

Bridging Practice and Research: Secure Data Management in the Classroom

Work in Progress

Richard Savacool and Rajendra K. Raj

Rochester Institute of Technology, Computer Security and Information Assurance Program

Abstract—In recent years, information security and assurance has received considerable attention from the computing community, with universities revamping course offerings in areas such as cryptography, network security, enterprise systems security, secure coding, and digital forensics. Although secure data management is a major aspect of overall information systems security, it has received less attention than it deserves. This has not been the case at the Rochester Institute of Technology (RIT) where a course in secure database systems has been offered since 2003. The course has undergone revisions over the years and it was recently converted to be an online course.

The authors have worked in the area of information security from different perspectives. The first author, who has over a decade of practical experience in security in computer systems and holds several industry security certifications, recently took the secure database systems course at RIT as a student. The second author, who is the faculty member at RIT who developed and teaches the secure database systems course, previously worked as a software developer and manager in the financial services industry working with secure data in globally distributed systems. Based on their complementary experiences and using RIT's existing secure data management course, this work-in-progress paper describes how research and practice have been blended to create an effective course in secure data management.

Index Terms—Database security, computer security, secure data management, database systems.

I. INTRODUCTION

ALTHOUGH most organizations view their data assets as their *crown jewels* [3], keeping such data—typically stored in relational database systems—safe and secure, however, remains one of the weakest areas in information security. Security breaches in databases, including credit card and other personal financial information, have led to many states and countries enacting laws such as New York State's Information Security Breach and Notification Act to deal with such breaches. Other examples of the need for secure data management (SDM) include electronic voting systems (safety, security, and provenance of voting data), data privacy as exemplified by Health Insurance Portability and Accountability Act (HIPAA), and data outsourcing (preservation of confidentiality and integrity of personal, medical, and enterprise data). SDM is thus critical in the contemporary global marketplace, and computing students must be trained to meet its challenges.

With security having taken centerstage in an increasing number and variety of computing applications, it is

increasingly necessary to bridge the gulf between current research and industry practice in security and incorporate both into the modern computing curricula. It is also imperative that security curricula remain flexible to deal with changes in the computing world. Traditional approaches to teaching are therefore not sufficient to address the challenges of designing and developing secure computing systems. Appropriate pedagogical approaches are needed to integrate state-of-the-art security research and current practices from industry, government, and military.

Many universities have addressed the needs of information security by developing courses in cryptography, secure networking, and secure coding techniques. Few universities worldwide offer courses in SDM; the ones that do typically offer either graduate courses that focus on SDM research or undergraduate courses that present basic security features available in SQL and commercial database systems. Few attempts, if any, have been made to develop an SDM course that covers both commercial practice and current research in a holistic manner.

Rochester Institute of Technology (RIT) has offered a graduate course in SDM since 2003. This course, developed by the second author, covers basic concepts of secure data management and also includes a study of commercial practice and current research in this area. The first author who has extensive experience in the practice of secure systems recently took the course as a student. This report represents the collaboration between the two authors to identify the strengths and weaknesses of the current course and to identify opportunities to develop a more effective course that blends practice and research better.

In this work-in-progress report, we present our initial results of the effectiveness of blending practice with research in our course on secure data management. We begin by describing the course design and content, the pedagogical approach, and then discuss the most recent offering of the course. Initial evaluation of data collected over the last two offerings of the course is discussed next to determine the effectiveness of the course. We conclude by describing the current status and discuss future directions for the course.

II. COURSE DESIGN AND PEDAGOGY

A. Background

As stated in its course description, the Secure Database

Systems at RIT is a graduate course that “explores the policies, methods and mechanisms for protecting enterprise data. Topics include data reliability, integrity, and confidentiality; discretionary and mandatory access controls; secure database architectures; secure transaction processing; information flow, aggregations, and inference controls, and auditing; security models for relational, object-oriented, statistical, XML, and real time database systems.”

As prerequisite, students are expected to know fundamental database concepts including database design and modeling, architectures, database connectivity, and data organization and management. Students are also required to be competent programmers in Java, C, or C++. Students traditionally come from two majors: MS students in Computer Security and Information Assurance (CSIA) and MS students in Computer Science (CS). In recent years, graduate students in related disciplines such as Information Technology (IT) and undergraduate students in Computer Science have also opted to take this course.

Although the course includes an overview of overall system security, secure coding, network security, and basic privacy issues, it does not discuss those issues unless they arise in the context of secure data management. This approach permits focus on the many issues underlying secure database systems.

B. Course outcomes

With increased emphasis on student learning by accrediting commissions such as the Middle States Commission on Higher Education, courses at RIT emphasize course (learning) outcomes; thus, course content and student learning are guided primarily by the course outcomes. For the Secure Database Systems course, successful course completion means that a student must be able to do all of the following:

- 1) Explain basic concepts, policies, and mechanisms for building reliable and efficient secure relational database systems. [*Concepts*]
- 2) Explain how these concepts, policies, and mechanisms can be adapted for building reliable and efficient secure non-relational database systems. [*Non-relational*]
- 3) Demonstrate the design and implementation of secure policies and mechanisms to build a secure database system using a specific modern relational database system. [*Practice*]
- 4) Identify and investigate active areas of research in secure database systems. [*Research*]
- 5) Describe legal, privacy, and ethical issues in securing data and database systems. [*Ethics*]

Each outcome’s label, bracketed in italics above, serves as shorthand to remind faculty and students about the essence of the outcome. Course outcomes thus make it easier for both students and faculty to focus on the most important aspects of the course as they deal with various course activities.

C. Course topics

Course topics covered in the course essentially follow from the course outcomes. The course is structured to deal with the

two main categories of students: those with a strong security background (the CSIA students) and those with a strong database background (the CS and IT students). For the former, an overview of database topics is needed and for the latter, an overview of basic security concepts is needed. The topics covered in this course are broadly classified into three buckets.

- 1) General Topics. Included here are basic security concepts and terminology, access control mechanisms used in database systems, and integrity models and mechanisms. This is needed to introduce all students to standard terminology used in secure data management and ensure they are ready to read the current research papers in secure data management
- 2) Research. Included here is a historical research perspective covering basic multi-level secure (MLS) relational models and architectures for database systems, inference mechanisms in MLS and non-MLS systems, non-relational database systems, and a variety of topics selected from current research in secure data management. In the latest offering of the course, topics included the role of cryptography in database security, secure provenance, watermarking issues in relational databases, limiting disclosure through attribute security, forensic analysis of database logs, data inference in social networks, encrypting databases without altering structure, auditing the integrity of a database, and querying encrypted XML documents
- 3) Practice. Practitioner books by Ben Natan [3] and Litchfield [7] help to provide a list of suitable topics, which include the need to place database management system (DBMS) security within general security landscape; viewing the DBMS as a server; the role for secure communications between clients and servers; application security and proper database usage; database Trojans and database rootkits; regulations, compliance, and ethics; auditing; and DBMS case studies of Oracle, IBM DB2, or Microsoft SQL Server.

Table I shows the approximate time spent on the major topics in the course. Many of these topics have both a research and practice component. It should be noted that the time allocated is not necessarily proportional to the importance or relevance of the topic.

D. Pedagogic approach

From its inception in 2003, this course has used a pedagogic approach based on active learning (constructivist approach), and has minimized the use of passive learning (objectivist approach). Constructivism argues that students need to be active learners and construct knowledge individually based on what they already know [4]. This is arguably the only reasonable approach for learning in this course, given the constant updates to course materials as dictated by current research and industry practices. Section IV presents initial evaluation of the assessed data about this pedagogic approach.

For the past two years, this course has also been presented in a distance learning (or online) format, as well as a blended

TABLE I
APPROXIMATE TIME ALLOCATION TO EACH MAJOR COURSE TOPIC

Major Topic	Approximate Time Allocation
Reliability, integrity & confidentiality	15%
Access control and MLS	15%
Inferencing & information flow	10%
Secure database architectures	10%
Secure transaction processing	10%
Auditing and reporting	10%
Application data security	10%
Encryption	15%
Non-traditional DBMS	05%

format. Distance learning courses are conducted exclusively online and leverage collaborative learning tools such as web-based discussion forums, video conferencing, and instant messaging. The SDM course continues to use a constructivist approach to pedagogy, with traditional in-class active learning components such as discussions now moving online to discussion forums. Although the online SDM course tends to be less formal than a traditional one, students continue to be engaged in *active learning* online.

The SDM course is also offered in a *blended learning* format. Although blended learning is hard to define precisely and universally, a blended course at RIT [8] incorporates all of the following: (1) some online learning activities to complement face-to-face work, (2) around half the classroom time is replaced instructor-guided learning activities in asynchronous or synchronous interaction online, and (3) the online and face-to-face components of the course are integrated pedagogically valuable manner to ensure the best use of in-classroom and online aspects.

Both the totally online and blended sections of the course are integrated online, with the major difference being that the blended course has a weekly physical classroom meeting. The physical classroom “lecture” has tended to be student-demanded mini-lectures by the instructor and sometimes by other students or teams.

TABLE II
MAPPING GRADED COMPONENTS TO COURSE OUTCOMES

Graded Component	Course Outcomes Addressed
Team project	1 (Concepts), 3 (Practice), 4 (Research)
Research paper reviews and discussions	2 (Non-relational), 4 (Research), 5 (Ethics)
Cooperative discussions	1 (Concepts), 3 (Practice), 5 (Ethics)
Final exam	1 (Concepts), 3 (Practice), 4 (Research)

III. MOST RECENT COURSE OFFERING

The most recent offering of the course was in the Winter 2008-09 quarter. We present highlights of the course offering by describing the textbooks and other readings, online discussions, and the course project. Table II presents the various course components (in the left column) used for grading students in terms of the attainment of proficiency and in terms of the SDM course outcomes (in the right column).

Each component contributes to attainment of multiple course outcomes and each course outcome is dependent on multiple components. Section IV provides additional details about this mapping and how it is used to measure outcome achievement.

A. Textbooks and readings

The “textbooks” currently used in the course are practitioners’ books by Ben Natan [3] and Litchfield [7]. These books are supplemented by older textbooks to provide a historical perspective, for example the collection of essays on Information Security, edited by Abrams et al. [1]. Additional readings include a set of research papers, a set of industry reports, and reports of data breaches forming case studies. Previous course offerings used Afyouni [2] and Litchfield et al. [6] that are still used as supplementary sources.

Research papers are primarily accessed from the IEEE or ACM digital libraries, with the primary source of the papers being the premier database research conferences such as ICDE, SIGMOD, and VLDB, with additional content from security conferences that include data management.

B. Discussions

Because peer-based teaching [5] is a powerful learning tool, this course has been organized to foster learning through in-class discussions (for blended format students), online discussions, and research paper-driven discussions. Individual contributions are measured by both in-class and online discussions, while the paper-based discussions are more collaborative in nature.

In addition, the instructor provides the student teams each week with a set of questions on various research and practice concepts in SDM. It is left to the discretion of each collaborative team to develop a productive way of handling these discussions. Students are free to divide the work among teammates as they see fit, and often use one of two common approaches: (1) delegate specific questions to specific team members, and (2) collectively answer the questions as a group. Each group may adopt any approach for learning as long as each student has an opportunity to learn all of the material.

C. Group project

The group project requires the design and development of a two-tier or three-tier relational database application, typically built in using Oracle or Microsoft SQL Server. With the popularity of web interfaces and ease of web development, teams often choose to implement the front-end in a web-friendly format such as PHP or Java. Although much of the implementation is left to the student, the project has some minimum requirements:

- 1) Design and implement a multi-user database application.
- 2) Build an application front-end and database back-end.
- 3) Secure database applications using various best practices.
- 4) Document system architecture and security features.
- 5) Provide the project to two other teams in the class for detailed security analysis and penetration testing.
- 6) “Attack” the projects of two other teams to discover problems and report these problems to these teams.
- 7) Implement recommendations made by the two teams to mitigate discovered security issues.

Students are required to incorporate and experiment with SDM issues selected from research and practice. The project also includes developing an overall security policy for the team’s database system and application, supporting attribute- and tuple-level security and exploring auditing techniques such as custom audit triggers and fine-grained auditing.

Teams are also provided with detailed criteria for grading the team and each team member individually.

D. Use of the virtual machine lab

A notable aspect of this offering of the SDM course was the utilization of RIT’s virtual machine lab based on VMware Lab Manager. This agile development lab environment facilitates learning in security courses, for example, it is possible to have “real” network and database attacks. Systems can be run on different operating systems with different versions of database servers. The infrastructure provides increased flexibility to share workspaces among team members and to provide duplicate environments to attacking teams. The virtual machine lab provides several advantages over traditional physical machine infrastructure, for example, a quicker time-to-market, safety via snapshots in the event of failure, and a flexible software development environment.

IV. INITIAL ASSESSMENT AND EVALUATION

Our initial evaluation is based on the assessment data collected for 51 students from the last two offerings of the course as they were fairly similar in content. Results of the initial evaluation are presented here, but it should be noted that the evaluation phase is still in progress.

A. Course effectiveness

Only instructor-observed measurements of student performance were used for assessment because these direct measurements are more reliable than student self-assessments.

Table II showed how each course outcome maps to each graded course component. By mapping instructor-grades for various graded components, we can measure progress made by each student in attaining proficiency in each course outcome. It should be noted that the approach to assessment used here is the standard assessment approach used by the CS department for our accrediting agencies. That is, data being gathered and evaluated here follows our standard approach to assessment data handling used for Middle States accreditation.

Levels of proficiency reached by students on each course outcome are computed based on performance on appropriate

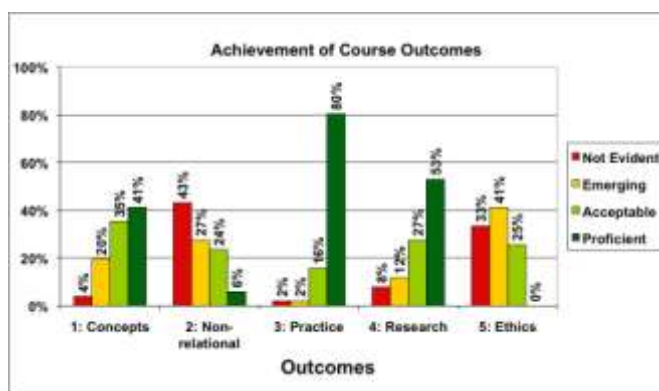


Fig. 1. Student achievement by course outcome (number of students = 51). Each bar shows percentage of students achieving that level of proficiency.

grading components used to assess that outcome. The levels of proficiency are *Not Evident*, *Emerging*, *Acceptable*, and *Proficient*. Fig. 1 displays student achievement of proficiency in each of the SDM course outcomes. For example, for the first course outcome that focused on students learning basic concepts, 41% of the students were considered to be proficient, 35% to be acceptable, 20% had emerging knowledge, and 4% had not shown any measureable evidence of learning. In other words, 76% of the students attained an acceptable or proficient level for this outcome.

High levels of achievement are also observed for the Practice outcome (96% of students attaining acceptable or better) and the Research outcome (80% of students attaining acceptable or better). On the other hand, substantially lower levels of achievement are observed for the Non-relational outcome (30% attaining acceptable or better) or for the Ethics outcome (25% attaining acceptable or none achieving proficiency). Initial analysis of these unsatisfactory numbers reveals that fewer than usual papers covering XML databases and other non-relational data were assigned. For the Ethics outcome, the problem was not that legal, privacy, and ethics were not covered in the course, but that no grading activity (in discussions, paper reviews or final exam) had been assigned to assess this outcome. This initial analysis makes it obvious that additional papers, instructor-guided discussion questions, or perhaps an exam question are needed for secure management of non-relational data. Also indicated is a final exam question to assess student knowledge and application of ethics.

Fig. 2 displays mean proficiency achieved by students on each outcome and immediately confirms the improvements

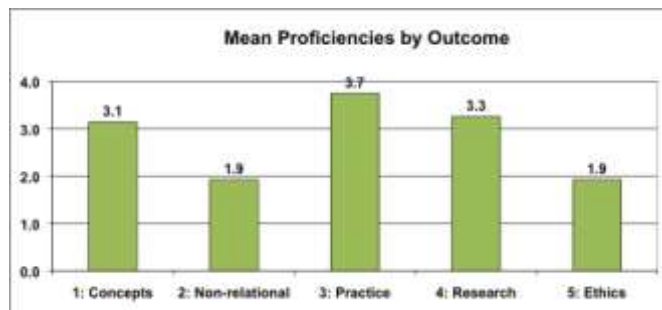


Fig. 2. Mean student proficiency by outcome (number of students = 51).

needed for Outcomes 2 and 5.

B. Students' assessment of the course

Assessment of student perceptions was also done using Likert scale measures in the earlier of the two offerings of the course. To improve objectivity, this student survey was online, anonymous, and optional, and did not contribute to the student grade. 14 out of a total of 26 students filled out this survey.

Although the sample size was small (14), an overwhelming majority—12 or higher out of 14 students—agreed or strongly agreed that the following were extremely useful in helping them understand the material: (a) individual discussions with the rest of the class, (b) paper reviews, (c) collaborative team discussions of the instructor-guided questions on research and practice, (d) collaborative teamwork on the course project, and (e) their team's attacks on other team projects. As the number of students who filled out the survey is fairly small, we do plan to assess students in future offerings of the course and report these results in a subsequent paper.

C. Challenges to learning

One of the challenges to learning often reported by students is RIT's quarter system. As each quarter is 10 weeks long, without disciplined attention to deliverables, it can be easy for students to fall behind. Careful pacing by the instructor is required to prevent course requirements from becoming overwhelming. To help mitigate this issue, an online calendaring system is used within the SDM course page to track deliverables and assist students with workload planning.

Class size also influences the level of effort associated with participatory activities such as online discussions. For example, in the most recent SDS course offering, 28 students generated nearly 400 postings for one week's online discussion; this is less of an issue with class sizes smaller than 20. Despite subsequent efforts to reduce online workload (for example, by encouraging quality over quantity in online discussions), this course achieved the dubious distinction of being rated the topmost heavy-hitter in RIT's computing college by RIT's Online Learning group that manages the university online courses environment.

Blending research with practice in the high-paced quarter schedule is indeed challenging, but student feedback indicates they prefer the current high-paced course instead of a two-quarter sequence of courses that allows both research and practice to be covered at a slower pace.

V. CONCLUSIONS

A. Current status

The course has been offered at RIT annually since 2003. It has been well received by the student community at RIT, with over 170 students having completed the course since its inception. A majority of these students have been MS students in Computer Science for whom it serves as an elective course for students specializing in Data Management. The course is required for MS students in Computer Security and

Information Assurance. Both types of students have brought diverse strengths to the course, from programming knowledge to overall secure systems knowledge, and have contributed to the overall dynamism in the course.

B. Contemplated changes

No structural changes are currently contemplated for the course. The set of research papers assigned for reading will continue to be revised each year, as will course discussions, which are based on the latest data breaches reported (which unfortunately continue to proliferate). The use of hardware (virtual or otherwise), software, and textbooks will continue to reflect the latest releases of database system and related system software.

Changes will be made to address shortcomings identified by our initial evaluation discussed in the Section IV. The next offering of the course will include a more thorough coverage of concepts and tools for secure management of non-relational data including XML data and unstructured data (e.g., blogs or web documents). While legal, privacy, and ethical issues are covered adequately at present, future course offerings will also conduct appropriate assessment for these components.

C. Final words

Over half of RIT's computer science, information technology, and computer security and information assurance graduates go on to develop, manage, or maintain systems containing sensitive data. It is therefore critical that computing students are grounded in the principles of secure data management.

RIT's SDM course bridges the divide between research and practice effectively by incorporating a variety of active learning or constructivist exercises such as the group project and cooperative discussions. The course also emphasizes how theory and pure research can be applied to solve a variety of real-world secure data management problems including data breaches and leakage. We are continuing to revise the course and working on improving our assessment of its effectiveness.

REFERENCES

- [1] M. D. Abrams, S. Jajodia, and H. J. Podell, *Information Security: An Integrated Collection of Essays*. IEEE Computer Society Press. Essays 0-3, 19-26. Available: <http://www.acsac.org/secshelf/book001/book001.html>
- [2] H. A. Afyouni, *Database Security and Auditing: Protecting Data Integrity and Accessibility*. Thomson Course Technology, Florence, KY, 2005.
- [3] R. Ben Natan, *Implementing Database Security and Auditing*, Elsevier Digital Press, Burlington, MA, 2005.
- [4] N. Crumpacker, "Faculty pedagogical approach, skill, and motivation in today's distance education milieu," *Online Journal of Distance Learning Administration*, State University of West Georgia, Distance Education Center, Winter 2001. Available: <http://www.westga.edu/~distance/ojdl/winter44/crumpacker44.html>.
- [5] M-J. Eisen, "Peer-based learning: a new-old alternative to professional development," *Adult Learning*, American Association for Adult and Continuing Education, Jan 2001. Available: http://www.accessmylibrary.com/coms2/summary_0286-7343233_ITM.
- [6] D. Litchfield, C. Anley, J. Heasman, and B. Grindlay, *The Database Hacker's Handbook: Defending Database Servers*. Wiley, Indianapolis, IN, 2005.

- [7] D. Litchfield, *The Oracle Hacker's Handbook: Hacking and Defending Oracle*. Indianapolis, IN: Wiley, 2007.
- [8] RIT Online Learning, Blended Courses. Available:
<http://online.rit.edu/faculty/blended/>.