The protection of statistical databases is the problem with the highest importance for many organizations, in particular the public administration. An appropriate method for statistical databases protection has to ensure a balance between safety and quality of data, taking into account the associated risks and business needs. To improve security of statistical databases integration of protection mechanisms with the enterprise security architecture (i.e. using SABSA methodology) is proposed. Coherent and comprehensive approach of the protection is a result of the proposal. It allows statistical databases safety to be in line with business requirements and ensure that statistical databases security mechanisms are integrated with existing enterprise security mechanisms and services.

Keywords: enterprise security architecture, SABSA.

1. Introduction

Statistical databases are used to store anonymous data. This data is available to the user only at the level of statistical groups, and reserved at a unit-level. The main difference between the statistical databases and relational databases is a very limited set of allowed queries. Inquiries are mainly limited to operations such as the number of a set, the sum, average, and several other statistical calculations that are performed on a subset relationship.

The problem of the statistical databases protection is the provision of statistics on people, while limiting the ability to access information about any specific individual. Achieving this is extremely difficult, because each statistic contains derivative information on the basis of which was calculated and, therefore, any provision of accurate statistics is a risk factor of disclosing private information [2]. In addition, the correlation of different statistics may lead to the disclosure of exact or approximate information of private data. An additional difficulty is the protection of sensitive personal data from the user having knowledge of the characteristic values of an individual acquired outside the statistical database [6].

Advanced researches in the field of protection of statistical databases, as well as existing solutions usually involve two basic approaches that are often used in combination. The first approach is to restrict access to database systems where access is granted only to authorized and authenticated individuals to carry out the approved analysis. Additional restrictions may also be introduced at the level of the results of statistical analysis, making it impossible to publish outside of a statistics organization (i.e. Eurostat, the statistical office in France, the Netherlands, Denmark, or Sweden) [13]. The second approach is to reduce or replace data, where less than a full set of data is published, or the data has been distorted in some way before publication. There are many methods of statistical databases by manipulating the data that are the object of analysis (while maintaining statistical accuracy of resulting analysis). The importance of the protection of privacy in statistical databases is growing nowadays in Poland mainly because of the national digitalization of health care and IT support for the Agricultural Census 2010 and National Census and Dwellings 2011.

For a coherent and comprehensive approach to the security of statistical databases, an integration of the existing security mechanisms with the enterprise security architecture is advisable. The security architecture is a coherent set of principles, guidelines, benchmarks and standards of security, along with a description of their interconnection, and taking into account the circumstances and the needs of business organizations. To implement the security architecture is to provide adequate mechanisms for business needs, which will reduce the risk of data breaches.
An example of the methodology used to manage the security architecture is SABSA. (Sherwood Applied Business Security Architecture) SABSA methodology is used by many organizations all over the world to develop the security architecture, in particular in the public administration. An example would be the United Kingdom’s Ministry of Defence, which has chosen the SABSA methodology to implement Defence Information Security Architecture [15].

SABSA methodology presents a business approach to security within an organization. It is used for a comprehensive management of security service and security architecture. It allows the use of well-known standards as components of the architecture in a consistent manner, which together provides a comprehensive model, based on business requirements.

By using the SABSA methodology it is possible to develop security mechanisms that can guarantee the security of statistical databases in a specific business organization. At the same time, such an approach ensures the integration of security mechanisms for statistical databases with other security services and mechanisms used in the organization.

In this paper, we propose an integration of statistical database security mechanisms with enterprise security architecture via SABSA methodology. In accordance with the principles of creating the security architecture the paper covers all SABSA layers. In following chapters a model overview is presented starting from the top layer – business requirements and going down to lower layers – the implementation of the security mechanisms of statistical databases, and elements of the operational layer.

2. SABSA Model

The SABSA model is based on six layers:
- Contextual – a description of the business and business requirements
- Conceptual – a high-level description of security
- Logical – security services
- Physical – mechanisms and security methods
- Component – implementation of the selected security method
- Operational – monitoring changes in the security architecture and security audits.

Operational layer is defined in parallel with the logical, physical and component layers. In each of these layers six aspects of security are defined, which form the SABSA matrix. The SABSA matrix structure is analogous to the Zachman framework [14], but it should be noted that they were developed independently. Creating enterprise security architecture is to develop various aspects of each layer starting from the highest to the lower ones.

The approach presented in this document is based on a layered SABSA model. Selected SABSA matrix elements are used, because not all elements need to be addressed. Selected elements fully show the ease of use and usefulness of the SABSA methodology in order to integrate the security of statistical databases with the security architecture.

Fig. 1. SABSA matrix
Analysis presented in the paper are carried through all SABSA layers from top to bottom. The results of the analysis of items resulting from the previous layer are a direct source of the development of elements of subsequent layers. It should be emphasized that the result of the elaboration of the various elements of each layer will vary for different organizations.

**Contextual Layer**

The contextual layer represents the business view on the organization’s security. The type of security architecture that organization needs is defined in this layer. The aim of the contextual layer is to collect business requirements for security and create a business risk model for statistical databases.

The first step to create the security architecture is to define business requirements. Business requirements are all important aspects of security from the perspective of the organization. Defining the business requirements must be carried out, particularly in the following areas: Legal requirements – the most important source of business requirements are legal provisions in force within the country. The Polish constitution of 2nd April, 1997, introduced into the Polish legislation the principle of the protection of personal data. The principle was specified by the act of August 29th, 1997 on the protection of personal data, which is an implementation of the directive of the European Parliament and the European Council of October 24th, 1995 on the protection of individuals with regard to the processing of personal data and the free exchange of such data. Persons whose data is being processed were provided with instruments to control the process of processing of their data, while the database administrators imposed a series of obligations that maintain data security.

In the face of progressive informatisation of society, the development of new technologies and globalization issues, data safety is one of the major challenges of modern information science. It should be noted that in addition to personal data, there are other types of data, for which protection is especially important for those affected. An example could be any confidential company information. It is to protect the statistical database that provides any kind of data. However, in view of the critical safety of personal data, they are most important.

In addition, every organization is forced to comply with the laws specific to their business. An example would be the Polish Law on the National Census, where in cooperation with the Polish General Inspector of Data Protection introduced a lot of requirements to increase data security:

- The Polish Central Statistics Office may entrust personal data to other entities
- Introduced additional security of information systems used during the census
- Prohibited from sharing personal data by the Central Statistical Office for purposes other than statistical.

It is worth stressing the latter points. The Polish Central Statistical Office is authorized to release personal data exclusively for statistical requirements in the absence of adequate (in the statistical sense) secure database entities to which data are available. This is an example of the lack of an architectural approach to data security i.e. spot data protection, without taking into account the comprehensive approach to ensure real security.

- Business Strategy, including the mission, vision and objectives of the organization, which can be found, among others in:
  - State Informatisation Plan for years 2007–2010
  - Polish National Security Strategy
  - National Development Plan.
- The key resources that are relevant to the organization and must be protected by appropriate safeguards for business information and business information systems.
- Business processes to which it is necessary to implement information security. These are all processes related to:
  - Business interaction
  - Business transactions
  - Business communication.

A complete set of business requirements is a contribution to the creation of a business risk model. Creating the business risk model is to verify the business requirements for meeting them through a standard database. In each case, if the requirement is not met the risk corresponding to this requirement has to be defined, and then put to the business risk model. A full risk model describes all the business risks associated with the use of standard databases to provide statistical data. One can easily draw the conclusion that all risk mitigation is a key to the implementation of security mechanisms to protect statistical data. Risk mitigation drives an implementation of the security mechanisms protecting statistical data.

The impact and the vulnerability are described for each threat. The final risk
evaluation is the product of both components. This value determines the priority with which the risk should be mitigated. Each organization must be examined in order to select the value below, which the risk may be regarded as acceptable.

Conceptual Layer

Within the conceptual layer the business requirements are translated into a security concept. At this stage, the concept of a security method that meets business needs is created. The concept represents the vision of full security, and not its individual components.

Methods of statistical database protection affect the value of the obtained statistics. In addition to the security relationship between the data and the accuracy of the statistics obtained, an important element is the cost of implementing the method and its impact on the effectiveness of the database.

In order to define all the relevant factors affecting the quality of the operation of the statistical database, the element of the conceptual layer is used, i.e. the business attributes profile. The business attributes profile is a set of attributes describing the concept of security – in our case, factors affecting the quality of the operation of statistical databases. Profile results directly from business requirements. Normally, global standards like ISO 17799/27001 defines key business attributes: integrity, availability and confidentiality. SABSA methodology defines 85 business attributes for the entire organization. In accordance with best practices and global studies, there shall be five main attributes for statistical databases: security, information loss, cost, bias, consistency. Please note that it is possible to describe regular databases with a larger set of attributes, but only specific attributes for statistical databases were selected. Other attributes should be taken into account when creating the security architecture (for other elements of the organization).

Attributes specified for the purpose of statistical database security are selected and described below. First three attributes are of the highest importance from the business requirements perspective. Last two attributes are specific to methods based on data perturbation. Each proposed attribute is presented along with a measurement method. According to business requirements the database administrator, who is responsible for information assurance, sets target performances of every business attributes.

Measurement methods of these attributes are used for verification if the target performance criterion is met. Attributes:

- Security – a business attribute describing a statistical database system ability to mitigate the risk of compromising confidentiality. The measuring method selected for the Security business attribute is a combination of the following methods: Distance-based Linkage Disclosure (DLD), Probabilistic Linkage Disclosure (PLD) and Interval Disclosure (ID) as defined in [10, 11]. The combination of these methods is constructed as follows: Security = 0.25 (DLD) + 0.25 (PLD) + 0.5 (ID). The measurement method described for the Disclosure Risk Assessment in [10] is proposed for Security business attributes among other defined and proposed metrics in literature because of its general approach and because a lot of security methods have been evaluated using that combination.

- Information Loss – a business attribute describing the difference between original data and data released to the legitimate user. Information Loss stands for quality and usefulness of released data. Measuring the method proposed for the SABSA methodology is an application of Information Loss Measures from [10, 11]. The method is based on averaging the mean variations of differences between original and released matrices representing original and released datasets respectively, as well as of differences of their covariance matrices, diagonals of covariance matrices, vectors of variable averages. In the method the mean absolute error of the difference between correlation matrices of original and released matrices is taken into account.

- Cost – a business attribute that reflects cost of an initial implementation and cost of overhead processing per query. The measuring method measures an extra effort consumed on the initial implementation of the protection mechanism expressed in money and a difference in time consumed on processing a given set of queries on the database with the protection mechanism and without. Importance of any of these two factors is organizational specific.

- Bias – a business attribute describing how much actual value of a statistic differs from a perturbed value returned by the statistical database. The measurement method...
proposed is to empirically create a set of the most popular statistic queries (based on historical logs) and use that set for checking the variance between original and perturbed statistics.

- Consistency – a business attribute that represents the lack of:
  - Contradictions – i.e. repetitions of the same query yield different results, or when an additive statistic corresponding to a row or column sum differs from that obtained by adding the statistics in the row or column of the table.
  - Paradoxes – negative values are returned for counts. Proposed measuring method is based on knowledge about a specific protection mechanism. If the specific mechanism makes possible creation of contradictions or paradoxes set value 1 for the method, otherwise 0.

Another key element of a conceptual layer is to develop the control objectives, from which one can reduce the risk (from the business risk model in the contextual layer) to an acceptable level (defined also in the contextual layer). For each risk from the model, control objectives are defined. Typically, the control objectives are selected from the standards such as COBIT, ISO 17799, NIST 800-53. In that case, the objectives are so unique that we need to define it according to a business risk model. From the list of control objectives the set of control objectives that must be met to secure statistical databases in accordance with business requirements will be created. In addition, it is the basis for a security audit (as part of the operational layer).

Logical Layer

Logical layer contains a comprehensive description of the organization’s functional requirements to security resources. The logic of security is defined independently on the mechanisms that implement it as defined in the physical layer.

The main product of this layer is to define the security services – a set of activities for security. Compliance with security services will ensure security in the organization at the required level (arising from the business requirements). In our case, the service is to protect statistical databases, taking into account the compliance with control objectives and implementation of business attributes.

This layer describes in details the problem of the protection of statistical databases – how to create a slice of the security architecture for a statistical database. In addition, the link between the upper layer products with the proper implementation of the security method is created.

Physical Layer

The physical layer covers the selection of a specific security method to be used in order to meet the requirements of the organization. The main products of the physical layer are security mechanisms – different ways of realization of security services defined in the logical layer.

There are many mechanisms for the privacy protection of statistical databases (while maintaining statistical accuracy of derived statistics), which can be divided into four main groups:

- Query set reduction
- Data perturbation
- Result based perturbation
- Others.

In order to design a security method that addresses the security service as defined in the logical layer, analysis of hybrids of security mechanisms that ensure that the control objectives are met, must be performed. The next step is to evaluate hybrids in accordance to the business attributes profile.

The challenge is to ensure a high level of data security, while maintaining satisfactory accuracy of the statistics calculated. For this purpose it is necessary to analyze various mechanisms and a combination of selected mechanisms into the security methods for achieving this goal.
It is assumed in the model that weights of all business attributes are equal. The weight of each business attribute has to be set depending on organization business requirements. Security and Information Loss are attributes that are opposite to each other so their importance are supposed to be equally weighted. Cost attribute is the third most important business attribute and its weight fully depends on organization resources.

Component Layer

The component layer covers the selection of the implementation technique of the security method, which will be used to secure data. Implementation of the security method that has been selected in the physical layer takes place in the component layer. Implementation can take place in various ways. The selected implementation technique of the security method must work together to ensure consistency and completeness of the security infrastructure.

The product of this layer is consistent with the requirements of business organizations and will be used for statistical data provision.

Operational Layer

Operational layer describes the processes and procedures through which information systems of organizations operate in a safe manner. The operational layer is defined
in parallel with the logical, physical and component layers.

The first element of the operational layer is to define a process of monitoring changes to the business requirements and business risk model. On the basis of these changes, the security architecture is updated, and thus, also the implementation technique of the security method developed for statistical databases must be updated.

The second important operational element is to carry out periodic safety audits of statistical databases, in accordance with the control objectives defined in the conceptual layer. In case of discrepancies specific actions should be introduced that would allow to meet the control objectives.

3. Summary

The aim of this paper was to propose integration of existing mechanisms securing statistical databases with the security architecture of the organization based on SABSA methodology.

A coherent and comprehensive approach is ensured this way. It allows the integration of the security method developed for statistical databases with existing security mechanisms and services of other architectural elements of the organization. In addition, it allows for a security method development that complies with business requirements – sometimes ensuring the maximum value of each attribute is not consistent with the requirements of an organization.

Defining the lower SABSA layers allows increasing details of the safety aspect. On the other hand, implemented security mechanisms can be linked to business justifications by moving towards top layers.

Future work will be focused on finding appropriate methods for securing statistical databases. Verification and analysis of proposed techniques is not only based on properties of these methods, but also their applications to specific databases. The usability of methods highly depends on database structure and characteristics of the data it contains. The challenge is to find methods of ensuring a balance between safety and quality of data for a statistical database, taking into account database applications and the associated risks.

4. Bibliography

Bezpieczeństwo statystycznych baz danych jako element architektury bezpieczeństwa

Ł. ŚLEŻAK, J. BUTANOWICZ

Ochrona statystycznych baz danych jest problem najwyższej wagi w wielu organizacjach, w szczególności jednostkach administracji publicznej. Odpowiednia metoda ochrony statystycznych baz danych musi zapewnić równowagę pomiędzy bezpieczeństwem oraz jakością danych, biorąc pod uwagę ryzyka i wymagania biznesowe. Najlepszym sposobem osiągnięcia tego celu jest integracja mechanizmów ochrony statystycznych baz danych z architekturą bezpieczeństwa przedsiębiorstwa (np. z wykorzystaniem metodologii SABSA). W ten sposób zapewnione jest spójne i kompleksowe podejście. Pozwala to na wdrożenie mechanizmów ochrony statystycznych baz danych zgodnych z wymaganiami biznesowymi i zapewnienie, że te mechanizmy są zintegrowane z istniejącymi usługami oraz rozwiązaniami z zakresu bezpieczeństwa.

Słowa kluczowe: architektura bezpieczeństwa, SABSA.