Risk-Aware Information Disclosure*

Alessandro Armando[1,2], Michele Bezzi[4],
Nadia Metoui[2,3], and Antonino Sabetta[4]

[1] DIBRIS, University of Genova, Italy
[2] Security & Trust Unit, FBK-Irst, Trento, Italy
[3] DISI, University of Trento, Italy
[4] Product Security Research, SAP Labs, Sophia-Antipolis, France

Abstract. Risk-aware access control systems grant or deny access to resources based on some notion of risk. In this paper we propose a model that considers the risk of leaking privacy-critical information when querying, e.g., datasets containing personal information. While querying databases containing personal information it is current practice to assign all-or-nothing access to avoid the disclosure of sensitive information. Using our model, access-control decisions are based on the disclosure-risk associated with a data access request and, differently from existing models, we include adaptive anonymization operations as risk-mitigation methods. By applying these operations, a request that would otherwise be rejected, is permitted after reducing the risk associated with the returned dataset.

1 Introduction

The increasing availability of large and diverse datasets (*Big Data*, such as customer data, transactions, demographics, product ratings) helps businesses to get insights on their markets and their customers' needs, and predict what is next. It is also boosting the creation of new *data monetization* businesses, where companies package their data and sell them to other organizations. According to IDC [17] the market for Big Data business will reach 16.9 billion USD by 2015, up from 3.2 billion USD in 2010.

The full exploitation of big data raises various issues on the possible disclosure of sensitive or private information. In particular, big data often contain a large amount of personal information, which is subject to multiple and stringent regulations (EU data protection directive, HIPAA[5], etc.). These regulations impose strong constraints on the usage and transfer of personal information, which make their handling complex, costly, and risky from a compliance point of view. As a consequence, personal data are often classified as confidential information, and only a limited number of business users (e.g., high level managers) have access to them, and under specific obligations (e.g., within the perimeter of the

* This work has been partly supported by the EU under grant 317387 SECENTIS (FP7-PEOPLE-2012-ITN)

[5] HIPAA: Health Insurance Portability and Accountability Act

company network, no transfer to mobile devices, etc.). As a matter of fact, because of the difficulty of dealing with the potential privacy implications in an efficient and systematic way, an *all-or-nothing* decision is often followed; by using this approach, many business users are just prevented from retrieving data from databases as soon as these databases contain, even if only in few specific tables, some personal information. However, many business applications (e.g., business analytics and reporting, recommendation systems) do not need all the personal details on specific individuals, and an *anonymized version* of the dataset is still an asset of significant value that can address the business requirements in most cases.

Anonymization methods can be applied to obfuscate the personal identifiable information, such as suppressing part of or entire records; generalizing the data, i.e., recoding variables into broader classes (e.g., releasing only the first two digits of the zip code) or rounding numerical data; replacing identifiers with random values (e.g., replacing a real name with a randomly chosen one), randomly swapping some attributes in the original data records, applying permutations or perturbative masking, i.e., adding random noise to numerical data values.

To assess the level of anonymity, several metrics have been proposed in the literature (see [3, 8] for a review). These metrics differ in a number of ways, but they all express the risk of disclosing personal-identifiable information when releasing a given dataset. Based on these metrics, several anonymization methods have also been put forth [7]. These methods increase protection by lowering the privacy risk and by enabling a wider exploitation of the data, but they assume the accepted risk level is statically given. In practice the accepted risk level may depend on a number of factors that can only be computed at run-time, e.g., the trustworthiness or the competence of the user or the quality of the security context used to issue the query.

In this paper we propose an access control model for risk-aware information disclosure. In our model access-control decisions are based on the disclosure-risk associated with a data access request and, differently from existing models, we include adaptive anonymization operations as risk-mitigation methods. By applying these operations, a request that would otherwise be rejected, is permitted after reducing the risk associated with the returned dataset.

Structure of the paper. In the next section we provide a representative, real-world scenario that illustrates the motivation for risk-aware information disclosure. In Section 3 we recall some background notions on risk-aware access control and privacy preserving information disclosure. In Section 4 we present our access control model for risk-aware information disclosure and in Section 5 we illustrate its application on the scenario introduced in Section 2. In Section 6 we discuss the related work. We conclude, in Section 7, with some final remarks.

2 Scenario

Employee surveys are a widely used instrument for organizations to assess job satisfaction, quality of management, people motivation, etc. Considering the

possible sensitiveness of data, surveys should be anonymous, meaning that the organization and management should not be able to identify how a specific employee responded. Usually, the organization – say, a large company – conducting the survey outsources the data collection to a third-party. When processing the data, the third-party has access to individual-level information, whereas this data is not accessible to the company. To protect the anonymity of the survey, the company can access the data under the condition that *(i)* identifiers are removed and *(ii)* the number of respondents is larger than a certain thresholds (usually between 10 and 25). Different splits of data can be requested (e.g., per organization, per job profile, etc.), but data are accessible only if the query results contains a number of respondents that is larger than the fixed thresholds. On top of that, additional access control rule can be enforced, e.g., a manager would only see data referring to his/her team or department (provided that conditions *(i)* and *(ii)* are also fulfilled); an employee would be allowed to see overall (company results) only. As an example, consider a question like "Do you respect your manager as a competent professional?" with a five points scale (1 to 5). A manager could see the response of his/her team if at least, say, 10 people answered to it. If the manager decides to refine the analysis asking for data related to the people in his/her team AND with a "developer" role, again the response should be made available only if at least 10 respondents with that role answered to the question.[6]. Current systems typically do not provide any data if the number of respondents is below the defined thresholds (for the specific role). In other words, in order to avoid the risk of disclosing too much information, an overly conservative approach is taken and problematic queries are not permitted altogether. Ideally, the access control system should be able to provide the largest possible amount of information (still preserving anonymity) for any query. In practice, in presence of queries that might cause anonymity issues (i.e., not enough respondents, or more generally, too small result set), the system should be able to quantify the disclosure risk associated with the query and compare it with whatever risk level has been set as the acceptable threshold. If the threshold is exceeded, the system could apply, for example, a "generalization" operation (making the query less specific), thus increasing the cardinality of the result set and reducing the risk of disclosing the identity of respondents. Of course, applying such operation would not yield the *exact* data set the user asked for, but this method would: 1) provide some relevant (i.e., as close as possible to the original query) information to the user, and 2) preserve anonymity according to some pre-defined disclosure-risk levels (possibly linked to the requestor trust or role).

In the next section,we discuss how to implement such a system using risk-based access control, and anonymization mitigation strategies.

[6] In real surveys single records are actually never shown, but just percentages, in this example it would be something like 10% answered 1, 25% answered 2, etc. Since the number of respondents is known, in practice, for one question, this equivalent of getting the data with no identifiers.

3 Background

In this section, we present a Risk Aware Access Control model introduced in earlier work by Chen et al. [5, 4]. We also present some privacy concepts and the "k-anonymity" model for preserving privacy [18], since it is the mostly used metrics for anonymity for surveys.

3.1 Risk-Aware Access Control

We provide a brief presentation of the formal model for Risk-Aware Access Control (RAAC) that has been introduced in [5]. We use this model as the basis of our access control model for risk-aware information disclosure that is presented in Section 4.

Formally, a RAAC consists of the following components:

- a set of users U;
- a set of permissions P, usually representing action-object pairs;
- a set of access requests Q, modeled as pairs of the form (u, p) for $u \in U$ and $p \in P$;
- a set of *risk mitigation methods* $\mathcal{M}$, i.e., actions that are required to be executed to mitigate risk;
- a function π mapping permissions into *risk mitigation strategies*, i.e., lists of the form $[(l_0, M_0), (l_1, M_1), \ldots, (l_{n-1}, M_{n-1}), (l_n, M_n)]$, where $0 = l_0 < l_1 < \cdots < l_{n-1} < l_n \leq 1$ and $M_i \in \mathcal{M}$ for $i = 0, \ldots, n$;
- a set of *states* Σ, i.e., tuples of the form (U, P, π, τ) where τ abstracts further specific features of the state; for instance, in the Risk-Aware Role-Based Access Control (R^2BAC) model [4], τ comprises the set of roles R, the user-role assignment relation $UA \subseteq U \times R$, the role-permission assignment relation $PA \subseteq P \times R$, the role hierarchy $\succeq \subseteq R \times R$, and the user trustworthiness $\alpha : U \to (0..1]$, the user-role competence function $\beta : U \times R \to (0..1]$, and the role-permission appropriateness function $\gamma : R \times P \to (0..1]$;
- a *risk function* $risk : Q \times \Sigma \to [0..1]$ such that $risk(q, \sigma)$ denotes the risk associated to granting q in state σ;
- an *authorization decision function* $Auth : Q \times \Sigma \to D \times 2^{\mathcal{M}}$ with $D = \{\mathsf{allow}, \mathsf{deny}\}$ such that if $q = (u, p)$ and $\pi(p) = [(l_0, M_0), \ldots, (l_n, M_n)]$, and σ the current state, then

$$Auth(q, \sigma) = \begin{cases} (d_i, M_i) & \text{if } risk(q, \sigma) \in [l_i, l_{i+1}), i < n, \\ (d_n, M_n) & \text{otherwise} \end{cases}$$

where $d_i \in D$. Intuitively, if the risk associated with access request (u, p) is l, then $Auth$ returns an authorization decision and a set of risk mitigation methods corresponding to the interval containing l.

3.2 Privacy Preserving Information Disclosure

From a data privacy standpoint, the data stored in database tables and the columns (data attributes) of the tables can be classified as follows.

- *Identifiers.* These are data attributes that can uniquely identify individuals. Examples of *identifiers* are the Social Security Number, the passport number, the complete name.
- *Quasi-identifiers (QIs) or key attributes* [9]. These are the attributes that, when combined, can be used to identify an individual. Examples of *quasi-identifiers* are the postal code, age, job function, gender, etc.
- *Sensitive attributes.* These attributes contain intrinsically sensitive information about an individual (e.g., diseases, political or religious views, income) or business (salary figures, restricted financial data or sensitive survey answers).

Various anonymity metrics have been proposed so far (see [3, 8] for a review). In this paper we concentrate on a very popular metric, k-Anonymity [18]. Other metrics are presented in Section 6. k-Anonymity condition requires that *every* combination of quasi-identifiers is shared by at least k records in the anonymized dataset. A large k value indicates that the anonymized dataset has a low identity privacy risk, because, at best, an attacker has a probability $1/k$ to re-identify a record (i.e., associate the sensitive attribute of a record to the identity of a person).

4 Risk-Aware Information Disclosure

We now refine the RAAC model of Section 3.1 into our model for Risk-Aware Information Disclosure. Let P be a set of database views (or virtual tables). If p is a view, then $|p|$ denotes the anonymity of p according to some given metrics (e.g. k-anonymity). The higher is the value of $|p|$, the smaller is the risk to disclose sensitive information by releasing p. Thus, for instance, we can define *the (privacy) risk of disclosing* p to be $1/|p|$ and *the (privacy) risk of disclosing* p *to* u *in* $\sigma = (U, P, \pi, \tau)$ to be

$$risk((u,p),\sigma) = \begin{cases} 1 & \text{if not } granted_\tau(u,p) \\ 1/|p| & \text{otherwise} \end{cases}$$

where $granted_\tau(u,p)$ holds if and only if u is granted access to p according to τ. For instance, if τ is an RBAC policy $(U, R, P, UA, RA, \succeq)$, then $granted_\tau(u,p)$ holds if and only if there exist $r, r' \in R$ such that $(u,r) \in UA$, $r \succeq r'$, and $(p,r') \in PA$.

When the risk associated to the disclosure of a certain view p is greater than the maximal accepted risk t, we can use obligations for obfuscating or redacting the view and thus bring the risk below t. In this paper we consider k-anonymization functions $\phi_k : P \to P$ for $k \in \mathbb{N}$ as risk mitigation methods, but functions based on other metrics can be used as well. Clearly $|\phi_k(p)| \geq k$

for all $p \in P$. We then consider risk mitigation strategies of the form $\pi(p) = [(0, \iota), (t, \phi_{\lceil 1/t \rceil}(.))]$, where $\iota : P \to P$ is the identity function (i.e. such that $\iota(p) = p$ for all $p \in P$) and the following authorization decision function:

$$Auth((u,p),\pi) = \begin{cases} (\mathsf{allow}, \iota) & \text{if } risk(u,p) < t, \\ (\mathsf{allow}, \phi_{\lceil 1/t \rceil}(\cdot)) & \text{if } risk(u,p) \geq t \end{cases}$$

that always grants access but yields an anonymized version of the requested view if the risk is greater that the maximal accepted risk t. In other words, if user u asks to access p, then access to p is granted unconditionally if $risk(u,p) < t$, otherwise an anonymized version of p, say $\phi_{\lceil 1/t \rceil}(p)$, is computed and returned to u.

Example 1. To illustrate assume Alice asks for a view p_1 such that $|p_1| = 4$ and that $\pi(p_1) = [(0, \iota), (t, \phi_{\lceil 1/t \rceil}(.))]$ with $t = 0.1$, i.e. $\pi(p_1) = [(0, \iota), (0.1, \phi_{10}(.))]$. It is easy to see that $risk(Alice, p_1) = 0.25$ and that $Auth((Alice, p_1), \pi) = \phi_{10}(p_1)$.

Alice then asks for a view p_2 such that $|p_2| = 20$ and that $\pi(p_2) = \pi(p_1) = [(0, \iota), (t, \phi_{\lceil 1/t \rceil}(.))]$ with $t = 0.1$, i.e. $\pi(p_2) = [(0, \iota), (0.1, \phi_{10}(.))]$. It is easy to see that now $risk(Alice, p_2) = 0.05$ and therefore that $Auth((Alice, p_2), \pi) = \iota(p_1) = p_1$.

The following results state that the risk of disclosing the view returned by our authorization decision function is never greater than the maximum accepted risk.

Proposition 1. *Let* $(D, M) = Auth((u,p),\pi)$. *Then* $risk(u, M(p)) \leq t$.

In many situations of practical interest, we want the risk of a query $q = (u,p)$ to depend also on the trustworthiness of the user u. This can be done by (re)defining the risk function as follows:

$$risk((u,p),\sigma) = \begin{cases} 1 & \text{if not } granted_\tau(u,p) \\ \max\{0, \frac{1}{|p|} - \alpha(u)\} & \text{otherwise} \end{cases} \quad (1)$$

where $\alpha : U \to (0..1]$ is a function that assigns a trust value to users.

When roles correspond to job functions, it is natural to assign trust to roles and to derive the trust of a user from the trust assigned to the roles assigned to that user in the following way:

$$\alpha(u) = \max\{\alpha(r') : (p, r') \in PA \text{ and } \exists r \succeq r' \text{ s.t. } (u,r) \in UA\}.$$

5 Application of Risk-Aware Role-Based Access Control

We now show how our risk-aware information disclosure model can be used to support the scenario of Section 2. This will be done by setting appropriate values to the parameters occurring in the definition of the risk function (1).

For sake of simplicity we consider a small company, with 8 employees and one manager. The company runs an employee survey, with one single question with answer ranging in a five points scale (from 1 to 5) (*sensitive attribute*, cf. Section 3.2), and collecting user names[7] (the *identifiers*), as well as the job title and the location of the office (the *quasi-identifiers*). The actual dataset is in Table 1(a). To preserve privacy we set the maximal acceptable risk to $t = 0.125$.

Table 1. The Employee Survey Example

(a) Original dataset

Survey Administrator view $\lvert p_{all}\rvert = 1$			
Name	Job	Location	**Answer**
Timothy	SeniorDeveloper	Houston	4
Alice	Support	Houston	5
Perry	JuniorDeveloper	Rome	5
Tom	Admin	Rome	3
Ron	SeniorDeveloper	London	4
Omer	JuniorDeveloper	London	4
Bob	Support	Houston	5
Amber	Admin	Houston	3

(b) Anonymized version: *identifiers* and *quasi-identifiers* are suppressed

Employee View $\lvert p_{supp}\rvert = 8$			
Name	Job	Location	**Answer**
***	***	***	4
***	***	***	5
***	***	***	5
***	***	***	3
***	***	***	4
***	***	***	4
***	***	***	5
***	***	***	3

The outsourcing company collecting the data is considered fully trusted and will therefore have access to all the information. We model this by setting the trust of the `admin` role to 1, i.e. $\alpha(\texttt{admin}) = 1$. Thus, an administrator can access the original dataset, say p_{all} with anonymity $|p_{all}| = 1$ (i.e., all distinct values, see Table 1(a)), since $\alpha(\texttt{admin}) = 1$ and the risk value is smaller than the threshold, i.e., $1 - 1 = 0 < 0.125$. If we set the trust value of the `manager` role to 0.21, i.e. $\alpha(\texttt{manager}) = 0.21$ (corresponding to access views with anonymity $k \geq 3$), than a manager cannot access p_{all} as is, since $1 - 0.21 > 0.125$ and some anonymization, as risk mitigation strategy, must be carried out on the data to decrease the risk. For example, if we suppress the identifier attribute (*Name*) and the quasi-identifiers (*Job* and *Location*), we obtain the view p_{supp} shown in Table 1(b). The view p_{supp} corresponds to an anonymity level $|p_{supp}| = 8$ and since $0.125 - 0.21 < 0.125$, access is granted to the manager.[8] The manager can also ask for more granular views of the results. For example, if she wants to know the distribution of the answers in one location, say Houston, $|p_{Houst}| = 4$, the risk $0.25 - 0.21 = 0.04$ is still smaller than $t = 0.125$. On the other hand, if she asks for the result in Rome, $|p_{Rome}| = 2$, then the risk associated with the view for

[7] In real cases they are typically user IDs

[8] In real surveys the result will appear as a report like: 37.5% answered 5, 37.5% answered 4 and 25% answered 3. For a single question this is equivalent to the view in Table 1(b).

Table 2. Views of the employee survey for the Rome location

(a) Before generalization.

View: Location=Rome, $\lvert p_{Rome}\rvert = 2$			
Name	Job	Location	**Answer**
***	***	Rome	5
***	***	Rome	3

(b) After generalization

View: Location=Rome Anoymized $\lvert p_{EMEA}\rvert = 4$			
Name	Job	Location	**Answer**
***	***	EMEA	5
***	***	EMEA	3
***	***	EMEA	4
***	***	EMEA	4

the manager is $0.5 - 0.21 > 0.125$ and the access is granted only if appropriate anonymization is performed. In this case, location could be generalized from Rome to EMEA (so including London workforce), as shown in Table 2(b). The resulting view has anonimity $|p_{EMEA}| = 4$ and since the risk is smaller than $t = 0.125$, then the manager is allowed to see the view.

Table 3. Views of the employee survey for Rome and JuniorDeveloper

(a) Before generalization of location and job

Loc=Rome AND Job=JuniorDeveloper $\lvert p_{Rome+JuniorDeveloper}\rvert = 1$			
Name	Job	Location	**Answer**
***	JuniorDeveloper	Rome	5

(b) After generalization of location and job

View Loc=Rome AND Job=JuniorDeveloper Anonymized $\lvert p_{EMEA+Dev}\rvert = 3$			
Name	Job	Location	**Answer**
***	Dev	EMEA	5
***	Dev	EMEA	4
***	Dev	EMEA	4

Similarly, if the manager wants to see the results per location and per job function (say in Rome for JuniorDeveloper only, see Table 3(a)), the anonymity level is low, $|p_{Rome+JuniorDeveloper}| = 1$, and the associated risk is greater than $t = 0.125$. Again, instead of simply denying access, the system can perform generalization on both the quasi-identifiers, *Job* (generalized to the job family developer) and *Location*, thereby increasing the anonymity ($|p_{EMEA+Dev}| = 3$) and decreasing the risk ($risk(manager, p_{EMEA+Dev}) = 0.123$) to an acceptable level for a manager (see Table 3(b)).

Finally, employees should have access to the global results only. The trust value is therefore set to $\alpha(\texttt{employee}) = 0.125$ and the only view permitted is with suppression of all identifiers and quasi-identifiers, which has $|p_{supp}| = 8$, see Table 3(b).

6 Related Work

Risk-aware access control (see, e.g., [4–6, 10, 19]) has received a growing attention in the last few years. However, little attention is given to privacy aspects. The approaches that address privacy (see, e.g., [16, 14]) do so by adding privacy policy enforcement on top of the access control evaluation process. In our approach privacy risk as well as access risk are evaluated for every access request.

Risk Aware Access Control Models generally determine the risk as a function of the likelihood of a permission misuse and the cost of the permission authorized and misused. The likelihood of misuse can depend on the user trustworthiness and competence [4], the user behavior [1], and the uncertainty of the access decision [15]. The quantification of the cost of permission misuse has been addressed by several researches. Cheng et al. [6], in their assign a sensitivity label to every resource. The value of a resource is then determined according to its sensitivity. The cost of a misused permission depends on the resource's value. Molloy et al. [15] and Baracaldo et al. [1] propose to evaluate the cost in term of financial gain and damage. Chen and Crampton [4] do not explicitly calculate the permission misuse cost in their model, but mention that the cost of misuse is valued and used to define risk thresholds and risk mitigation strategies for every permission. In our model the risk results from the likelihood of identity disclosure which depends on the sensitivity of the requested information and the requestor trustworthiness.

Chen et al. [5, 12] propose to use, both user and system obligations as risk mitigation methods. An obligation describes some actions that have to be fulfilled by the subject, the system or a third part (e.g.an administrator), in a specific time window. In the literature we can distinguish between two categories of obligations: *provisions* or *pre-obligations* [2] are actions that must be executed prior to making an authorization decision; *post-obligations* are actions that must be fulfilled after the authorization decision is made. Unlike Chen et al. models that use post-obligations, monitor the fulfillment of these obligations after granting access and reward or punish users according to whether they have succeed or not to fulfill the required action, in our model we use provisions to enforce the risk mitigation strategy at run-time.

In this paper we consider only k-anonymity as anonymity metrics, but alternative metrics do exist. A group (with minimal size of k records) sharing the same combination of quasi-identifiers could also have the same sensitive attribute, so even if the attacker is not able to re-identify the record, he can discover the sensitive information (attribute disclosure). To capture this kind of risk ℓ-diversity was introduced [13]. The ℓ-diversity condition requires that for *every* combination of key attributes there should be at least ℓ values for each confidential attribute. Although, ℓ-diversity condition prevents the possible attacker from inferring exactly the sensitive attributes, he may still learn a considerable amount of probabilistic information. More specifically, if the distribution of confidential attributes within a group sharing the same key attributes is very dissimilar from the distribution over the whole set, an attacker may increase his knowledge on sensitive attributes (*skewness attack*, see [11] for details). To overcome the prob-

lem, t-closeness [11] estimates this risk by computing the distance between the distribution of confidential attributes within the group and in the entire dataset. These measures provide a quantitative assessment of the different risks associated to data release, and each of them (or a combination thereof) can be applied to estimate privacy risk depending on the use case at hand.

7 Conclusions

We have presented a model for information disclosure where access-control decisions are based on the risk associated with a data access request. Anonymization operations are used as risk-mitigation methods to compute views satisfy the accepted level of risk. This allows for granting access to requests that would otherwise be rejected. Our model leverages existing modes for Risk-Aware Access Control (most notably [5, 4]) but it also shows how they can be adapted so to support the controlled disclosure of privacy-sensitive information.

References

1. Nathalie Baracaldo and James Joshi. A trust-and-risk aware rbac framework: Tackling insider threat. In *Proceedings of the 17th ACM Symposium on Access Control Models and Technologies*, SACMAT '12, pages 167–176, New York, NY, USA, 2012. ACM.
2. Claudio Bettini, Sushil Jajodia, X. Sean Wang, and Duminda Wijesekera. Provisions and obligations in policy management and security applications. In *Proceedings of the 28th International Conference on Very Large Data Bases*, VLDB '02, pages 502–513. VLDB Endowment, 2002.
3. Michele Bezzi. An information theoretic approach for privacy metrics. *Transactions on Data Privacy*, 3(3):199–215, 2010.
4. Liang Chen and Jason Crampton. Risk-aware role-based access control. In Catherine Meadows and Carmen Fernandez-Gago, editors, *Security and Trust Management*, volume 7170 of *Lecture Notes in Computer Science*, pages 140–156. Springer Berlin Heidelberg, 2012.
5. Liang Chen, Jason Crampton, Martin J. Kollingbaum, and Timothy J. Norman. Obligations in risk-aware access control. In Nora Cuppens-Boulahia, Philip Fong, Joaquín García-Alfaro, Stephen Marsh, and Jan-Philipp Steghöfer, editors, *PST*, pages 145–152. IEEE, 2012.
6. Pau-Chen Cheng, Pankaj Rohatgi, Claudia Keser, Paul A. Karger, Grant M. Wagner, and Angela Schuett Reninger. Fuzzy multi-level security: An experiment on quantified risk-adaptive access control. In *IEEE Symposium on Security and Privacy*, pages 222–230. IEEE Computer Society, 2007.
7. V. Ciriani, S. De Capitani di Vimercati, S. Foresti, and P. Samarati. Theory of privacy and anonymity. In M. Atallah and M. Blanton, editors, *Algorithms and Theory of Computation Handbook (2nd edition)*. CRC Press, 2009.
8. Chris Clifton and Tamir Tassa. On syntactic anonymity and differential privacy. *Trans. Data Privacy*, 6(2):161–183, August 2013.
9. T. Dalenius. Finding a needle in a haystack-or identifying anonymous census record. *Journal of official statistics*, 2(3):329–336, 1986.

10. Luke Dickens, Alessandra Russo, Pau-Chen Cheng, and Jorge Lobo. Towards learning risk estimation functions for access control. In *In Snowbird Learning Workshop*, 2010.
11. Ninghui Li, Tiancheng Li, and S. Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on*, pages 106–115, April 2007.
12. Timothy J. Norman Liang Chen, Luca Gasparini. XACML and risk-aware access control. Technical report, 2013.
13. Ashwin Machanavajjhala, Johannes Gehrke, Daniel Kifer, and Muthuramakrishnan Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. In *ICDE '06: Proceedings of the 22nd International Conference on Data Engineering (ICDE'06)*, page 24, Washington, DC, USA, 2006. IEEE Computer Society.
14. L.D. Martino, Q. Ni, D. Lin, and E. Bertino. Multi-domain and privacy-aware role based access control in ehealth. In *Pervasive Computing Technologies for Healthcare, 2008. PervasiveHealth 2008. Second International Conference on*, pages 131–134, Jan 2008.
15. Ian Molloy, Luke Dickens, Charles Morisset, Pau-Chen Cheng, Jorge Lobo, and Alessandra Russo. Risk-based security decisions under uncertainty. In *Proceedings of the Second ACM Conference on Data and Application Security and Privacy*, CODASPY '12, pages 157–168, New York, NY, USA, 2012. ACM.
16. Qun Ni, Alberto Trombetta, Elisa Bertino, and Jorge Lobo. Privacy-aware role based access control. In *Proceedings of the 12th ACM Symposium on Access Control Models and Technologies*, SACMAT '07, pages 41–50, New York, NY, USA, 2007. ACM.
17. IDC Report. Worldwide big data technology and services 2012-2015 forecast. *IDC Report*, 2012.
18. Pierangela Samarati. Protecting respondents' identities in microdata release. *IEEE Trans. Knowl. Data Eng.*, 13(6):1010–1027, 2001.
19. Riaz Ahmed Shaikh, Kamel Adi, and Luigi Logrippo. Dynamic risk-based decision methods for access control systems. volume 31, pages 447–464, 2012.