АНОТАЦІЯ

Вілігура В. В. Моделі і методи забезпечення безпеки баз даних з універсальним базисом відношень. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 125 Кібербезпека та захист інформації (Галузь знань 12 Інформаційні технології). – Харківський національний університет імені В. Н. Каразіна Міністерства освіти і науки України, Харків, 2024.

Дисертація присвячена розробці, удосконаленню та використанню моделей та методів забезпечення безпеки баз даних.

Метою дисертаційної роботи є підвищення ефективності захисту баз даних, побудованих на основі схеми з універсальним базисом відношень, шляхом розробки та застосування моделей, методів та засобів забезпечення їхньої безпеки.

У першому розділі дисертації (*Сучасний стан, проблеми та завдання забезпечення безпеки баз даних*) виконано аналіз сучасного стану, основних проблем забезпечення безпеки баз даних (БД) та постановку задач дослідження. Зокрема, було проведено аналіз підходів та досягнень у галузі забезпечення та оцінки безпеки інформаційних систем загалом та баз даних, як їх основного функціонального компонента, зокрема, в тому числі проведено аналіз формальних моделей управління доступом та забезпечення цілісності даних, як методологічної основи побудови систем захисту та оцінки їхньої безпеки. За результатами аналізу виявлено недоліки та невирішені питання, що стосуються безпеки баз даних та її оцінки, виходячи з яких сформульовано задачі дисертаційного дослідження.

У другому розділі (*Розробка моделі захисту та методу оцінки безпеки бази даних*) вирішено завдання розробки та обґрунтування моделі захисту бази даних на основі системи безпеки з повним перекриттям та методу оцінки безпеки бази даних. Завдяки розширенню моделі Клементса–Гофмана (Clements–

Hoffman model) за рахунок включення безлічі вразливостей об'єктів (що дозволяє більш адекватно оцінювати ймовірність небажаного інциденту (реалізації загрози) у двофакторній моделі), певному інтегральному показнику захищеності БД (як величини зворотної сумарному залишковому ризику, складові компоненти якої представляються у вигляді відповідних лінгвістичних змінних), розробленому методу оцінювання основних компонент бар'єрів безпеки та захищеності бази даних в цілому, що спирається на теорію нечітких множин та ризику, стає можливою кількісна оцінка безпеки бази даних, що аналізується. Отримано *перший та другий наукові результати – удосконалено*:

– модель системи безпеки з повним перекриттям Клементса–Гофмана, яка відрізняється від відомої розширеною, за рахунок доповнення моделі множиною вразливостей об'єктів, як окремо об'єктивно існуючої категорії, та конкретизованим для баз даних складом компонент, що дозволяє більш адекватно оцінювати ймовірність небажаного інциденту (реалізації загрози) та захищеність бази даних у цілому;

– метод оцінювання основних компонент бар'єрів безпеки та захищеності бази даних в цілому, який на відміну від відомих, за рахунок комплексування вдосконаленої моделі Клементса–Гофмана, введеного інтегрального показника безпеки, положень теорії нечітких множин та ризику, дозволяє адаптуватися до нових умов функціонування та прозоро, комплексно та кількісно оцінювати безпеку баз даних з різними моделями даних.

У третьому розділі (*Розробка методів маскування даних*) вирішено завдання розробки та обґрунтування методів маскування даних, що дозволяють зменшити ймовірність реалізації загрози логічного висновку та забезпечити більш ефективне приховування коду критично важливих модулів, що постійно зберігаються, яке вимагає значно більших обчислювальних і часових витрат на його розкриття зловмисником, ніж при використанні існуючих способів, що надаються розробниками деяких сучасних систем керування базами даних (СКБД). Отримано *третій, четвертий та п'ятий наукові результати*:

удосконалено метод маскування MOBAT, що відрізняється від відомого, можливістю обфускації (англ. obfuscation) даних, шляхом математичних

перетворень на основі обчислення операцій за модулем, що застосовуються до елементів даних не тільки числового, а й широко поширеного в базах даних рядкового типу, що дозволяє суттєво розширити охоплення різноманітних маскованих з метою утруднення реалізації зловмисником загрози умовиводу даних БД;

отримали подальший розвиток:

– метод маскування елементів даних не ключових полів кортежів таблиць виробничої бази даних, що відрізняється від відомих оригінальним підходом до процесу перемішування з можливістю випадкової заміни елементів даних різного типу всередині заданого поля рядка та використання технології динамічного маскування, що дозволяє при менших обчислювальних витратах на перетворення і без зміни формату вихідних даних забезпечити ефективне приховування даних, яке ускладнює реалізацію загрози логічного висновку;

– метод приховування коду збережених у базі даних програм, який на відміну від відомих, дозволяє за рахунок випадкової перестановки (що спирається на сучасний варіант алгоритму тасування Фішера-Йейтса) символів коду з можливою заміною кожного з них на інший випадково вибраний із стандарту Unicode забезпечити більше ефективний (якій вимагає значно більших обчислювальних витрат) захист коду від його розкриття зловмисником, при цьому гарантуючи цілісність коду.

Отримано *перший практичний результат*: розроблений метод маскування елементів даних не ключових полів кортежів таблиць виробничої бази даних, орієнтований на заплутування, псевдонімізацію (англ. pseudonymisation) даних та ускладнення реалізації загрози логічного висновку, дозволяє зменшити час на відповідні операції перетворення на (10-17) % щодо методу класичного шифрування, при цьому не наводячи до зміни формату та збільшення розмірності даних, що зберігаються. Даний метод може бути також використаний у невиробничих базах даних, розширюючи можливості так званого статичного маскування даних.

У четвертому розділі (*Розробка методу контролю цілісності та справжності збережених програм, заснованого на можливостях технології*

блокчейн) вирішено завдання розробки та обґрунтування методу контролю цілісності та справжності модулів, що постійно зберігаються, заснованого на можливостях технології блокчейн. Отримано *шостий науковий результат*: *вперше* запропоновано метод моніторингу, що ґрунтується на можливостях технології блокчейн, який на відміну від відомих дозволяє за рахунок використання створеної зумовленої структури, правил формування первинного та наступних блоків у блокчейновому ланцюжку, організації зберігання цієї структури в рамках реляційної моделі даних, способів обчислення кореня геш-дерева, суворо контролювати набір програм БД, їх цілісність, справжність при менших обсягах збережених для цього даних і необхідних ресурсів процесора.

Отримано *другий практичний результат*: запропонований метод моніторингу модулів БД, що постійно зберігаються, вимагає менших обсягів збережених для цього даних і ресурсів процесора, ніж відомий метод контрольних сум, який для підтримки аналогічного контролю цілісності та справжності PSM вимагає виконання процедур гешування та цифрового підпису із збереженням відповідних даних для кожного конкретного PSM у конкретній схемі БД, причому однаково, не забезпечуючи контроль всього набору PSM загалом.

У п’ятому розділі (*Реалізація методів і засобів захисту в базах даних, побудованих на основі схеми з універсальним базисом відношень*) вирішено завдання обґрунтування та систематизації реалізованих заходів захисту, що забезпечують конфіденційність, цілісність даних та постійно збережених модулів баз даних з універсальним базисом відношень. Ці заходи ґрунтуються як на загальних формальних моделях управління доступом, забезпечення цілісності даних, методах, засобах, механізмах, що підтримуються СКБД, на платформі якої запропонована схема реалізується, так і на власних, розроблених у рамках створення інваріантної до предметних областей схеми БД. Отримано *третій практичний результат*: розроблені в процесі створення схеми БД з УББВ спеціальні заходи у вигляді відповідних методів, реалізованих об'єктів схеми (тригерів, процедур, пакетів, таблиць, функцій) та правил їх використання підвищують безпеку таких БД, забезпечуючи високий ступінь контрольованості

доступу до даних (аж до конкретного елемента), необхідну конфіденційність, цілісність даних та об'єктів схеми БД, на відміну від традиційних реляційних БД, що не володіють подібними заходами та функціональністю.

У шостому розділі (*Оцінка безпеки бази даних з універсальним базисом відношень*) здійснено оцінку безпеки бази даних з універсальним базисом відношень та наведено порівняльний аналіз захищеності баз даних, побудованих за традиційною технологією та на основі універсального базису відношень. Порівняльний аналіз показав, що використання запропонованих у роботі рішень дозволить підвищити ефективність / результативність захисту баз даних, побудованих на основі схеми з універсальним базисом відношень, більш ніж у 1.5 рази щодо традиційних реляційних БД.

Ключові слова: інформаційна система, база даних, база даних з універсальним базисом відношень, об'єкт бази даних, безпека, кібербезпека, модель безпеки, модель системи захисту з повним перекриттям, ризик, генератор псевдовипадкових чисел, маскування даних, шифрування, криптографічний захист інформації, цілісність, блокчейн.

ABSTRACT

Vilihura, V. V. Models and methods for ensuring the security of databases with the universal basis of relations. – Qualifying scientific work as a manuscript.

Thesis for the degree of Doctor of Philosophy in specialty 125 Cybersecurity and information protection (Field of knowledge 12 Information Technology). – V. N. Karazin Kharkiv National University of the Ministry of Education and Science of Ukraine, Kharkiv, 2024.

The thesis is devoted to the development, improvement and use of models and methods for ensuring database security.

The purpose of the thesis is to increase the efficiency of protection of databases built based on the scheme with the universal basis of relations, through the development and application of models, methods and means of ensuring their security.

In the first section of the thesis (*Current State, problems and tasks of database security*) the analysis of the current state, the main problems of ensuring the security of databases (DB) and the formulation of research tasks are carried out. In particular, an analysis of approaches and achievements in the field of ensuring and assessing the security of information systems in general and databases, as their main functional component, was carried out, in particular, an analysis of formal models of access control and ensuring the integrity of data as a methodological basis for building protection systems and assessing their security. Based on the results of the analysis, disadvantages and unresolved issues related to the security of databases and its assessment have been identified, on the basis of which the tasks of the thesis research have been formulated.

In the second section (*Development of the protection model and the method for assessing database security*), the problem of developing and justifying the database security model based on a full overlap security system and the database security assessment method is solved. Due to the extension of the Clements–Hoffman model due to the inclusion of a set of vulnerabilities of objects (which allows for a more adequate assessment of the probability of an undesirable incident (threat realization) in a two-factor model), a certain integral indicator of database security (as the inverse of

the total residual risk, the constituent components of which are represented in the form of corresponding linguistic variables), the developed method for assessing the main components of security barriers and security of the database as a whole, based on the theory of fuzzy sets and risk, it becomes possible to quantify the security of the analyzed database. The *first and second scientific results were obtained – improved*:

– the model of the security system with a complete overlap of Clements-Hoffmann, which differs from the well-known extended one, due to the addition of the model with a set of vulnerabilities of objects, as a separate objectively existing category, and a specified composition of components for databases, which allows for a more adequate assessment of the probability of an unwanted incident (threat implementation) and the security of the database as a whole;

– the method for assessing the main components of security barriers and security of the database as a whole, which, unlike the known ones, due to the integration of the improved Clements-Hoffmann model, the introduced integral security indicator, the provisions of the theory of fuzzy sets and risk, allows adapting to new operating conditions and transparently, comprehensively and quantitatively assessing the security of databases with different data models.

The third section (*Development of data masking methods*) solves the problem of developing and justifying data masking methods that reduce the probability of the threat of logical inference and ensure more effective hiding of the code of critically stored modules, which requires much more computational and time costs for its disclosure by an attacker than when using the existing methods provided by the developers of some modern database management systems (DBMS). The *third, fourth and fifth scientific results were obtained*:

the method of masking MOBAT, which differs from the known one, in the possibility of obfuscation of data, *has been improved* by means of mathematical transformations based on the calculation of operations by modulus, applied to data elements not only numeric, but also widespread in databases of string type, which allows to significantly expand the coverage of various masked in order to complicate the threat of inference of database data by an attacker;

have been further developed:

– the method of masking data elements of non-key fields of tuples of tables of the production database, which differs from the known ones in the original approach to the shuffle process with the possibility of random replacement of data elements of different types within a given field of the row and the use of dynamic masking technology, which allows with lower computing costs for transformation and without changing the format of the original data to ensure effective data hiding, which makes difficult to implement the threat of a logical conclusion;

– the method of hiding the code of programs stored in the database, which, unlike the known ones, allows, due to the random rearrangement (based on the modern version of the Fisher-Yates shuffle algorithm) of the code symbols with the possible replacement of each of them with another randomly selected from the Unicode standard, to provide more effective (which requires much higher computing costs) protection of the code from its disclosure by an attacker, while guaranteeing the integrity of the code.

The first practical result was obtained: the developed method of masking data elements of non-key fields of tuples of tables of the production database, focused on obfuscation, pseudonymisation of data and complication of the implementation of the threat of logical inference, allows to reduce the time for the corresponding conversion operations by (10-17)% regarding the method of classical encryption, while not leading to a change in the format and an increase in the dimension of the stored data. This method can also be used in non-production databases, expanding the possibilities of so-called static data masking.

In the fourth section (*Development of the method for controlling the integrity and authenticity of stored programs based on the capabilities of blockchain technology*), the problem of developing and substantiating a method for controlling the integrity and authenticity of permanently stored modules, based on the capabilities of blockchain technology, is solved. *The sixth scientific result was obtained: for the first time* a method of monitoring based on the capabilities of blockchain technology is proposed, which, unlike the known ones, allows, through the use of the created predetermined structure, rules for the formation of the primary and subsequent blocks in the blockchain chain, the organization of storage of this structure within the

framework of a relational data model, methods of calculating the root of the hash tree, to strictly control the set of database programs, their integrity, authenticity with smaller volumes of data stored for this and the necessary resources processor.

The *second practical result is obtained*: the proposed method of monitoring permanently stored database modules requires smaller amounts of data and processor resources stored for this purpose than the well-known method of checksums, which, in order to maintain similar control over the integrity and authenticity of PSM, requires the execution of hashing and digital signature procedures with the preservation of the corresponding data for each specific PSM in a specific database schema, and in the same way, without providing control of the entire PSM set as a whole.

In the fifth section (*Implementation of methods and means of protection in databases built based on the scheme with the universal basis of relations*) the problem of substantiation and systematization of the implemented protection measures that ensure confidentiality, data integrity and permanently stored database modules with the universal basis of relations is solved. These measures are based on both general formal models of access control, data integrity, methods, means, mechanisms supported by the DBMS, on the platform of which the proposed scheme is being implemented, and on our own, developed within the framework of creating a database scheme invariant to subject areas. The *third practical result* was obtained: special measures developed in the process of creating a database schema with UBV in the form of appropriate methods, implemented schema objects (triggers, procedures, packages, tables, functions) and rules for their use increase the security of such databases, providing a high degree of control of access to data (up to a specific element), the necessary confidentiality, integrity of data and objects of the database scheme, in contrast to traditional relational databases, that do not have such measures and functionality.

In the sixth section (*Assessment of the security of database with the universal basis of relations)* an assessment of the security of a database with the universal basis of relations is carried out and a comparative analysis of the security of databases built on the traditional technology and on the basis of the universal basis of relations is provided. A comparative analysis has shown that the use of the solutions proposed in

the work will increase the efficiency / effectiveness of the protection of databases built on the basis of the scheme with the universal basis of relations, more than 1.5 times relative to traditional relational databases.

Keywords: information system, database, database with the universal basis of relations, database object, security, cyber security, security model, full overlap security system model, risk, pseudorandom number generator, data masking, encryption, information cryptographic protection, integrity, blockchain.