Condensed Matter Physics 2008, Vol. 11, No 4(56), pp. 761–765

Information and data protection within a RDBMS

Y.Khmelevsky

Computer Science Department, Okanagan College, 1000 KLO Road, Kelowna, BC, V1Y 4X8, Canada

Received January 31, 2008

Security issues for some special large data, such as binary and image files, as well as video and audio files and streams still require a special development, especially for the industrial database systems (Oracle, MS SQL, DB2, etc). New encryption methods should be used additionally to traditional encryption methods and other protection solutions, such as authentication, authorization, access control, security monitoring and audit. The purpose of this article is to present the research results regarding information security and data protection, as well as some practical aspects of the encryption by CrypTIM algorithm, developed by Prof. V. Ustimenko in the last decade [Ustimenko V., Lecture Notes In Computer Science, 2001, **278**, 2227]. This text additionally proposes a practical utilization of the Model Driven system design for large objects (LOB) encryptions within a database, used to store some special large binary files, such as images, sound files, movies, special binary files in order to improve maintenance and data protection. Novel problems and trends in providing security against criminal activities in the current Cyberspace are analyzed.

Key words: *private key cryptography, stream ciphers, graph based encryption*

PACS: *02.10.Hh, 02.10.Ox*

1. Introduction

1.1. Security problems and data protection

For the last decade we can see a strong interest to the data protection in corporate information systems. And this interest grows, especially following each new scandal dealing with financial or corporate data leak. Let us consider just a few cases:

1. 160 laptop computers were lost or stolen in less than four years from FBI, including at least 10 that contained sensitive or classified information [2].

2. A backup computer file containing information on almost half a million of Talvest Mutual Funds clients has gone missing. The missing data was in a file that disappeared "while in transit between the offices". The file had personal and financial details on current and former clients of Talvest Mutual Funds, which is a CIBC bank subsidiary in Canada. The information may have included client names, addresses, signatures, dates of birth, bank account numbers, beneficiary information and/or Social Insurance Numbers [3].

3. The data security breach happened on the network of CardSystems Solutions. The program captured credit card data. The CardSystems breach follows several high-profile data loss incidents that potentially exposed American consumers to identity theft, including the loss of CitiFinancial tapes containing unencrypted information on 3.9 million customers. More than 40 million credit card accounts were exposed by the breach. About 22 million of those are Visa cards and 13.9 million are MasterCards. The remaining accounts were linked to other brands, including American Express and Discover [4].

We can continue the list of security leaks, breaches, scandals etc. for many pages. Almost every week we have new reports about similar security problems. Every year "unbreakable information systems" are broken and almost every information system and any technology have security vulnerabilities, connected with potential information leaks. Moreover, the organized crime in the

Cyberspace is winning the Internet security war. Specialists warned at the world's foremost gathering of computer hackers in Las Vegas in August 2006: "The online peril is no longer brilliant young social outcasts penetrating networks for notoriety; it is international crime rings swiping billions of US dollars with keystrokes and malicious computer codes", cyber cops agreed [5]. That is a real situation with information system security and data protection.

1.2. Information security engineering and data protection integration into a model-driven software development process

One of the goals of this article is to investigate the information security engineering and data protection integration into a model-driven software development process, reducing the gap between security analysis and the integration of reliable information system security and data protection mechanisms, security models and systems design models integration. This text is an attempt to help incorporate information security and data protection topics into database management, SW engineering and other Computer Science courses taught at colleges and universities, and introduce the information security and data protection methods at the modelling level within SW engineering and design.

2. Information systems security and data protection model driven approach

Possible solutions for information security and data protection in GIS as well as in Database Management Systems for large binary objects on the Oracle 9i DBMS were investigated in [6–9] and [10]. The investigation of possible solutions of data protection for the GIS systems was performed at different levels, including data encryption by traditional and new algorithms within DBMS. Our results were presented at international conferences [6–8] and [9], in journal publications [10] as well as in a MSc research work at the University of the South Pacific, Fiji.

A lot of interesting papers related to data protection and information security have been published in the recent few years. Some of them suggest new approaches to the improvement of information system and data protection. Highly complicated security of the n-tier multi-component systems can be drastically improved by implementing an appropriate encryption technology at a low data protection level and by implementing a Model Driven Security approach to building secure systems by specifying system models along with their security requirements and using the tools in order to automatically generate system architectures from the models, including complete, configured access control infrastructures [11]. Paper [11] proposed a general scheme for constructing languages that combines the languages for system modelling with languages for security modelling. The models in the combined languages permits to automatically generate access control infrastructures for server-based applications, built from declarative and programmatic access control mechanisms. A UML-based approach is presented in [12], which could be utilized for the n-tier architecture and some special data storage systems solution [9]. The feasible architecture included a client tier in which user services resided, where the client tier was represented by a Web browser or wireless device (thin client), and either Web browser with Java applets or ActiveX components or a Java application (thick client) [13,14]. The middle tier was divided into two or more subsystems (or layers) with different functions and security features, including SSL encryption, authentication, user's validation, single-sign logon server, and digital signature. Web services perform specific application functions, and special queries, and can be integrated as a part of the middle-tier application server [15].

In order to transform the already discussed deployment architecture [9] into UML deployment models, we should use the developed UML approach for Access Control Infrastructures analyses in [11] and UML models for information systems performance analyses in [16]. The primary model represents the basic functionality without any security mechanisms. The components of a web-system are logically divided into three tiers: a) a set of emulated web browsers (EB), b) web tier including Web Servers, Image Servers and Web Caches and c) persistent storage [16]. Figure 1 presents a general UML deployment diagram of the feasible Information System, investigated in the [9]. The UML model has a transition from 3-tier architecture to 4- and n-tier and

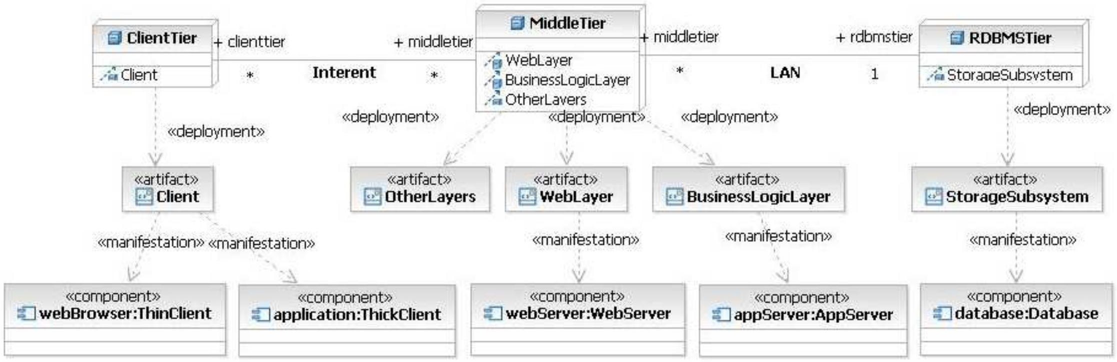

Figure 1. A general UML deployment diagram of the feasible Information System.

encryption implementation for all data transmission, as an important part of security solution. The UML diagrams implementation in our research gives a valuable tool for the security performance investigation and appropriate security implementation by a Model Driven Security [11] implementation, where designers specify the system models along with their security requirements and use tools in order to automatically generate system architectures from the models, including complete, configured access control infrastructures. From the models, the access control infrastructures for server-based applications can be automatically generated, built from declarative and programmatic access control mechanisms [11].

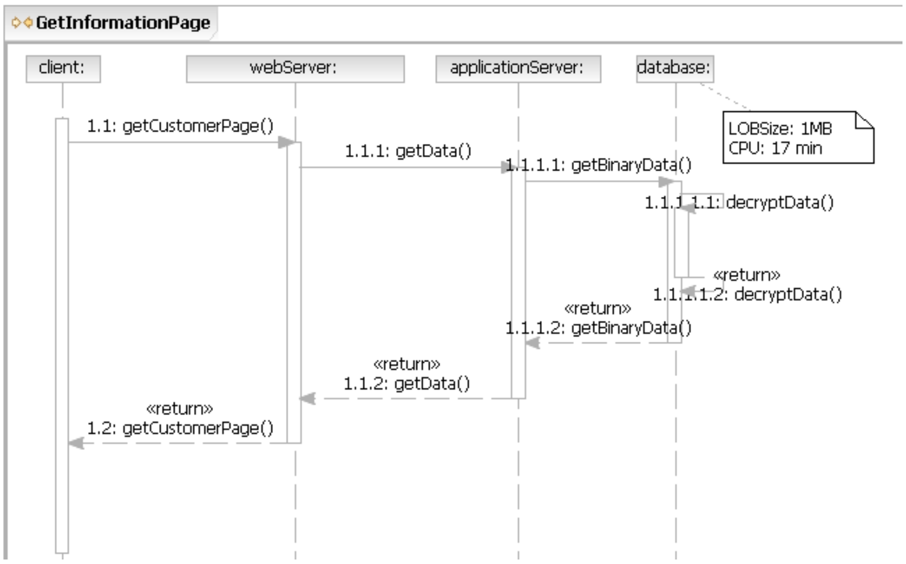

Figure 2. An example of scenario GetInformationPage with extra encryption implementation and time delays.

The next UML diagram in figure 2 shows an example of scenario GetInformationPage with extra encryption implementation and time delays for the encryption/decryption of the information, compare with [16]. In accordance with our research, the encryption/decryption process can be the most consumable part of a system performance and can create very high workload for the data processing. Encryption should be used within n-tier information systems architecture for different purposes of protection [6,9]. The first purpose of such protection is encryption of a user's identity for authentication and authorization services. For a typical case at present, this relies on the transport layer for security via the SSL protocol, which also provides data integrity and strong authentication of both clients and servers. SSL provides message-level security and non-repudiation. The second, encryption can be used for the protection of large binary data, such as images, video and audio, in transit. There are a few solutions for protecting Web-based data in transit over the Internet. These solutions include virtual private networks (VPNs) for secure point-to-point connectivity or SSL layer. The third, cryptography can be used to encrypt sensitive data stored on a data storage system [6,9], including caches. Special cryptography software on a server can be applied for these purposes. Security on a thick client can be implemented by using the categories of protection similar to thin client and the elements of security infrastructure.

A RDBMS is a core system in any organization or should be a core system, and as was discussed in [9] has got a powerful mechanism for storing different types of information with different access rights and sophisticated security solutions. Every year new products emerge on the market, which raise the possibilities to utilize legacy RDBMS for a critical data storage. The data encryption can significantly degrade the system performance and application response time. In our previous publications [6,9,10] we investigated the Oracle 9i DBMS_OBFUSCATION.TOOLKIT with a comparison to new CryptTIM algorithm [10]. The performance for large data was relatively slow, because we used loops in the encryption/decryption implementation. The new CrypTIM version implementation uses a linear approach and is supposed to give a much faster performance, acceptable for implementation in industry.

2.1. Large oracle object encryption

Special approaches were developed using encryption for large files in Oracle. To encrypt LOB data objects, the procedure splits the data into smaller binary chunks, then encrypts and appends them to the LOB object back. Once the encrypted binary data have been allocated into LOB segments, they were decrypted by chunks and written back to BLOB object. For the read-only binary data, additional LOB object once encrypted should always be kept. This saved time for encryption procedure during log off. The decrypted binary data were simply replaced by read-only encrypted data in the main permanent storage during log off.

As was already discussed in [10], the encryption/decryption time has linear correlation to the file size. Roughly it takes about 60 seconds for 51 KB file encryption within 16 byte length password by using PL/SQL functions, and for 1MB – about 17 minutes. By utilizing two core multi-processor workstations and Java programming language for the encryption/decryption functions, encryption time was further decreased many times. This permits to use the CrypTIM algorithm for practical implementation in industries.

3. Conclusion

N-tier architectures and Web Services are making the application layer more complex and fragmented, but Information Technology security and data protection problems are a real challenge to the current system development. The one solution to a better security and protection lies in applying the Model Driven Security approach as well as UML to building secure systems by specifying system models along with their security requirements and using the tools to generate system architectures from the models, including complete, configured access control infrastructures, as well as the security framework to all subsystems and components of n-tier system. On the other hand, to protect the data within an information system, a reliable encryption, such as CrypTIM, RSA, etc. has to be implemented on all information system layers and within a DBMS.

Furthermore a disadvantage of any security and data protection system is its low performance, which has to be investigated and monitored at all the levels of a system design and development.

Acknowledgements

This text would not be possible if Prof. Ustimenko had not developed his unique CrypTIM encryption algorithm. Hence, the author would like to thank Vasyl Ustimenko for his new algorithm, patience and supervision of the research conducted.

Thanks to the University of the South Pacific in Fiji for the research support and the Okanagan College for the time to write this paper.

References

1. Ustimenko V., Lecture Notes In Computer Science, 2001, **278**, 2227.
2. Eggen D., Washington Post, 2007, A06.
3. CBC News. http://www.cbc.ca/money/story/2007/01/18/cibc.html, (2007).
4. Evers J., Silicon.com. http://software.silicon.com/security/0,39024655, 39131314,00.htm, (2005).
5. Theage.com.au, http://www.theage.com.au/articles/2006/08/06/ 1154802739105.html, (2006).
6. Govorov M., Khmelevsky Y., Ustimenko V., Khorev A. – In: 11th International Symposium on Spatial Data Handling Fisher, Peter F. (Ed.), Springer-Verlag, 2005, p. 71.
7. Govorov M., Khmelevsky Y., Sharma P., InterCarto/InterGIS, 2004, **10**, 476.
8. Sharma P., Govorov M., Khmelevsky Y., Dhanjal S., Human Perspectives in the Internet Society, 2004, **489**.
9. Govorov M., Khmelevsky Y., Ustimenko V., Khorev A. – In: Proc. of the 21st International Cartographic Conference (ICC), 2003, p. 1784.
10. Ustimenko V.A., Khmelevsky Y.M., The South Pacific Journal of Natural Science, 2002, **34**, 20.
11. Basin D., Doser J., Lodderstedt T., ACM Trans. Softw. Eng. Methodol., 2006, **39**, 15.
12. Jurjens J. Springer-Verlag, 14, (2004) (explanation from author!)
13. ISO 19125 (DIS), ISO/TC 211, N1002 (2000). (explanation from author!)
14. OpenGIS Project Document, 00-028 (2000).(explanation from author!)
15. Web Services Handbook, IBM (2003).(explanation from author!)
16. Petriu D.C., Woodside C.M., Petriu D.B., Xu J., Israr T., Georg G., France R., Bieman J.M., Houmb S.H., Jurjens J., ACM Press. 91, (2007). (explanation from author!)

Інформаційна безпека та захист даних в рамках RDBMS

Ю.Хмелевський

Факультет комп'ютерних наук, Коледж Оканаган, Канада

Отримано 31 січня 2008 р.

Аспекти безпеки деяких спеціальних даних великого об'єму, таких, як бінарні, відео та аудіо файли і потоки, досі потребують уваги і спеціальних досліджень, що, зокрема, стосується індустріальних баз даних (Oracle, MS SQL, DB2 та інші). Нові методи кодування повинні додатково використовуватися разом з традиційними методами та іншими засобами безпеки, такими, як аутентифікація, авторизація, захист доступу, моніторинг і аудит рівня безпеки. Метою публікації є огляд досліджень з інформаційної безпеки та захисту даних, а також – практичних аспектів кодування за алгоритмом Cryp TIM, розробленим проф. Устименком за останнє десятиріччя. Ця публікація пропонує практичне використання модельно-рухомої системи, створеної для кодування великих об'єктів (LOB), що розміщені в базі даних для зберігання деяких спеціальних файлів, таких як зображення, аудіо файли, фільми, спеціальні бінарні файли, – для покращення можливостей користування та підвищення рівня безпеки даних. Розглянуто нові проблеми та напрямки в забезпеченні захисту від кримінальних дій в сучасному глобальному інформаційному просторі.

Ключові слова: *криптографія з приватним ключем, цифровий потік, кодування на основі графів*

PACS: *02.10.Hh, 02.10.Ox*