Applying Data Privacy Techniques on Published Data in Uganda

Kato Mivule[1], Claude Turner[2]
Computer Science Department
Bowie State University
14000 Jericho Park Road Bowie, MD 20715, USA
[1]mivulek0220@students.bowiestate.edu, [2]cturner@bowiestate.edu

Abstract - *The growth of information technology (IT) in Africa has led to an increase in the utilization of communication networks for data transaction across that continent. Thus, many in Africa have become increasingly dependent on the Internet for data transactions. In the country of Uganda, for example, exponential growth in data transaction has presented a new challenge. Namely, what is the most efficient way to implement data privacy? While studies on data privacy have been done for developed nations such as in the European Union, studies for data privacy implementation in emerging markets have been minimal. It is with such background that we discuss data privacy challenges in Uganda. We also present an implementation of data privacy techniques for a published Ugandan dataset and suggest how this approach may be generalized to provide data privacy in the country.*

Keywords: Data Privacy; Database Security; Statistical Disclosure Control; k-anonymity; Tabular data.

1. Introduction

The exponential growth of Information Technology (IT) in Africa has led to an increase in data transaction across Africa's communication networks, with 110 million Internet users and 500 million mobile phone subscriptions as of 2010[1]. In Uganda's case, higher education institutions routinely post student admission and graduation data online and grant access to student records online [2]. The Ugandan Electoral Commission posted the national voter's register online [3][4]. While the Uganda Bureau of Statistics publishes statistical data routinely, and takes great care to remove personal identifiable information (PII), a review of the published datasets from other Ugandan entities such as educational institutions and the Electoral Commission of Uganda show PII was included in published datasets. At the same time a growing number of young Ugandans are fans of large Online Social Networks (OSN) like Facebook, resulting in large amounts of PII leaked from online auxiliary data sources.

While case studies on data privacy have been done for developed nations such as in the European Union, studies for data privacy and security implementation in emerging markets such as Uganda have been minimal [48]. Yet with the growth of the globalized economy and multinational entities, demands for data privacy and security while transacting in business in the emerging markets is critical. Therefore in this paper, we take a look at current data privacy and security laws and present an implementation of data privacy techniques for a published Ugandan dataset and suggest how this approach may be generalized to provide data privacy in the country.

The rest of this paper is organized as follows. Section 2 looks at current data privacy and security policies in Uganda. Section 3 describes related work on data privacy and security in Uganda. Section 4 looks at the essential data privacy terms used in this paper. Section 5 gives an overview on data privacy techniques discussed in this paper. Section 6 discusses the implementation while Section 7 presents the results; and finally, Section 8 provides the conclusion.

2. Data Privacy and Security Policies

In developed countries like the USA, data gathering institutions are bounded by state and federal privacy laws that require that privacy of individuals be protected. One example in the USA is the Privacy Act of 1974, Health Insurance Portability and Accountability Act (HIPAA) of 1996, and the Personal Data Privacy and Security Act of 2009, requiring entities to protect and secure PII in data [5][6][7]. The Ugandan constitution defines the rights of an individual to privacy in terms of interference, stating that no person shall be subjected to interference with the privacy of that person's home, correspondence, communication or other property, however, no precise definition is given in the context of PII, data privacy, and computer security [8]. Ugandan Bureau of Statistics Act of 1998 describes Ugandan government policy on data collected by the Ugandan Bureau of Statistics (UBS). Absent from that description is how non-governmental entities collect and disseminate data. The Ugandan Bureau of Statistics Act of 1998 does not discuss what PII is in the Ugandan context. The only close reference is the "removal of identifiers" before data is granted to researchers [9]. In this case "identifiers" is ambiguous and could perhaps reference 'names' but not 'geographical location'. However, UBS with expert care does publish de-identified micro datasets online but at the same time, many entities in Uganda publish non de-identified tabular datasets.

A look at documents from authorities that govern communication technology in Uganda, the Uganda Communications Commission (UCC) and the Ministry of Information and Communications Technology (ICT) show that policies on data privacy and security have not been clearly formulated [9][10][11][12][13][14]. In the USA for instance, PII could include an individual's social security number yet in Uganda, social security numbers are non-existent; thus, the set of PII in the USA differs from that in Uganda. Therefore, there is a need to expand Uganda's policy on how government and non-government entities collect and disseminate data. To date, no clear legal and technological data privacy framework exists in Uganda. Despite the absence of any clearly formulated policy on data privacy in Uganda, this work suggest the application of data privacy techniques that could be utilized to provide basic data privacy in this context.

3. Related work on data privacy in Uganda

Our study of the literature reveals that work on data privacy in Uganda and much of sub-Saharan Africa is sparse. To date and to the best our knowledge, this work’s focus on the application of data privacy techniques to the Ugandan context might be novel. While research on computer security in Uganda exists, most of the work centers on network accessibility control methodologies [15][16][17][18][19]. For example, Mutyaba [20] and Makori [21] offer an excellent presentation on cryptographic methodologies for computer security, and Okwangale and Ogao [22] discuss data mining techniques; however, privacy preserving data mining (PPDM) methodologies are not discussed. Bakibinga [23] has articulated the need for electronic privacy in Uganda from a policy view point. Frameworks for secure management of electronic records have been proposed by Luyombya [24], Ssekibule and Mirembe [25], and Kayondo [26]; however, these works focus on data security and access control. But data privacy differs from data security in that data privacy has to do with the confidentiality of data, while data security focuses on its accessibility. Even when a database system is physically secured, an inference attack could occur on published datasets [27]. It should be noted that the Ugandan Bureau of Statistics Act of 1998 does provide a legal framework for data privacy that focuses on data gathered by the UBS. What is absent from the Ugandan computational literature is the data privacy technological framework that entities other than the Ugandan Bureau of Statistics, such as health, academia, and private business could employ [28]. To date, no work has come to our attention on if data privacy methodologies employed by UBS have been applied to private sector. Therefore, it is in this light that we make the case for data privacy in Uganda and the need for more research on data privacy and PPDM methodologies tailored to the Ugandan and African context.

4. Essential data privacy terms

The following definitions will be important in the sequel: *Data privacy* is the protection of an individual’s data against unauthorized disclosure while *Data security* is the safety of data from unauthorized access [29] [30]. *Personally identifiable information* (PII) is any data about an individual that could be used to construct the full identity of that individual [31][32]. *Data De-identification* is a process in which PII attributes are removed such that when the data is published, an individual's identity cannot be reconstructed [33] [34]. *Data utility verses privacy* has to do with how useful a published dataset is to a consumer of that published dataset [35] [36]. Often the usefulness of data is lost when PII and quasi-attributes, are removed or transformed; a balance between privacy and data utility is always sought [37]. It has been determined that achieving optimal data privacy while not distorting data utility is a continual NP-hard challenge [38]. *Statistical databases* are published data sets that do not change, in many cases released in aggregated format [39]. *Attributes* in statistical databases, are field names or columns [29]. *PII attributes* are properties that uniquely identify an individual; an example includes social security number. *Quasi-attributes* are attributes not in the PII category but can be used to reconstruct an individual's identity in conjunction with external data. *Confidential attributes* are attributes not in the PII and quasi-attributes category but contain sensitive information, such as salary, HIV status, etc. *Non confidential attributes* are attributes that individuals do not consider sensitive as causing disclosure. However, non-confidential attributes can still be used to re-identify an individual given auxiliary data, thus making the explicit description of what PII is and is not even more challenging [40]. *Inference and reconstruction attacks* are methods of attack in which separate pieces of data are used to derive a conclusion about a subject, in this case, reconstruct their identity [41].

5. Data privacy techniques

Data privacy methods are categorized as *non-perturbative* techniques in which original data is not modified, some data is suppressed or some sensitive details removed while with *perturbative techniques*, original data is altered or disguised so as to protect PII and sensitive data [29]. While a number of data privacy techniques exist, we focus on application of *k-anonymity*, *suppression*, and *generalization*. *Suppression is* a popular data privacy method in which data values that are unique and can be used to establish an individual's identity are omitted from the published dataset [42][43]. *Generalization* is a data privacy method in which attributes that could cause identity disclosure are made less informative. An example includes replacing the gender attribute value with “person” instead of “Male” or “Female” [44]. *K-anonymity* is a data privacy enhancing mechanism that utilizes *generalization*, and *suppression* as outlined extensively by Samarati [45] and Sweeney [27]. *k-anonymity* requires that for a dataset with quasi-identifier attributes in database to be published, values in the quasi-identifier attributes be repeated at least ***k*** times to ensure privacy; that is, $k > 1$ [27]. However, achieving the optimal *k-anonymized* dataset has been shown to be an NP-Hard problem [46].

6. Data privacy implementation

In this section, we describe our implementation of basic data privacy algorithms on a Ugandan dataset, utilizing open source technologies that are freely available for all to download. In this way, nations from emerging markets such as Uganda could incur minimal costs when it comes to data privacy implementation. We express our implementation using the set theory notation, relational database notation, and lastly MySQL implementation. The initial step was to de-identify a Ugandan dataset of 1200 records from a Makerere University student admission list that is published publicly online by the University, by removing PII as defined by the US data privacy laws [3]. While no explicit data privacy laws exist in Uganda, we utilized the definitions of what constitutes PII as defined by the US data privacy laws (HIPAA), considering that they could be universally applicable. We employed SQL, utilizing MySQL Sever, an open source tool freely available for download.

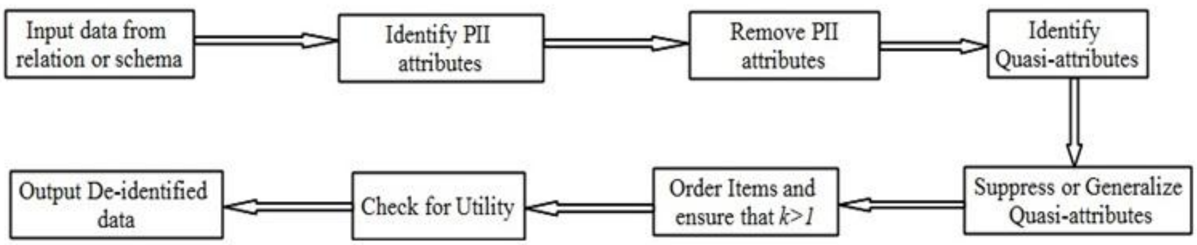

Figure 1*: A Data De-identification procedure utilizing k-anonymity*

RegNo	StudentNo	Lname	Fname	Mname	Sex	BirthDate	Nationality	Hall	Program	IndexNo	Year
09/U//EVE	20900	Annet	Anna		F	01/01/67	UGANDAN	AFRICA	LIS	U0166	2008
09/U//EVE	20901	Green	RICE		F	01/01/80	UGANDAN	MARY STUART	ARM	U0763	2008
09/U//EVE	20902	Timothy	NICE		F	01/01/81	KENYAN	MARY STUART	BLE	U0063	2007
09/U//EVE	20903	Jones	Jane	GRACE	F	01/01/73	TANZANIA	MARY STUART	LIS	U0198	2007
09/U//EVE	20904	Carter	James		M	01/01/74	UGANDAN		RAM	U0160	2007
09/U//EVE	20905	Brown	Britain	N	F	01/01/83	KENYAN	AFRICA	ARM	U0715	2008
09/U//EVE	20906	Sams	Sam		F	01/01/84	TANZANIA	MARY STUART	RAM	U0725	2007
09/U//EVE	20907	Faster	Master		M	01/01/85	UGANDAN		BLE	U1148	2008
09/U//EVE	20908	Uhuru	Kenya		F	01/01/90	UGANDAN	COMPLEX	ARM	U0062	2007
09/U//EVE	20909	Vineyard	Martha		M	01/01/88	KENYAN	AFRICA	ARM	U1017	2008

Table 1*: Admission List with PII – BirthDate, IndexNo, and RegNo are generalized*

Steps in the Data Privacy Procedure shown in Figure 1:
INPUT: Data from relation or schema
OUTPUT: Data privacy preserving published tabular dataset

1. Identify PII Attributes
2. Remove PII Attributes
3. Identify quasi-identifier attributes
4. Generalize or Suppress quasi-identifier attributes
5. Check that $k>1$ in tuples
6. Check for single values that cannot be grouped together to achieve $k>1$
7. If single values and outliers exist, Generalize or Suppress until k-anonymity at $k>1$
8. Check for utility
9. Publish tabular dataset

We borrowed from set theory notation to describe how we implemented the data privacy procedure on the Ugandan data set as follows:

- The original Ugandan published dataset included the following attributes, in which we let the following:
  - *A* = { *RegNo, StudentNo, Lname, Fname, Mname, Sex, BirthDate, Nationality, Hall, Program, IndexNo, Year* }, the relation *admission list* that included all attributes in the published dataset.
  - We let *B* = { *Lname, Fname, Mname, StudentNo, IndexNo, RegNo*}, the set of all PII attributes that we identified in the published dataset.
  - We let *C* = { *Nationality, Sex, BirthDate,*}, the set of all quasi-identifier attributes identified in the dataset.
  - We let *D={Hall, Program, Year}*, the set of all non-sensitive attributes.
  - *Lastly, we let E={ }*, the set of all sensitive attributes.
- Thus, we have $B \subset A$, $C \subset A$, $D \subset A$ and $E \subset A$;
  - *Therefore* $A = B \cup C \cup D \cup E$, *and* $A = \{ B, C,$

D, E}.

- ○ *By r*emoving PII, we get *A ={ C, D, E}.*
- ○ The de-identification of the *Admission List* set involves a complement of the *PII* set: $(B)^c = U - B = A - B = C + D + E$. Therefore, we remained with the *quasi attributes, non-sensitive attributes, and sensitive attributes; w*here *U* is the universal set, which in this case is all the *Admission List attributes.*

- We suppressed or generalized the *quasi attributes:* suppress or generalize (*C*).
  - ○ We then applied *k-anonymity*: *k-anonymity(* $(B)^c$ *).*
  - ○ *Finally,* we ordered values of $(B)^c$.
  - ○ If $k = 1$, we suppressed or generalized *C* until $k > 1$.

Relational model view: For a formal relational model view implementation, we applied the following notation:

- ○ we let *π <attribute list>*$^{(R)}$ *,*
  - ○ *w*here ***π*** is the projection or selecting of attributes from a relation (Table),
  - ○ *<attribute list>* is the list of attributes from *Admission List*
  - ○ , $^{(R)}$ is the relation from which we select attributes.

The original projection with all attributes is:

- ○ *π<RegNo, StudentNo, Lname, Fname, Mname, Sex, BirthDate, Nationality, Hall, Program, IndexNo, Year >*$^{(\,Admission\ List\,)}$.
- ○ The projection void of PII attributes is:
  - ○ *To_Be_Published_List ← π< Sex, BirthDate, Nationality, Hall, Program, Year >*$^{(\,Admission\ List)}$.
- ○ We apply *k-anonymity* to the list that is to be published:
  - ○ *k-anonymity(To_Be_Published_List).*

7. Results

We generalized the *BirthDate* attribute to further prevent any reconstruction attacks by first developing a domain generalization hierarchy (DGH). We chose the DGH based on the oldest person in the dataset, and built our DGH to B_4 = {196*}, giving protection for the individuals born in 1967 [43], as shown in *Figure 2.*

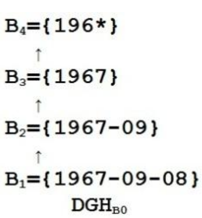

Figure 2*: Domain generalization hierarchy structure*

The SQL Implementation: We implemented data de-identification in SQL by creating a SQL View and doing SELECT on the view by choosing only attributes that remain in the *Admission List* after removing PII. We created SQL Views that are void of PII attributes:

```
CREATE VIEW V2 AS SELECT  Sex,
BirthDate, Nationality, Hall,
Program, Year FROM
Admission_List;
```

Generalization: Utilizing the SQL functions, CREATE, SELECT, and UPDATE, we further generalized the *Program* attribute so as not to grant such information to a researcher. We generalized the *BirthDate* attribute to additionally prevent any reconstruction attacks.

Sex	BirthDate	Nationality	Hall	Program	Year
F	196*	UGANDAN	AFRICA	LIS	2008
F	196*	UGANDAN	MARY STUART	ARM	2008
F	196*	KENYAN	MARY STUART	BLE	2007
F	196*	UGANDAN	MARY STUART	LIS	2008
M	196*	UGANDAN		RAM	2007
F	196*	KENYAN	AFRICA	ARM	2008
F	196*	TANZANIA	MARY STUART	RAM	2007
M	196*	UGANDAN		BLE	2008
F	196*	UGANDAN	COMPLEX	ARM	2007
M	196*	TANZANIA	AFRICA	ARM	2008

Table 2: *Results after generalization and suppression*

MySQL implementation:

```
CREATE table V2_Generalize1
SELECT Sex, BirthDate,
Nationality, Hall, Program, Year
FROM V2;

UPDATE V2_Generalize1 set
BirthDate ='1950-99' WHERE
BirthDate BETWEEN 1950-01-01 AND
1999-12-31';
```

Suppression: In the case of achieving *k-anonymity*, we had to suppress some values that appeared once, yet still we had to ensure the utility of the data set, as too much suppression would kill the utility of the published dataset.

Sex	BirthDate	Nationality	Hall	Program	Year
F	196*	UGANDAN	AFRICA	LIS	2008
F	196*	UGANDAN	MARY STUART	ARM	2008
F	196*	KENYAN	MARY STUART	BLE	2007
F	196*	TANZANIA	MARY STUART	LIS	2008
M	196*	UGANDAN		RAM	2007
F	196*	KENYAN	AFRICA	ARM	2008
F	196*	TANZANIA	MARY STUART	RAM	2007
M	196*	UGANDAN		BLE	2008
F	196*	UGANDAN	COMPLEX	ARM	2007
M	196*	KENYAN	AFRICA	ARM	2008

Table 3: *Results after suppression, highlighted values to be further suppressed until k>1*

MySQL implementation:

```
UPDATE V2_Generalize1 set Hall
=' 'WHERE Hall ='Complex';
```

Check for *k-anonymity* that *k >1* by ordering data:

MySQL implementation:

```
SELECT Sex, BirthDate, Nationality, Hall, Program, Year FROM V2 ORDER BY Sex, Program, Hall;
```

k-anonymity achieved at *k>1*, where k is each value in the quasi attributes repeated at least *k >1* times.

Sex	BirthDate	Nationality	Hall	Program	Year
F	196*	UGANDAN	AFRICA	LIS	2008
F	196*	UGANDAN	MARY STUART	ARM	2008
F	196*	KENYAN	MARY STUART	BLE	
F	196*	TANZANIA	MARY STUART	LIS	
M	196*	UGANDAN		RAM	2007
F	196*	KENYAN	AFRICA	ARM	2008
F	196*	TANZANIA	MARY STUART	RAM	
M	196*	UGANDAN		BLE	2008
F	196*	UGANDAN		ARM	2007
M	196*	KENYAN	AFRICA	ARM	2008

Table 4: *Results after we achieve k-anonymity at k>1*

Removing names and student numbers entirely diminishes utility, in that the data becomes meaningless to students who simply want to view it to see if their names are on the university admission list. One way this problem can be dealt with is by publishing a list that includes the *student number* or *student names* while obscuring other PII data. However, in both scenarios, the issue of balancing data utility and data privacy remain quite challenging and demands tradeoffs [47].

8. Conclusion

We have made the case for the need to revamp Uganda's data privacy policy to encompass both private and government sectors on how to gather and disseminate data, and the need to implement data de-identification techniques. With the growth of data transaction in Uganda, there is a need for more research on how to implement privacy preserving data publishing and privacy preserving data mining methodologies tailored to the Ugandan context, with applications ranging from academia, government, health sector, and private sector. We have shown that with freely available open source technologies, some level of data privacy can be implemented on datasets from emerging markets. However, the problem of what PII constitutes in the emerging market nations still remains. Although no set of PII has been proposed in Uganda, we suggest that PII include any information that could specifically identify an individual in the Ugandan context. This could include: full names, face, fingerprints, handwriting, genetic data such as DNA, national ID number, driver's license number, passport number, credit and debit card numbers birth-date, birth place, village of residence, city of residence, county of residence, phone number, and student examination numbers. Applying the *k-anonymity* procedure might be practicable in the Ugandan context; however, achieving optimal privacy while maximizing utility continues to be an NP-hard problem, as data is lost through generalization and suppression process. Therefore more studies need to be done on various implementations of optimal data privacy tailored to Ugandan context; with consideration that PII differs in Uganda from other geographical locations.

9. References

[1] International Telecommunications Union, ITU Free statistics, 2009.

[2] International Telecommunications Union, The World In 2010 The Rise of 3G, 2010.

[3] MUK, Makerere University 2010 Admission List, Academic Registrar's Department, 2010.

[4] The Electoral Commission of Uganda, Online Voter's Register, 2010. http://www.ec.or.ug/

[5] USDOJ, “The Privacy Act of 1974. 5 U.S.C. § 552a”, 1974.

[6] USGPO, HIPAA of 1996-H. Rept.104-736, U.S. Govt Printing Office, 1996.

[7] US Library of Congress, 2009. Personal Data Privacy and Security Act of 2009– S.1490, THOMAS (Library of Congress).

[8] Embassy of the Republic of Uganda, Washington DC, The Constitution of The Republic of Uganda, 1995.

[9] UBS, The Bureau Of Statistics Act 12 1998, Uganda Gazette No.36 Volume XCI, 11th June, 1998.

[10] UCC, Uganda Communications Commission Regulations, 2010.

[11] Privacy International, PHR2006 - Republic Uganda, Constitutional Privacy Framework, 2007.

[12] Ministry of ICT, Ministerial Policy Statement for Ministry of ICT 2007/2008 Presented to Paliament, June 2006.

[13] Ministry of ICT, Ministerial Policy Statement for Ministry of ICT 2009/2010 Presented to Paliament, June 2009.

[14] Ministry of Works, National Information and Communication Technology Policy, October 2003.

[15] Nakyeyune, F., An Internal Intrusion Prevention Model, Makerere University Research Repository, 2009.

[16] Mutebi, R.M., and Rai, I.A., An Integrated Victim-based Approach Against IP Packet Flooding Denial of Service, IJCIR 2010. pp. 295-311.

[17] Makori, A.C. and Oenga, L., A Survey of Information Security Incident Reporting for Enhanced Digital Forensic Investigations, IJCIR 2010. pp.19-31

[18] Kizza, J.M., et al., Using Subgraph Isomorphism as a Zero Knowledge Proof

Authentication in Timed Wireless Mobile Networks, IJCIR 2010. pp. 334-351.
[19] Mirembe, D.P. and Muyeba, M., Security Issues in Ambulatory Wireless Sensor Networks (AWSN): Security Vs Mobility, IJCIR 2009. pp.289-301.
[20] Mutyaba R.B., Improving the RSA cryptographic algorithm using double encryption, Makerere Univ Research Repository, 2009.
[21] Makori, A.C., Integration of Biometrics with Cryptographic Techniques for Secure Authentication of Networked Data Access. IJCIR 2009. pp. 1-13
[22] Okwangale, F.R., and Ogao, P., Survey of Data Mining Methods for Crime Analysis and Visualisation, IJCIR 2006. pp. 322-327
[23] Bakibinga, E.M., Managing Electronic Privacy in the Telecommunications Sub-sector: The Ugandan Perspective. Africa Electronic Privacy and Public Voice Symposium, 2004.
[24] Luyombya, D., Framework for Effective Public Digital Records Management in Uganda. Doctoral Thesis, UCL(University College London), 2010.
[25] Ssekibule, R., and Mirembe, D.P., Security Analysis of Remote E-Voting," Makerere University Research Repository, 2007.
[26] Kayondo, L.F., A Framework for Security Management of Electronic Health Records By, Makerere University Research Repository, 2009.
[27] Sweeney, L., k-anonymity: A Model for Protecting Privacy, IJUFKS, 2002. pp. 557-570.
[28] UBS, The Bureau Of Statistics Act 12 1998, Acts Supplement No.7, The Uganda Gazette No.36 Volume XCI, 11th June, 1998.
[29] Ciriani, V., et al, Secure Data Management in Decentralized System, Springer, ISBN 0387276947, 2007, pp 291-321, 2007.
[30] Denning, D. E. and Denning, P.J., Data Security, ACM Computing Surveys, Vpl. II,No. 3, September 1, 1979.
[31] U.S. DHS, Handbook for Safeguarding Sensitive PII at The DHS, October 2008.
[32] McCallister, E. and Scarfone, K., Guide to Protecting the Confidentiality of PII, Recommendations of the NIST, 2010.
[33] Ganta, S.R., et al, 2008. Composition attacks and auxiliary information in data privacy, Proceeding of the 14th ACM SIGKDD 2008, p. 265.
[34] Oganian, A. and Domingo-Ferrer, J., On the complexity of optimal micro-aggregation for statistical disclosure control, Statistical Journal of the United Nations Economic Commission for Europe, Vol. 18, No. 4. (2001), pp.345-353.
[35] Rastogi et al, The boundary between privacy and utility in data publishing, VLDB ,September 2007, pp. 531-542.
[36] Sramka et al, A Practice-oriented Framework for Measuring Privacy and Utility in Data Sanitization Systems, ACM, EDBT 2010.
[37] Sankar, S.R., Utility and Privacy of Data Sources: Can Shannon Help Conceal and Reveal Information?, presented at CoRR, 2010.
[38] Wong, R.C., et al, Minimality attack in privacy preserving data publishing, VLDB, 2007. pp.543-554.
[39] Adam, N.R. and Wortmann, J.C., A Comparative Methods Study for Statistical Databases: Adam and Wortmann, ACM Comp. Surveys, vol.21, 1989.
[40] Narayanan, A. and Shmatikov, V., Myths and fallacies of "personally identifiable information". Comm. ACM. 2010, 24-26.
[41] Brewster, K.F., 1996. The National Computer Security Center (NCSC) Technical Report - 005 Volume 1/5 Library No. S-243,039, 1996.
[42] Bayardo, R.J., AND Agrawal, R., Data Privacy through Optimal k-anonymization, ICDE, 2005. pp. 217-228.
[43] Ciriani, V., et al, Theory of privacy and anonymity. In Algorithms and theory of computation handbook (2 ed.), 2010.
[44] Samarati, P. and Sweeney, L., Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression, IEEE Symp on Research in Security and Privacy, 1998, pp. 384–393.
[45] Samarati, P., Protecting Respondent's Privacy in Microdata Release. IEEE on TKDE, 2001. pp. 1010-1027.
[46] Meyerson, A., and Williams, R., On the complexity of optimal K-anonymity. ACM PODS, 2004. pp. 223-228.
[47] Rastogi et al, The boundary between privacy and utility in data publishing, VLDB, September 2007, pp. 531-542.
[48] C. Kuner, European data protection law: corporate compliance and regulation. Oxford University Press, ISBN 9780199283859, 2007.