Providing database encryption as a scalable enterprise infrastructure service

U. T. Mattsson
Protegrity Corp.

Abstract

As databases become networked in more complex multi-tiered applications, their vulnerability to external attack grows. We address scalability as a particularly vital problem and propose alternative solutions for data encryption as an enterprise IT infrastructure component. In this paper we explore a new approach for data privacy and security in which a security administrator protecting privacy at the level of individual fields and records, and providing seamless mechanisms to create, store, and securely access databases. Such a model alleviates the need for organizations to purchase expensive hardware, deal with software modifications, and hire professionals for encryption key management development tasks. Although access control has been deployed as a security mechanism almost since the birth of large database systems, many still look at database security as a problem to be addressed as the need arises – this is often after threats to the secrecy and integrity of data have occurred. Instead of building walls around servers or hard drives, a protective layer of encryption is provided around specific sensitive data items or objects. This prevents outside attacks as well as infiltration from within the server itself. This also allows the security administrator to define which data stored in databases are sensitive and thereby focusing the protection only on the sensitive data, which in turn minimizes the delays or burdens on the system that may occur from other bulk encryption methods.
Keywords: isolation, intrusion tolerance, database security, encryption, privacy, VISA CISP, GLBA, HIPAA.

1 Introduction

Although access control has been deployed as a security mechanism almost since the birth of large database systems, for a long time security of a DB was

considered an additional problem to be addressed when the need arose, and after threats to the secrecy and integrity of data had occurred [23]. Now many major database companies are adopting the loose coupling approach and adding optional security support to their products. You can use the encryption features of your Database Management System (DBMS), or perform encryption and decryption outside the database. Each of these approaches has its advantages and disadvantages. Adding security support as an optional feature is not satisfactory, since it would always penalize system performance, and more importantly, it is likely to open new security holes. Database security is a wide research area [26, 23] and includes topics such as statistical database security [21], intrusion detection [34], and most recently privacy preserving data mining [22], and related papers in designing information systems that protect the privacy and ownership of individual information while not impeding the flow of information, include [22, 23, 24, 25].

2 Choosing the point of policy enforcement and data protection (PEPP)

Encryption is the perfect technique to solve this problem. Prior work [7] [2] does not address the critical issue of performance. But in this work, we have addressed and evaluated the most critical issue for the success of encryption in databases, performance. To achieve that, we have analysed different solution alternatives. There are two dimensions to encryption support in databases. One is the granularity of data to be encrypted or decrypted. The field, the row and the page, typically 4KB, are the alternatives. The field is the best choice, because it would minimize the number of bytes encrypted. However, as we have discovered, this will require methods of embedding encryption within relational databases or database servers. The second dimension is software versus hardware level implementation of encryption algorithms. Our results show that the choice makes significant impact on the performance. The loss of granular protection will impact the security level. This is discussed in more detail in [18]. Choosing the point of implementation not only dictates the work that needs to be done from an integration perspective but also significantly affects the overall security model. The sooner the encryption of data occurs, the more secure the environment—however, due to distributed business logic in application and database environments, it is not always practical to encrypt data as soon as it enters the network. Encryption performed by the DBMS can protect data at rest, but you must decide if you also require protection for data while it's moving between the applications and the database. How about while being processed in the application itself? Particularly if the application may cache the data for some period. Sending sensitive information over the Internet or within your corporate network as clear text, defeats the point of encrypting the text in the database to provide data privacy. Good security practice protects sensitive data in both cases – as it is transferred over the network (including internal networks) and at rest. Once the secure communication points are terminated, typically at the network perimeter, secure transports are seldom used within the enterprise. Consequently,

information that has been transmitted is in the clear and critical data is left unprotected. This is discussed in more detail in [18].

2.1 Data privacy

Database-level encryption allows enterprises to secure data as it is written to and read from a database. This type of deployment is typically done at the column level within a database table and, if coupled with database security and access controls, can prevent theft of critical data. Database-level encryption protects the data within the DBMS and also protects against a wide range of threats, including storage media theft, well known storage attacks, database-level attacks, and malicious DBAs. Storage-level encryption enables enterprises to encrypt data at the storage subsystem, either at the file level (NAS/DAS) or at the block level SAN. This type of encryption is well suited for encrypting files, directories, storage blocks, and tape media. In today's large storage environments, storage-level encryption addresses a requirement to secure data without using LUN (Logical Unit Number) masking or zoning. While this solution can segment workgroups and provides some security, it presents a couple of limitations. It only protects against a narrow range of threats, namely media theft and storage system attacks. However, storage-level encryption does not protect against most application- or database-level attacks, which tend to be the most prominent type of threats to sensitive data. Current storage security mechanisms only provide block-level encryption; they do not give the enterprise the ability to encrypt data within an application or database at the field level. Consequently, one can encrypt an entire database, but not specific information housed within the database.

2.2 Encryption scheme alternatives

We considered several possible combinations of different encryption approaches, namely; software and hardware level encryption, and different data granularity. We started with software encryption at field level. We then developed search acceleration support to index encrypted fields, and experienced a low performance overhead when searching on encrypted fields, including primary index fields. We also directed our experiments hardware level encryption only for master key encryption.

2.3 Basic software level encryption

Initially we considered several encryption algorithms AES, RSA [10] and b) Blowfish [11] for the implementation. We conducted experiments using these algorithms and found that the performance and security of the AES algorithm is better than the RSA implementation and the Blowfish algorithm implementation. AES is fast, compared to other well-known encryption algorithms such as DES [12]. Detailed description of the algorithm is given in [12].

2.4 Hardware level encryption

We studied the use of HSM FIPS-140-1 level 3 Hardware Security Modules with a mix of hardware and software keys. The master key was created and encrypted

/ decrypted on HSM. The master key is not exposed outside the HSM. The cost of encryption/decryption consists of start up cost, which involves function and hardware invocation, and encryption/decryption algorithm execution cost, which is depended on the size of the input data. This implies that the start up cost is paid every time a row is processed by encryption. We used specialized encryption hardware from different vendors, including IBM, Eracom, nCipher, and Chrysalis for this test. On of our test beds used the IBM S/390 Cryptographic Coprocessor, available under IBM OS/390 environment with Integrated Cryptographic Enterprise IT infrastructure component Facility (ICSF) libraries. IBM DB2 for OS/390 provides a facility called ”editproc” (or edit routine), which can be associated with a database table. An edit routine is invoked for a whole row of the database table, whenever the row is accessed by the DBMS. We registered an encryption/decryption edit routine for the tables. When a read/write request arrives for a row in one of these tables, the edit routine invokes encryption/decryption algorithm, which is implemented in hardware, for whole row. We used the DES [3] algorithm option for encryption hardware. The loss of granular column-level protection will impact the security level. This is discussed and evaluated earlier.

2.5 Encryption penalty

If we compare the response time for a query on unencrypted data with the response time for the same query over the same data, but with some or all of it encrypted, the response time over encrypted data will increase due to both the cost of decryption as well as routine and/or hardware invocations. This increase is referred to as the encryption penalty. An observation according to recent studies is that, different fields have different sensitivity [16]. It is possible for Hybrid to support encryption only on selected fields of selected tables. Encryption, by its nature, will slow down most SQL statements. If some care and discretion are used, the amount of extra overhead should be minimal. Also, encrypted data will have a significant impact on your database design. In general, you want to encrypt a few very sensitive data elements in a schema, like Social security numbers, credit card numbers, patient names, etc. Some data values are not very good candidates for encryption -- for example booleans (true and false), or other small sets like the integers 1 through 10. These values along with a column name may be easy to guess, so you want to decide whether encryption is really useful. Creating indexes on encrypted data is a good idea in some cases. Exact matches and joins of encrypted data will use the indexes you create. Since encrypted data is essentially binary data, range checking of encrypted data would require table scans. Range checking will require decrypting all the row values for a column, so it should be avoided if not tuned appropriately with an accelerated search index.

2.6 Query rewrite to improve encryption overhead

We implemented limited support for rewrite of a query, and experienced significant optimisation capabilities when searching on encrypted columns. A

method for common sub-expression elimination (CSE) needs to be applied to expensive user defined functions for a query. Common sub-expression detection and elimination are well known in compiler optimisation [1] [9].

3 Scalability of different encryption architectures

Each of these approaches has its advantages and disadvantages. Adding only central security and encryption support is not satisfactory, since it would always penalize system performance, and more importantly, it is likely to open new security holes. Database security is a wide research area [6, 3] and includes topics such as statistical database security [21], intrusion detection [19, 4], and most recently privacy preserving data mining [13], and related papers in designing information systems that protect the privacy and ownership of individual information while not impeding the flow of information, include [13, 14, 5, 8].

3.1 Performance considerations

We studied the industry standard SQL benchmark [15] as a model for workloads. Some simple sample tests on Oracle and DB2. The first benchmark was focus on a particular customer scenario. Subsequent benchmarks used a workload combined from multiple customer case studies. The technological aspects of developing database privacy as an enterprise IT infrastructure component lead to new research challenges. First and fore-most is the issue of encryption key management. Most corporations view their data as a very valuable asset. The key management system would need to provide sufficient security measures to guard the distributed use of encryption keys. We propose a combined hardware and software based data encryption system as the solution to this problem. A distributed policy and audit capability is proposed for the control the use of different encryption keys. Detailed investigation of this solution is presented below. Since the interaction between the database and the enterprise IT infrastructure component there are potential over-heads introduced by encryption. Therefore the sources of performance degradation and its significance should be determined.

3.2 Network Attached Encryption

The Network Attached Encryption is implemented as a Network Attached Encryption Appliance that scales with the number of Network Attached Encryption Appliances available. The benchmarks showed a throughput of between 440 and 1,100 row-decryptions per second. The benchmarks showed a linear scalability of this topology when adding additional database servers. A system with twelve database servers performed at 4,200 row-decryptions per second with five Network Attached Encryption Appliances. In prior work with IBM Research [46] we addressed some critical performance issues when using HSM support. A coming paper will address how to avoid the problems of

performance and scalability when using HSM support, and also how to prevent API level attacks when using HSM support, including Network Attached Encryption Appliances.

3.3 The Hybrid system

The Hybrid system is implemented as distributed processes that scale with the number of processors and database server available. The Hybrid solution can also utilize an optional HSM in a way that allows the total encryption system to scale with the number of processors available on the database servers. Our DB2 benchmarks at IBM showed a typical throughput of 187,000 row-decryptions per second, with 20 concurrent users. This translates to an ability to decrypt 187,000 database table rows per second. The test tables included 80 bytes of encrypted data per row. We saturated all six RS6000 processors at 100% utilization when we tested with 1,000 concurrent users. Some additional benchmarks with DB2, and Oracle showed a typical throughput in the range of 66,000 to 110,000 row-decryptions per second, on a two processor, 3 GHz system with 3 GB RAM, running a Windows operating system. The benchmarks also showed a linear scalability of this topology when adding additional database servers. A system with twelve database servers performed at 2,100,000 row-decryptions per second. Additional tuning by adding an accelerated search index for encrypted columns, reduced the response-time and the number of rows to decrypt, by a factor between 10 and 30 for some of the queries in our Oracle test. This can be viewed as enabling a 'virtual throughput' in the range of 660,000 to 1,100,000 'virtual row-decryptions' per second, when comparing to a solution that is not using an accelerated search index for encrypted columns. Some preliminary benchmarks with SQL Server showed a typical throughput in the range of 3,000 to 32,000 row-decryptions per second, depending on a optimised combination of column level encryption and table level encryption, and the amount of cached table data, The initial SQL Server 2000 test used a low-end test system running Windows with a 1.6 GHz processor, 1 GB Physical RAM, and 3 GB Virtual RAM. Additional details from the ongoing benchmarks will be discussed in a coming paper.

4 Policy management

Current commercial RDBMSs support many different kinds of identification and authentication methods, password-based authentication [32], host-based authentication [24, 32, 31], PKI (Public Key Infrastructure) based authentication [39], third party-based authentications such as Kerberos [37], DCE (Distributed Computing Environment [43]) and smart cards [40]. Essentially, all methods rely on a secret known only to the connecting user. It is vital that a user should have total control over her/his own secret. For example, only she/he should be able to change her/his password. Other people can change a user's password only if they are authorized to do so. In a DB system, a DBA can reset a user's password upon the user's request, probably because the user might have forgotten her/his

password. However the DBA can temporarily change a user's password without being detected and caught by the user, because the DBA has the capability to update (directly or indirectly) the system catalogs. This is discussed in more detail in [18].

5 Auditability

Technically, if we allow a DBA to control security without any restriction, the whole system becomes vulnerable because if the DBA is compromised, the security of the whole system is compromised, which would be a disaster. However if we have a mechanism in which each user could have control over their own secrecy, the security of the system is maintained even if some individuals do not manage their security properly. Access control is the major security mechanism deployed in all RDBMSs. It is based upon the concept of privilege. A subject (i.e., a user, an application, etc.) can access a database object if the subject has been assigned the corresponding privilege. Access control is the basis for many security features. Special views and stored procedures can be created to limit users' access to table contents. However, a DBA has all the system privileges. Because of their ultimate power, a DBA can manage the whole system and make it work in the most efficient way. However, they also have the capability to do the most damage to the system. With a separated security directory the security administrator sets the user permissions. Thus, for a commercial database, the security administrator (SA) operates through separate middle-ware, the Access Control System (ACS), used for authentication, verification, authorization, audit, encryption and decryption. The ACS is tightly coupled to the database management system (DBMS) of the database. The ACS controls access in real-time to the protected fields of the database. Such a security solution provides separation of the duties of a security administrator from a database administrator (DBA).

6 Encryption key management

One of the essential components of encryption that is often overlooked is key management - the way cryptographic keys are generated and managed throughout their life. Because cryptography is based on keys that encrypt and decrypt data, your database protection solution is only as good as the protection of your keys. Security depends on two factors: where the keys are stored and who has access to them. When evaluating a data privacy solution, it is essential to include the ability to securely generate and manage keys. This can often be achieved by centralizing all key management tasks on a single platform, and effectively automating administrative key management tasks, providing both operational efficiency and reduced management costs.

7 Conclusion

We addressed scalability as a particularly vital problem and propose alternative solutions for data encryption as an enterprise IT infrastructure component. In this

paper, we introduced the Hybrid, a database privacy solution built on top of all major relational databases. The Hybrid model introduces many significant challenges primary of which are the additional overhead of searching on encrypted data an infrastructure to guarantee data privacy, and management of such an enterprise IT infrastructure component. We have addressed these issues. Our experiments using several benchmarks showed that the overhead is tolerable when using a suitable encryption architecture. The Hybrid model implements a scalable approach for data privacy and security in which a security administrator protecting privacy at the level of individual fields and records, and providing seamless mechanisms to create, store, and securely access databases. Such a model alleviates the need for organizations to purchase expensive hardware, deal with software modifications, and hire professionals for encryption key management development tasks. We proposed, implemented, and evaluated different encryption schemes. We showed the drastic decrease in query execution times from distributed software level encryption. We believe, from our experience, database privacy as an infrastructure service is a viable model and has a good chance of emerging as a successful offering for most applications.

References

[1] A. Aho, S. Johnson, and J. Ullman. Code generation for expressions with sub-expressions. Journal of ACM, Jan., 1977.

[2] G. Davida, D. Wells, and J. Kam. A database encryption system with subkeys. ACM Transactions on Database Systems, 6(2), 1981.

[3] DES. Data encryption standard. FIPS PUB 46, Federal Information Processing Standards Publication, 1977.

[4] T. F. Lunt. A survey of intrusion detection techniques. Computer & Security, 12(4), 1993.

[5] Agrawal, J. Kiernan, R. Srikant and Y. Xu. Implementing P3P using database technology. In Proc. of the 19th Int'l Conference on Data Engineering, Bangalore, India, March 2003.

[6] G. Hamilton and R. Cattell. JDBC: A Java SQL API. http://splash.javasoft.com/jdbc/.

[7] J. He and M. Wang. Encryption in relational database management systems. In Proc. Fourteenth Annual IFIP WG 11.3 Working Conference on Database Security (DBSec'00), Schoorl, The Netherlands, 2000.

[8] R.Agrawal, J.Kiernan, R.Srikant and Y.Xu. AnXPath-based preference language for P3P. In Proc. of the 12th Int'l World Wide Web Conference, Budapest, Hungary, May 2003.

[9] S. Muchnick. Advanced Compiler Design and Implementation. Morgan Kaufmann Publishers, 1997.

[10] R. L. Rivest, A. Shamir, and L. M. Adleman. A method for obtaining digital signatures and public key cryptosystems. Communications of the ACM, 21(2):120–126, 1978.

[11] B. Schneier. Description of a new variable-length key, block cipher (blowfish), fast software encryption. In Cambridge Security Workshop Proceedings, pages 191–204, 1994.

[12] B. Schneier. Applied Cryptography. John Wiley & Sons, Inc., 1996.
[13] R. Agrawal and J. Kiernan. Watermarking relational databases. In 28th Int'l Conference on Very Large Databases, Hong Kong, China, August 2002.
[14] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. Hippocratic databases. In Proc. of the 28th Int'l Conference on Very Large Databases, Hong Kong, China, August 2002.
[15] SQL. Benchmark Specification. http://www.tpc.org.
[16] A. F. Westin. Freebies and privacy: What net users think. Technical report, Opinion Research Corporation, http://www.privacyexchange.org/iss/surveys/sr990714.html, 1999.
[17] N. R. Adam and J. C. Wortman. Security-control methods for statistical databases. ACMComputing Surveys, 21(4):515– 556, Dec. 1989.
[18] Mattsson, Ulf T.,' A DATABASE ENCRYPTION SOLUTION ', LinuxSecurity.com, 28 July 2004, http://www.linuxsecurity.com/content/view/116068/65/
[19] Mattsson, Ulf T., Social Science Research Network, 'A Real-time Intrusion Prevention System for Commercial Enterprise Databases', http://papers.ssrn.com/sol3/papers.cfm?abstract_id=482282
[20] Mattsson, Ulf T., Search Security and Techtarget http://searchsecurity.techtarget.com/whitepaperPage/0,293857,sid14_gci1014677,00.html.
[21] N. R. Adam and J. C. Wortman. Security-control methods for statistical databases. ACMComputing Surveys, 21(4):515– 556, Dec. 1989.
[22] R. Agrawal and J. Kiernan. Watermarking relational databases. In 28th Int'l Conference on Very Large Databases, Hong Kong, China, August 2002.
[23] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. Hippocratic databases. In Proc. of the 28th Int'l Conference on Very Large Databases, Hong Kong, China, August 2002.
[24] Agrawal, J.Kiernan, R.Srikant and Y.Xu. Implementing P3P using database technology. In Proc. of the 19th Int'l Conference on Data Engineering, Bangalore, India, March 2003.
[25] R.Agrawal, J.Kiernan, R.Srikant and Y.Xu. AnXPath-based preference language for P3P. In Proc. of the 12th Int'l World Wide Web Conference, Budapest, Hungary, May 2003.
[26] D. E. Denning. Cryptography and Data Security. Addison-Wesley Publishing Company, Inc., 1982.
[27] T. Dierks and C. Allen. The TLS Protocol - Version 1.0, Internet-Draft. November 1997.
[28] A. Freier, P. Karlton, and P. Kocher. The SSL Protocol Version 3.0, Internet-Draft. November 1996.
[29] S. Garnkel and G. Spa ord. Web Security & Commerce. O'Reilly & Associates, Inc., 1997.
[30] S. B. Guthery and T. M. Jurgensen. Smart Card Developer's Kit. Macmillan Technical Publishing, 1998.

[31] Informix. Informix-Online Dynamic Server Administrator's Guide, Version 7.1. INFORMIX Software, Inc., 1994.
[32] G. Koch and K. Loney. Oracle8: The Complete Reference. Osborne/McGraw-Hill, 1997.
[33] J. C. Lagarias. Pseudo-random number generators in cryptography and number theory. In Cryptology and Computational Number Theory, pages 115{143. American Mathematical Society, 1990.
[34] T. F. Lunt. A survey of intrusion detection techniques. Computer & Security, 12(4), 1993.
[35] National Bureau of Standards FIPS Publication 180. Secure Hash Standard, 1993.
[36] National Bureau of Standards FIPS Publication 46. Data Encryption Standard (DES), 1977.
[37] B. C. Neuman and T. Ts'o. Kerberos: An authentication service for computer networks. IEEE Communications, 32(9):33{38, 1994.
[38] San Jose Mercury News. Web site hacked; cards being canceled, Jan. 20, 2000.
[39] Oracle Technical White Paper. Database Security in Oracle8i, November 1999.
[40] W. Rankl and W. E_ng. Smart Card Handbook. John Wiley & Sons Ltd, 1997.
[41] R. Rivest. The MD5 Message-Digest Algorithm, RFC1321 (I). April 1992.
[42] R. L. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signature and public key cryptosystems. Communications of the ACM, 21:120{126, February 1978.
[43] W. Rosenberry, D.Kenney, and G. Fisher. Understanding DCE. O'Reilly & Associates, Inc., 1992.
[44] A. Shamir. How to share a secret. Communication of the ACM, 22(11):612{613, 1979.
[45] D. R. Stinson. Cryptography; Theory and Practice. CRC Press, Inc., 1995.
[46] M. Lindemann and SW Smith, Improving DES Hardware Throughput for Short Operations, IBM Research Report, 2001, http://www.research.ibm.com/secure_systems_department/projects/scop/papers/rc21798.pdf.