack may come from both the participants and the service provider. We differentiate two cases: Protection against participants" attack. Each contributor may accept some slices, when it is preferred as a slice deliver for participants met. Similar with the peripheral attacker, the participant cannot decrypt the slice for delivering shield against service provider"s attack. Given that the service provider has full access to the sensing records contributed by the participants, it can easily infer secret information about the participants, if proper privacy - preserving design is not provided. Still, SLICER can achieve the k-anonymity and protect participants" privacy information against the service provider. Therefore, we can draw the following theorem. The disadvantages of this paper are Privacy on query processing is not supported.

### C. Innovative Schemes for Resource Allocation in the Cloud for Media Streaming Applications (2015)

The system provides streaming resources from the cloud to the media content providers Prediction Based Resource Allocation (PBRA) algorithm is used to manage streaming resource allocation for content distribution. The reservation plan, the media substance provider reserves possessions in advance and pricing is exciting before the possessions are utilized (upon getting the appeal at the cloud provider, i.e., prepaid possessions). The ondemand plan, the media content supplier allocates streaming possessions ahead required. Pricing in the on-demand plan is stimulating by pay-per-use starting point. In general, the prices (tariffs) of the reservation research are cheaper than those of the ondemand plan (i.e., time reduce rates are only offered to the reserved (prepaid) resources). The Disadvantage of this paper are



Media content security and privacy are not considered.

#### **D. ENF-Based Region-of-Recording Identification for Media Signals (2015)**

Electrical Network Frequency (ENF) signals are used to verify the location of multimedia content recording process. Multiclass Region-of-recording Classification scheme is adapted to verify the recording location of multimedia contents. Dimensionality reduction schemes are identified to be helpful to assist efficient implementations of machine wisdom in many applications concerning a high dimension of features. To examine their effects on our difficulty, we have experimented with dissimilar dimensionality reduction schemes, purposely, the Fisher's Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) [20]. Using LDA, we trim down the dimensionality of our data from the original dimensionality of 16 to  $M - 1$ , where  $M$  is the amount of classes. Grids With Varying Profiles A probable factor impacting the efficiency of ENF-based spot classification is the unpredictability of a grid's ENF attributes with time. The power profile of a positive grid may exhibit different types of behaviours, depending on the time of a day or the term of a year. For our dataset, we have made sure to collect data from both daytime and nighttime, and when probable, we have try to collect data from different times in a year. The disadvantages of this paper are Security and privacy are not provided for multimedia data.

#### **E. Joint Physical-Application Layer Security for Wireless Multimedia Delivery (2014)**

Wireless multimedia data delivery scheme is secured with stream based data security models

Joint physical application layer security system provides the security for multimedia data delivery process. Watermarking technology extend the guard of multimedia content behind decryption by embedding a mark that cannot be apparent in the multimedia content. In other words, one embeds an subjective watermark into multimedia contents by applying unnoticeable changes to the original multimedia contents. Fundamentally, these alterations depend on a private key at the detector, and are tested and examine at the decoder. In preparation, the decoding idea uses the received watermarked multimedia content and a private key to approximate and test the watermark this is called a blind detection architecture. The disadvantages of this paper are Device level authentication and energy factors are not considered.

#### **F. A Hybrid Scheme for Authenticating Scalable VideoCode streams (2014)**

Scalable video coding (SVC) scheme is adapted to secure video transmission process hybrid authentication (HAU) scheme employs both cryptographic authentication and content-based authentication techniques for video authentication. The authentication tag generation procedure includes MAC production and feature extraction, where MACs are constructed by taking the encoded image of the bottom layer and a secret key as inputs, while the features of each frame are extracted from the maximum quality and resolution images of SVC bitstreams. Note that if an SVC codestream has spatial enhancement layers, HAU should originally downsample the main resolution to the same resolution as the base layer, then extract features from the downsampled one. Hence, extent of the feature value is allied to the base layer's resolution for each AU (Access Unit). Tag Conveyance is required to carry the authentication



tag (i.e., feature hash  $V$  and MAC  $\phi$ ) to the recipient together with the SVC codestream for confirmation. In HAU, we summarize the tag into SVC user data as a new SEI (Supplement Enhancement Information) NALU as was done in. The payload type of the new SEI is Unregistered User Data communication so as to preserve SVC format. The disadvantages of this paper are User level privacy is not supported.

### III. SYSTEM MODEL

#### A. Design Requirements

Multimedia big data systems face multifaceted design challenges such as scalability, mobility and overall verifiability of underlying policy. Scalability deals with provisioning of security for large number of users and devices interacting across the whole system. Given the sheer volume of data, the question from scalability perspective is whether it is a feasible approach for multimedia big data to compose one global policy for all users or compose individual policies for users. Since composition and verification of a policy can be non-linear in terms of complexity with respect to number of users, the computational complexity of composing one global policy is much higher than composing individual policies. Therefore, composing individual policies for multimedia big data is not a force fitting but rather a natural and innovative approach. With respect to mobility, the unique benefits include ubiquity, relevance and immediacy which are magnified by big data. To fully leverage these attributes, multimedia big data solutions need to be location-aware (ubiquitous), real-time (immediate), and context-aware (relevant). Moreover, since user can be naive, the composed policy should be verified with low complexity in order to avoid conflicts. The verification of fully composed, one global policy

can produce unpleasant consequences (i.e. may not be able to verify the policy at all). Hence the verification process should be local to individual policies. This further highlights the fact that composing individual policies is a natural approach for multimedia big data as mentioned above. The only limitation in verifying individual policies is that a local decision in verification may not present all the optimal solutions available and can overwhelm a user. In addition, such systems should provide support for privacy requirements such as Control, Collection, Intention Specification and Recording. *Control* implies an organization which collects personal data should control access and use of data while *Collection* refers to collecting data in a disciplined fashion in conformance with the privacy policies. *Intention Specification* requires that data access requests must specify the intention or intentions for how and by whom the data is going to be used. Collected data must therefore be used only for specified intentions under given context. *Recording* requires those who collect and release data should record each release or use to facilitate audits, inquiries and requests for access and revision. Finally, an orthogonal design issue for multimedia big data systems is search functionality that would allow a user to quickly locate their data. Although we don't implement the search feature, but the search feature can be supported by implementing the underlying GST-RBAC model used to capture the security and privacy requirements on SQL as done by Microsoft Azure.

#### B. Context-Aware Access Control

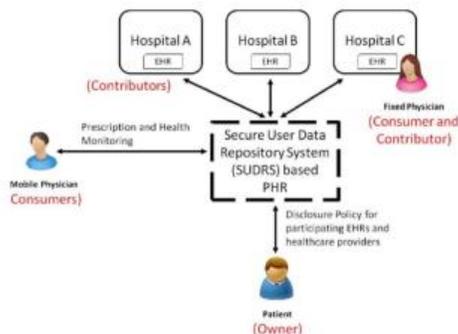
To deal with the context requirements related to privacy and security of multimedia big data systems, we use a Generalized Spatio-temporal Role Based Access Control (GST-RBAC) model.

GST-RBAC uses the basic RBAC model by taking into account the environmental contexts, such as location and time to provide a comprehensive and generalized approach to security and privacy management. In other words, GST-RBAC is a spatio-temporal extension of RBAC, a de-facto model for specifying security requirements of large organizations. RBAC model consists of four basic components including a set of users/devices, a set of roles, a set of permissions, and a set of sessions. A user can be a human being, an autonomous agent, a task, a physical device or a subsystem. RBAC is distinguished by its inherent support for principle of least privilege which requires a user to be given no more privileges than necessary to perform a task. Role enabling constraint in GST-RBAC defines the relationship between locations and time for which the role can be enabled. A role is enabled at a certain location and time, while it is not enabled at other locations or other times. Spatial constraints in GST-RBAC take location and temporal constraints take time as a context parameter validating an access request. The GST-RBAC model allows specification of spatial and temporal constraints on role enabling, user-role assignment, temporal constraints on role enabling, user-role assignment, and role-permission assignments, activation, runtime events, constraint enabling expressions and triggers .

Fig 1. Secure User Data Repository System

### C. Proposed System Design

The aforementioned security challenges are addressed by presenting a scalable generic architecture using a hybrid approach for the proposed SUDRS depicted in Fig. 1. The architecture consists of four components and several databases. We now describe the high-level architecture of Fig. 1 and discuss the functionality of the components. Note that different users can view a given multimedia data object differently depending on the users authorizations. Fig. 1 illustrates the generation of multiple views for the same multimedia data based on different authorizations. In this scenario, Robert Shaw's primary physician is authorized to see all of his medical records. The primary physician could consult another physician for an expert opinion. However, the patient's privacy policy might not allow releasing any information that can identify the patient to anyone other than the primary physician. Therefore, the system creates a filtered version of the original view where the patient's name changes from Robert Shaw to Bob, and the exact date of birth changes to an age attribute. The architecture in Fig. 1 shows that that the raw data and filtered data are stored separately in different databases. The proposed SUDRS allows the composition and enforcement of disclosure rules for personal data maintained by it. These rules are specified by the contributors, consumers and owner of data. The contributors of data are the originators of data generated as a consequence of the user's interaction with contributors. An example in this case is a health-care provider (doctors, nurses etc.) who generate health related information concerning a patient (owner) and contribute this information to be stored in the SUDRS. In this case, healthcare



provider is a contributor and patient is the owner of data. In addition to contributing data, the contributor also generates Originator Disclosure Rules (ODRs) for each element of data. The consumers of data in Fig. 1 are entities interested in accessing data. Consumers can also be contributors of data. For example, physician who is a consumer is adding records to patient profile thus acting also as a contributor. Consumers provide Access Rules (ARs) to the SUDRS defining the access times, locations, and other context parameters that govern a particular access for generic data types. An example in this regard is a physician defining the times of day as well as the location of access (e.g. from his clinic) for viewing pathology reports of a particular patient. Alternatively, this access by the physician can be from a remote location using a tablet or a mobile device. The consumers in Fig. 1 also provide their profile information [Consumer Profile (CP)] such as name, credentials and affiliation.

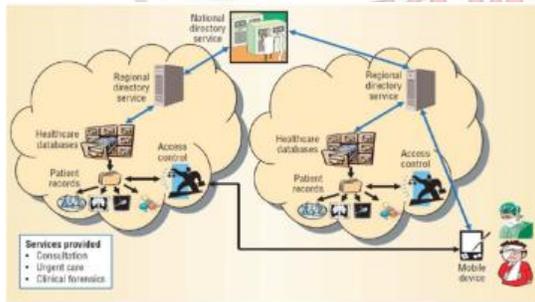


Fig 2. Typical Health Care System

The owner of data in Fig. 1 composes disclosure rules based on his or her Disclosure Intentions (DI), ODRs, ARs and CPs. For example, a patient can specify that the physician can access patient's pathology report only from hospital and only at certain times. These spatio-temporal access rules are then stored in the Disclosure Rule Base (DRB)

thus constituting a privacy driven GST-RBAC policy. Any access to the user's personal data by the consumers is thus evaluated by the GST-RBAC model against the stored rules and access is granted accordingly. The required data is extracted from the data storage system and sent to the consumer. Each request, either satisfied or denied, is also logged in the auditing subsystem. This logging allows the owners to track accesses to their personal records as well as evaluate accesses not satisfied by the disclosure rules in the system. Additionally, the audit subsystem keeps track of all disclosure rules that allowed access to a certain part of data with a view of facilitating the owner to adapt disclosure rules.

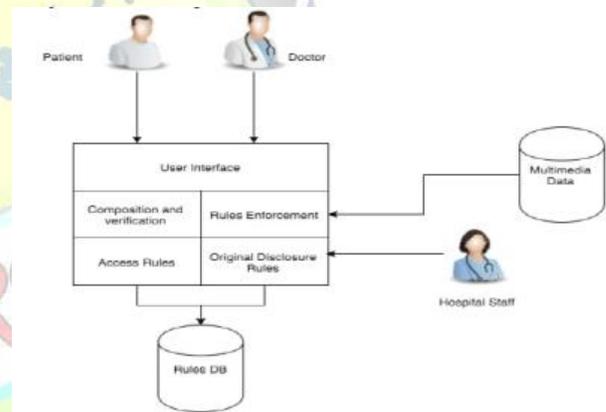


Fig 3. System Architecture

The following are the four critical components depicted in Fig. 3(a) that enable SUDRS to maintain privacy and security of data.

Composition and verification of Access Rules (ARs)

*Component 1:* This component allows consumers of multimedia data to compose and verify access rules held in the SUDRS.

*Composition and verification of Originator Disclosure Rules (ODRs) - Component 2:* This component facilitates the contributor to compose



and verify the ODRs related to multimedia data contributed to the SUDRS.

*Composition and verification of disclosure rules – Component 3:* This component enables the composition and verification of disclosure rules by owners based on ODRs, ARs, and Disclosure Intentions (DIs).

*Enforcement of disclosure rules - Component 4:* This component evaluates request of data stored in SUDRS based on context of request.

#### **D. Modules Description:**

PHR Owner Module

Cloud Server Module

Attribute based Access Policy Module

Data confidentiality Module

#### **PHR Owner Module**

The main goal of our framework is to provide secure patient-centric PHR access and efficient key management at the same time. The key idea is to divide the system into multiple security domains (namely, *public domains* (PUDs) and *personal domains* (PSDs)) according to the different users' data access requirements. The PUDs consist of users who make access based on their professional roles, such as doctors, nurses and medical researchers. In practice, a PUD can be mapped to an independent sector in the society, such as the health care, government or insurance sector. For each PSD, its users are personally associated with a data owner (such as family members or close friends), and they make accesses to PHRs based on access rights assigned by the owner.

Each data owner (e.g., patient) is a trusted authority of her own PSD, who uses a KP-ABE system to manage the secret keys and access rights of users in her PSD. Since the users are personally known by the PHR owner, to realize patient-centric access, the owner is at the best position to grant user access privileges on a case-by-case basis. For PSD, *data attributes* are defined which refer to the intrinsic properties of the PHR data, such as the category of a PHR file. For the purpose of PSD access, each PHR file is labeled with its data attributes, while the key size is only linear with the number of file categories a user can access. Since the number of users in a PSD is often small, it reduces the burden for the owner. When encrypting the data for PSD, all that the owner needs to know is the intrinsic data properties.

#### **Cloud Server Module**

In this paper, we consider the server to be semi-trusted, i.e., honest but curious as those in and . That means the server will try to find out as much secret information in the stored PHR files as possible, but they will honestly follow the protocol in general. On the other hand, some users will also try to access the files beyond their privileges. For example, a pharmacy may want to obtain the prescriptions of patients for marketing and boosting its profits. To do so, they may collude with other users, or even with the server. In addition, we assume each party in our system is preloaded with a public/private key pair, and entity authentication can be done by traditional challenge-response protocols.

#### **Attribute based Access Policy Module**

In our framework, there are multiple SDs, multiple owners, multiple AAs, and multiple users. In addition, two ABE systems are involved. We term



the users having read and write access as data readers and contributors, respectively.

#### Data confidentiality Module

The owners upload ABE-encrypted PHR files to the server. Each owner's PHR file is encrypted both under a certain fine grained and role-based access policy for users from the PUD to access, and under a selected set of data attributes that allows access from users in the PSD. Only authorized users can decrypt the PHR files, excluding the server.

#### IV. CONCLUSION

The sheer volume and rich semantics of multimedia big data as well as online management of user personal information introduce important security and privacy concerns. In this paper, we present the design of a system for composing and enforcing context-aware disclosure rules for preserving privacy and security of multimedia big data systems. The proposed system allows an online user to compose disclosure rules which are consistent and are verified for a set of verification properties. We also present the design of iPM prototype developed to implement the above mentioned design. Its various components and resulting policy is also presented. As future work, the proposed system can be extended in various directions. The prototype can be adapted for various applications such as secure Facebook, intelligent transportation systems etc. Furthermore, the prototype can be redesigned based on existing mature technologies such as Services-oriented Architecture (SOA). Future work will improve the security solution (implement HIPAA requirements, using HTTPS) and will evaluate the results through measuring the interoperability degree achieved by the presented solution.

#### REFERENCES

- [1] W. Zhu, C. Luo, J. Wang, and S. Li, "Multimedia cloud computing," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 59–69, May 2011.
- [2] S.-C. Chen, M.-L. Shyu, S. Peeta, and C. Zhang, "Learning-based spatio-temporal vehicle tracking and indexing for transportation multimedia database systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 4, no. 3, pp. 154–167, Sep. 2003.
- [3] S.-C. Chen, M.-L. Shyu, S. Peeta, and C. Zhang, "Spatiotemporal vehicle tracking: The use of unsupervised learning-based segmentation and object tracking," *IEEE Robot. Autom. Mag.*, vol. 12, no. 1, pp. 50–58, Mar. 2005.
- [4] R. Bhatti, B. Shafiq, M. Shehab, and A. Ghafoor, "Distributed access management in multimedia IDCs," *Computer*, no. 9, pp. 60–69, 2005.
- [5] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: Promise and potential," *Health Inf. Sci. Syst.*, vol. 2, no. 1, p. 3, 2014.
- [6] P. Weill and M. Vitale, *Place to Space: Migrating to eBusiness Models*. Cambridge, MA, USA: Harvard Univ. Press, 2013.
- [7] K. Zhang, X. Liang, X. Shen, and R. Lu, "Exploiting multimedia services in mobile social networks from security and privacy perspectives," *IEEE Commun. Mag.*, vol. 52, no. 3, pp. 58–65, Mar. 2014.
- [8] Privacy Rights Clearinghouse, San Diego, CA, USA, "A chronology of data breaches," , 2005 [Online]. Available: <http://www.privacyrights.org/data-breach>



- [9] I. Carrion, J. Aleman, and A. Toval, "Personal health records: New means to safely handle health data?," *Computer*, vol. 45, no. 11, pp. 27–33, Nov. 2012.
- [10] Christo Ananth, S. Esakki Rajavel, I. Anna Durai, A. Mydeen@SyedAli, C. Sudalai @ UtchiMahali, M. Ruban Kingston, "FAQ-MAST TCP for Secure Download", *International Journal of Communication and Computer Technologies (IJCCCTS)*, Volume 02 – No.13 Issue: 01, Mar 2014, pp 78-85
- [11] J. B. Joshi, Z. K. Li, H. Fahmi, B. Shafiq, and A. Ghafoor, "A model for secure multimedia document database system in a distributed environment," *IEEE Trans. Multimedia*, vol. 4, no. 2, pp. 215–234, Jun. 2002.
- [12] S. A. Chun and V. Atluri, "Protecting privacy from continuous highresolution satellite surveillance," in *Data and Application Security*. New York, NY, USA: Springer, 2001, pp. 233–244.
- [13] Christo Ananth, A. Ramalakshmi, S. Velammal, B. Rajalakshmi Chmizh, M. Esakki Deepana, "FASTR –SAFE AND SECURE", *International Journal For Technological Research In Engineering (IJTRE)*, Volume 1, Issue 12, August-2014, pp: 1433-1438
- [14] E. Bertino, B. Catania, M. L. Damiani, and P. Perlasca, "Geo-RBAC: A spatially aware RBAC," in *Proc. 10th ACM Symp. Access Control models Technol.*, 2005, pp. 29–37.
- [15] R. Bhatti, E. Bertino, and A. Ghafoor, "X-FEDERATE: A policy engineering framework for federated access management," *IEEE Trans. Softw. Eng.*, vol. 32, no. 5, pp. 330–346, May 2006.
- [16] I. Jacobson, G. Booch, and J. E. Rumbaugh, "Excerpt from the unified software development process: The unified process," *IEEE Softw.*, vol. 16, no. 3, pp. 82–90, Jan.–Feb. 1999.
- [17] J. Jürjens and G. Wimmel, "Security modelling for electronic commerce: The common electronic purse specifications," in *Proc. Towards E-Society*, 2001, pp. 489–505.
- [18] J. Jürjens, "Towards development of secure systems using UMLsec," in *Proc. Fundam. Approaches Softw. Eng.*, 2001, pp. 187–200.
- [19] J. P. Bowen, *Z: A Formal Specification Notation*. New York, NY, USA: Springer, 2001.
- [20] J. Zao, H. Wee, J. Chu, and D. Jackson, "RBAC schema verification using lightweight formal model and constraint analysis," in *Proc. SACMAT*, 2003 [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.93.1344&rep=rep1&type=pdf> [21] P. Ashley, S. Hada, G. Karjoth, and M. Schunter, "E-p3p privacy policies and privacy authorization," in *Proc. Workshop Privacy Electron.Soc.*, 2002, pp. 103–109.
- [22] G. Karjoth and M. Schunter, "A privacy policy model for enterprises," in *Proc. Comput. Security Found. Workshop*, 2002, pp. 271–281.
- [23] S. Jajodia, P. Samarati, M. L. Sapino, and V. Subrahmanian, "Flexible support for multiple access control policies," *ACM Trans. Database Syst.*, vol. 26, no. 2, pp. 214–260, 2001.
- [24] G. Karjoth, M. Schunter, and M. Waidner, "Platform for enterprise privacy practices: Privacy-enabled management of customer data," in *Proc. Privacy Enhancing Technol.*, 2003, pp. 69–84.
- [25] M. Schunter and C. Powers, "The enterprise privacy authorization language (EPAL 1.1)," IBM



ISSN 2394-3777 (Print)  
ISSN 2394-3785 (Online)  
Available online at [www.ijartet.com](http://www.ijartet.com)

*International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)*  
**Vol. 3, Special Issue 17, March 2016**

Res. Rep. RZ 3485 (#93951), 2003

[Online]. Available:

<http://www.w3.org/Submission/2003/SUBM-EPAL-20031110>

[26] S. Pearson and M. C. Mont, "Sticky policies: An approach for managing privacy across multiple parties," *Computer*, vol. 44, no. 9, pp. 60–68, 2011

