

Security-Related Research and Projects in Computing Promote Student Awareness of Security Issues

Charles C. Tappert and Sung-Hyuk Cha
CSIS, Pace University, Pleasantville, NY 10570, USA

Abstract

Security informatics represents a paradigm shift in university curricula in computing. In order to meet this challenge we will require a systemic curriculum change beyond the usual local course and program changes that have successfully handled smaller technological advances and shifts in the past. One of the novel approaches we use to teach information security at Pace University is to introduce security-related topics, research, and projects into our existing CSIS courses. We teach our masters and doctoral students how to conduct research and write dissertations in a number of areas of computing. Also, our student project teams at both the graduate and undergraduate levels are accustomed to developing real-world computer information systems for actual customers. In recent years, and especially since 9/11, we not only direct more of our faculty research toward security issues but also encourage more security-related student research and supervise more security-related student projects, as well as devoting more lessons to security topics in our courses. This paper gives an overview of our novel approach to teaching information security awareness through our research and projects.

Keywords: security education, authentication, biometrics, forensics

1. INTRODUCTION

Information assurance represents a paradigmatic shift in college and university curricula in the information technology and computing disciplines, and the related programs must rise to this challenge (Merritt, Stix, Sullivan 2004). Although faculty in information technology disciplines have experienced profound technology change over the short history of computing and are much more frequently introducing change into courses and programs than their colleagues in other disciplines, most of these changes to date is local to courses and programs. However, the systemic change needed to appropriately incorporate information assurance into curricula is so profound that the National Security Agency offers a model that is useful to faculty and institutions for assessing and developing information systems programs relative to the components of a

comprehensive information assurance curriculum (NSA 2004).

The School of Computer Science and Information Systems (CSIS) at Pace University has, or is connected with, several centers for research and development, and some of these activities are related to security. These centers include the Center for Advanced Media (CAM 2004), the Hudson Valley Center for Emerging Technologies (HVCET 2004), the Pervasive Computing Laboratory (Pervasive Computing Lab 2004), and the newly formed Information Assurance Education and Research Center (IAERC 2004). We also have courses specifically related to security and many lessons devoted to security-related topics in other courses. In addition to these centers and faculty research, the School of CSIS at Pace University has several research programs and courses that allow for student research. The Doctor of Professional Studies in Com-

puting (D.P.S.) program enables computing and information technology professionals to earn a doctorate in three years through part-time study while continuing in their professional careers (Merritt, et al. 2001). The M.S. in Computer Science program also gives students interested in research the option of completing a dissertation during their last year of study.

In this paper we (the authors) focus on our novel approach to teaching information security through research and course projects under our direction. Research is original, rigorous work that advances knowledge, improves professional practice, and/or contributes to the understanding of subject. Research methods depend upon the nature of the research: controlled experiment, empirical studies, theoretical analyses, or other methods as appropriate. We require research work to be of sufficient strength to be able to distill from it a paper worthy of publication in a refereed journal or conference proceedings.

Project work, on the other hand, uses known technology to develop systems according to specified requirements. We have our students serve the community – the internal university community, the greater university community, and the external non-profit local community – by developing real-world computer information systems for actual customers. Many of our security-related projects are developed to provide support for our faculty and student research, some of which is in collaboration with other universities and research centers. Thus, there is interplay between the project and research activities.

We included in this study the research and project activities that were either directly or indirectly related to security. For example, we included several medical applications that are indirectly related to security because they have potential uses in disaster management. Also, although the initial work on interactive visual image classification established the methodology with a flower classification application, we include this study because we have been extending the fundamental methodology to security-related applications like face recognition.

2. SECURITY-RELATED RESEARCH

In this section we describe our (the authors') faculty and student security-related research in CSIS at Pace University. Although our masters' and doctoral student dissertations provide the major portion of our student research, several courses, including our newly initiated research seminar, provide the opportunity for both graduate and undergraduate students to conduct small research studies.

Students entering our doctoral program must have an M.S. and at least five years work experience in computing in order to ensure that they have the background to conduct a significant piece of research. The M.S. dissertation students take the research seminar to learn the methodologies necessary to complete a research study during their last year in the M.S. program. Other graduate and some undergraduate students also take the research seminar where they learn how to conduct research and complete a small research study.

Our security-related research conducted in the last three years is highlighted in Table 1.

Table 1: Security-related research.

Research Topic	Research Level
Fundamental studies for biometric authentication (Cha and Srihari 2000; Cha 2001; Shrihari, et al. 2002; Trilok, Cha, Tappert 2004; Choi, et al. 2004)	Faculty Research
Interactive visual image classification (Evans, et al. 2003; Hart, Cha, Tappert 2004)	Faculty Research
Handwriting and Forensic Document Analysis (Cha and Srihari 2000; Cha 2001; Cha and Tappert 2002; Cha, et al. 2004; Shrihari, et al. 2002; Chen, et al. 2003)	Faculty Research
Wearable/Mobile/Pervasive Computing Research (Tappert et al. 2001; Kalia et al. 2002)	Faculty Research
Automatic Language Identification from Telephone Speech (Law 2002)	Doctoral Dissertation
A Pervasive Computing Solution to Asset, Problem and Knowledge Manage-	Doctoral Dissertation

ment (Kalia 2002)	
Information Assurance Strategic Planning: A Taxonomy	Doctoral Dissertation (in progress)
Stego-Marking in TCP/IP Packets	Doctoral Dissertation (in progress)
Intrusion Detection and Prevention	Doctoral Dissertation (in progress)
Computer Forensics and Cybersecurity Governance Model	Doctoral Dissertation (in progress)
Forged Handwriting Detection (Chen 2003)	M.S. Dissertation
Assessing the Discriminative Power of Voice (Trilok 2004)	M.S. Dissertation
Use of Histogram Distances in Iris Authentication (Choi, et al. 2004)	Research Seminar Study
Interactive Flag Identification (Hart, Cha, Tappert 2004)	Research Seminar Study
Handwriting style/nationality recognition	Research Seminar Study (in progress)
Uniqueness of disguised voices	Research Seminar Study (in progress)

In the remainder of this section we briefly describe each of the faculty research areas and the completed student research studies.

Fundamental studies for biometric authentication. We consider the task of establishing the distinctiveness of each individual in a population when there is a set of measurements that have an inherent variability for each individual. This task of establishing individuality can be thought of as showing the distinctiveness of the individual classes with a small error rate in discrimination. This is important in many forensic science applications such as writer, face, fingerprint, speaker, or bite mark identification. All these applications face the problem of scientifically establishing individuality, which is motivated by court rulings such as *Daubert vs. Merrell Dow Pharmaceuticals* (U.S. Supreme Court 1993). In Dr. Cha's dissertation work, supported by NIJ, he established the individuality of handwriting using distance statistics (Cha and Srihari 2000; Cha 2001; Srihari et al. 2002). In this research, we plan to provide proper statistical validation of the methodology used in

these earlier studies and are in the process of generalizing the results to other biometric and forensic domains (Trilok, Cha, Tappert 2004; Choi et al. 2004).

Interactive visual image classification.

We are interested in the role of human-computer interaction in applications of pattern recognition where higher accuracy is required than is currently achievable by automated systems, but where there is enough time for a limited amount of human interaction. This topic has so far received only limited attention by the research community. Our current, model-based approach to interactive recognition was originated at Rensselaer Polytechnic Institute (Nagy and Zou 2002), and then investigated jointly at RPI and Pace University (Evans, et al. 2003). Our success in recognizing flowers establishes a methodology for continued work in this area. We are currently in the process of extending our approach to the interactive recognition of flags (Hart, Cha, Tappert 2004), foreign signs, faces, and skin diseases.

Handwriting and Forensic Document Analysis.

Establishing the individuality (uniqueness) of a person's handwriting is a necessary precursor to the successful development and deployment of security-related handwriting applications, such as handwriting verification (Cha and Srihari 2000) and forged handwriting detection systems (Cha and Tappert 2002; Chen 2003; Chen, et al. 2003). We utilize pattern recognition and machine learning techniques to study forensic document analysis. Both authors are members of the International Graphonomics Society and actively collaborate with document examiners and analysts. Handwriting is a part of biometric authentication research and several student research topics are derived from this faculty research.

Wearable/Mobile/Pervasive Computing Research.

This research area, earlier investigated by one of the authors (Tappert, et al. 2001), overlaps with others in that many of the applications involve small mobile devices, wireless communication, and/or pervasive technologies. Several of the following student research topics, for example, are derived from this faculty research.

Automatic Language Identification from Telephone Speech. Fast automatic language identification, recognizing a speaker's language from a speech signal, is gaining increased importance in the contexts of economic globalization and of national security. While the most accurate systems use multiple, large vocabulary continuous speech recognizers, their scalability is limited because adding a new language entails the enormous training task required to develop a complete recognizer for that language. This dissertation (Law 2002) proposes a multi-stage approach: first an efficient clustering algorithm (least cost) selects the top candidates and then speech recognizers (high cost) identify the language. It presents an efficient first pass process that extracts test and reference patterns (acoustic feature vectors) from speech utterances using cepstral coefficients, and generates reference models from the reference patterns using a Vector Quantization (VQ) clustering algorithm. It examines various distance measures in the selection phase to find the best method. It shows that this first pass provides the top candidates with a combined accuracy of over 80 percent, and is a substantial improvement over existing fast systems in discriminating among different languages. Thus, the multi-stage approach should provide the best performance in terms of speed and cost; allow the flexibility and scalability of extending it to new languages, and serve as the foundation for a real-time application.

A pervasive computing approach to asset, problem, and knowledge management. This solution reduced the average time to complete all major helpdesk tasks, and led to faster data availability, improved data quality, improved help desk effectiveness, improved help desk Return on Investment (ROI), lower overall Total Cost of Ownership (TCO), and increased user satisfaction. This is a unique conceptualization, design, and construction of wireless laptop and PDA-based asset, problem and knowledge management systems. The systems work on a Lotus Notes/Domino-based server, supporting both wired desktops and wireless laptops and personal digital assistants (PDAs), to assist in the management of a divisional desktop support team across the U.S. (Kalia 2002, Kalia, et al. 2002)

Forged Handwriting Detection. Handwriting experts are usually required to differentiate between authentic and forged signatures. Therefore, it is important to develop an objective system to identify forged handwriting, or at least to identify those handwritings that are likely to be forged. Forgers often forge handwriting in terms of shape and size by carefully copying or tracing the authentic handwriting. This dissertation (Chen 2003; Chen, et al. 2003) hypothesizes, therefore, that good forgeries – that is, those that retain the shape and size of authentic writing – are usually written more slowly than authentic writing. It also hypothesizes that good forgeries are wrinklier (less smooth) than authentic handwriting. To examine these hypotheses both online and offline data from the same handwriting samples were collected. Experimental subjects wrote handwriting samples on paper mounted on a tablet digitizer, and the x-y coordinates of these online samples were used to calculate the speed of the handwriting. The writing speed of the good forgeries was found to be significantly slower than that of the authentic writings. The paper on which the handwriting samples were written was then digitally scanned at two different resolutions to calculate a measure of wrinkliness using a fractal number estimate of the jaggedness of the writing. The wrinkliness of the good forgeries was found to be significantly greater than that of the authentic writings, showing that it is possible to identify candidate forgeries from scanned documents. These studies employed the IBM ThinkPad TransNote, pen-enabled notebook computer.

Assessing the Discriminative Power of Voice. Establishing the individuality (uniqueness) of a person's voice is a necessary precursor to the successful development and deployment of security-related voice applications, such as speaker verification and speaker identification systems. Due to the large world population, however, the task of establishing the uniqueness of a person's voice is difficult, and one that has not yet been demonstrated. The approach here (Trilok 2004) is to use a dichotomous model that has previously been used to establish the individuality of handwriting and of fingerprints. This model transforms a many-class problem into a dichotomous one by using, as pattern classifier features, the dis-

tances between measurements extracted from utterances by the same and by different speakers. The resulting statistically inferable model allows experimental results based on a relatively small sample of the population to be generalized to the entire population.

Use of Histogram Distances in Iris Authentication. Quantitatively establishing the discriminative power of iris biometric data is considered. Multi-level 2D wavelet transform has been widely used for iris verification system. While previous approaches compute only means and variances, we propose using a histogram distance. We also use a methodology to establish a measure of discrimination that is statistically inferable. To establish the inherent distinctness of the classes and validate individuality, we transform the many-class problem into a dichotomy one by using a "distance" between two samples of the same person and between those of two different peoples. We demonstrate that using histogram matching results better performances than using only means and variances (Choi, et al. 2004).

Interactive Flag Identification. This study proposes an interactive system for identifying flags in photos taken from natural scenes. The system is interactive in two respects. First, because segmentation can be a difficult problem, users are asked to crop the flag portion from a photo. Second, the user makes the final decision by selecting one of the top choices obtained from the machine classification system. The proposed system utilizes a color-based image retrieval technique. For experimental purposes a large number of flag images are synthetically generated from a small number of original ones in order to increase the reference image database. A nearest neighbor classifier produces a sorted list of candidate choices. Recognition accuracy of these choices varies from 82% to 93% depending on whether the correct flag is among the first 8 or 18 top choices, respectively, from a set of 186 flags (Hart, Cha, Tappert 2004).

3..SECURITY-RELATED PROJECTS

This section describes the security-related projects in CSIS at Pace University under the authors' direction. We use real-world team projects to provide students with the educa-

tional experience of working together in teams, similar to what is done in industry, in order to design, build, and test computer information systems. Using real-world projects with real customers is not new; for example, the University of Southern California's Center for Software Engineering (USC CSE Website 2002) has been doing this for years in their capstone two-semester software engineering sequence.

Our student project work at the graduate level comes primarily from our two-semester, capstone course in Software Engineering where earlier coursework provides the students with proficiency in programming, networking, databases, and other skills useful for project work. Our student project work at the undergraduate level is specific to the Pattern Recognition and Artificial Intelligence courses where the students are taught how to build systems using software packages like MatLab.

Our security-related projects over the last three years are listed in Table 2. The customers were from various Pace University schools and departments (CSIS, Lienhard School of Nursing, Lubin School of Business, and Department of Information Technology), and from several outside organizations (Northern Westchester Hospital, Columbia Presbyterian Medical Center, IBM T.J. Watson Research Center, RPI).

Table 2: Security-related project systems.

Project	CSIS Course
Handwriting Forgery Quiz System	Software Engineering (M.S. program)
Eigenface Recognition System	Software Engineering (M.S. program)
PC Maintenance/Tracking System	Software Engineering (M.S. program)
Nurse Information System (NIS)	Software Engineering (M.S. program)
NIS Wireless Extension	Pervasive Computing (M.S. program)
Emergency Pre-Hospital Care Communication System	Software Engineering (M.S. program)
Medical Vital Sign Wearable Computer	Pervasive Computing (M.S. program)
Multimodal Radiological Reporting System	Software Engineering (M.S. program)
Interactive Visual System	Software Engineering (M.S. program)
VoiceXML Application Development Facility	Software Engineering (M.S. program)

VoiceXML Applications	Software Engineering (M.S. program)
Multimodal Voice/InkXML System	Software Engineering (M.S. program)
Spam Detection System	Artificial Intelligence (M.S. program)
Automatic Language Detection of Text Files	Pattern Recognition (B.S. program)
User Verification based on Keystroke and Mouse Movement	Artificial Intelligence (B.S. program)

Many of the project systems consisted of a Web interface to a backend database using the client/server architecture. The simpler ones simply allowed users to enter information into the database and retrieved the data for appropriate viewing. The more complex ones performed calculations on the data, which in some cases was substantial. The projects will be briefly described.

The **Handwriting Forgery Quiz System** was developed to investigate the detection of handwriting forgery. The team collected handwriting samples from ten subjects who wrote a set of words in their natural style and attempted forgeries of the other nine subjects' handwriting. These handwriting samples were digitally scanned and stored in an image database. A Web-based quiz was developed to gather statistical data from users trying to distinguish between the original and the forged handwriting samples. Novices and certified document examiners were asked to take the online quiz. Although many students took the quiz as novices, only a few document examiners were willing to do so, and the results were inconclusive. The implications of this and similar studies may have bearing on the admissibility of handwriting testimony in U.S. courts. The database of handwriting samples was also used in a study of handwriting forgery detection by machine (Cha and Tappert 2002; Chen 2003; Chen, et al. 2003).

The **Eigenface Recognition System** (Baker, et al. 2002) was developed to investigate combining several biometrics in order to develop a high-confidence user verification system. Such a subsystem can be embedded into any system that requires high security and can be developed using various combinations of different modalities; this project used four: face, fingerprint, handwriting, and voice. Each of the four verifica-

tion systems involves techniques of pattern recognition to compare signal samples from the input device with those from the database, and each can result in type I (valid user rejected) and type II (invalid user accepted) errors. Combining these four different verification techniques should achieve a high level of confidence. Initial work on this problem focused primarily on face recognition and resulted in a database of digitized photos and preliminary studies of verification by face image using the Eigenface technique. Experimental results show that the eigenface technique on color images yields better performance than on gray scale images.

The **PC Maintenance/Tracking System** assists in the security, maintenance, support, and tracking of PCs in CSIS at Pace University. Information is captured when a PC is reported as having a problem or needing an upgrade (information such as the date, the PC ID, the PC's location, the nature of problem or upgrade), when the problem is fixed (what repairs were made), when software is installed (the software, the installer), when a PC is moved, etc.

The **Nurse Information System (NIS)** (Palmer, et al. 2002) physical assessment application walks a student nurse through a physical assessment. A legacy system on a proprietary device running an obsolete operating system was ported to a Java 2 Micro Edition (J2ME) implementation. The application was completely redesigned and rewritten using object-oriented design techniques to create portable classes for reuse and for use in future applications. A small prototype was built and scaled up to a full application on Palm personal digital assistants (PDAs) running the PalmOS. A C++ conduit was also written to allow the prototype to transfer data from the Palm handheld devices to a PC for storage and further evaluation. The same students who developed the above-described NIS in the software engineering course extended that system with an **NIS Wireless Extension** in the pervasive computing course. They provided the earlier system with a wireless extension through the use of Java servlets on a Web server to allow the data to be e-mailed as an attachment to a user-specified e-mail address (Palmer, et al. 2002). This ensures that the NIS has a cross-platform future and can mi-

grate from one device to another to take full advantage of the J2ME networking features.

A preliminary study of an **Emergency Pre-Hospital Care Communication System** was conducted for Northern Westchester Hospital (Park, Pastore, Tappert 2002). Many hospital emergency departments are exploring new communication and information technologies that will assist them in providing higher quality of care by improving the speed of flow, the consistency, and the accuracy of information shared among all the parties involved in an emergency response team. This study describes some currently available or on-the-horizon communication and information technologies that may be appropriate for use in the emergency services field, discusses three systems currently being used by other hospitals, and provides three alternative approaches to obtaining the type of emergency communication system that will best fit the needs of Northern Westchester Hospital's emergency department. The study found that the solution to be chosen should depend on the time frame allotted for the project, the funds budgeted, and the number and skills of the staff members responsible for implementing the system.

The **Medical Vital Sign Wearable Computer** monitors a patient's vital signs, specifically heart rate and breath rate, which serve as indicators of a patient's immediate general health. If performed on a continuous basis and monitored remotely, it could assist medical support personnel. For example, each patient could be equipped with a medical status monitor that constantly reads blood pressure, pulse, respiration, and blood oxygen level, and a built-in expert system could determine when the patient's condition warrants attention and, if necessary, automatically transmit an alarm. This work extends that of a West Point student project in which a liquid-filled sensor pad placed close to the skin captured a signal that was processed by digital filtering to extract the heart and breath signals (Tappert, et al. 2001). As this project was nearing completion, we became aware of a similar product, the LifeShirt, which is a lightweight shirt with embedded sensors that measure respiratory function, heart rate, posture, and activity level, with optional peripheral de-

vices to measure blood pressure and blood oxygen saturation (VivoMetrics 2002).

The **Multimodal Radiological Reporting System** communicates between an end user and a computer. The Radiological Device gives the user the option of choosing the communication mode: voice, ink, or touch-tone. Once the information is entered in the system, it is interpreted into voice / ink XML and give the user the correct respond. The automated system will allow the user to modify records in the current database for patients with certain health problem. It will also query the database for information concerning the person's record in the database. This will be accomplished through the use of voice and ink engines that will obtain the results of what was input on the note pad, or through the microphone. The event handler will then decide on the appropriate action to take based on a multi-modal grammar that is specific to the ink, voice and DTMF action that the user took. The automated system would then return the action to the user that was taken, and it will update the database with the correct data if verified by the user. Or the user will have the option to correct the information that is entered and then re-submit the information. Once the user verifies the data for submission the system will update the appropriate data in the database. This data will then be able to be seen by a web interface that directly accesses the database to return user information.

Interactive Visual System. Mobile computing devices are being endowed with ever-increasing functionality. To demonstrate the augmentation of human cognition in an interactive visual recognition task, this project reengineered a PC-based system called CAVIAR (Computer Assisted Visual Interactive System) for a handheld computer with camera attachment. The resulting Interactive Visual System (IVS) (Evans, et al. 2003) exploits the pattern recognition capabilities of humans and the computational power of a computer to identify flowers based on features that are interactively extracted from an image and submitted for comparison to a species database. Whereas many of the more interesting mobile applications are communications based, IVS is autonomous. While IVS has similar functionality to that of CAVIAR, because it runs on a handheld computer it offers complete portability for

use in the field. We find that the handheld IVS and PC-based CAVIAR systems outperform humans alone both on speed and accuracy and machines alone on accuracy. This project is included here because it is our initial implementation of an interactive visual recognition system on a handheld computer, and our future plans include extending the underlying technique to interactive face recognition which is clearly security related.

The **VoiceXML Application Development Facility** was created to facilitate VoiceXML application development, and several small applications were used to test the system. The lab provides a gateway to develop voice applications, the interface for developing applications together with a template library, the capability to enable novices to use the code library to learn about and to build advanced applications, and facilitates both novice and expert developers in deploying multiplatform applications. The laboratory hardware consists of a Cisco router for Voice Over IP, the IBM Voice Server, a Local Area Network (LAN), the Public Switched Telephone Network (PSTN), and a Firewall for the Voice Server. A minimal prototype of the facility is now operational, and upon completion of the facility, the architecture of the VoiceXML studio should permit the registration, development, and deployment of an application. The studio essentially enables the user to load his/her application URL that can be referenced by the IBM Voice Server.

Another team worked on **VoiceXML Applications**. The primary application, a voice-enabled Web-based absentee system (Gallivan, et al. 2002), was implemented on the TellMe voice portal. It was tested by the class of software engineering students; students who intended to miss a class called the VoiceXML telephone number and were led through an automated dialog to record in a database his/her name, the date and time of the call, the courseID, and the date he/she would miss class. The system provides the instructor and other administrators with a permanent record of absentees that can be accessed and displayed through a Web interface. A more useful application of such a system would be for the university, or for any reasonably sized organization, to record absences or work time missed by its employees. Rather than calling a secretary, employees would telephone the system to

report their absences, and the employer could then access this information through the Web interface; other departments (or systems) needing the information, such as payroll, could also access the database.

A third XML system, the **Multimodal Voice/Ink XML System**, combined voice and pen input using the standard voice and ink XML formats (Trabelsi, et al. 2002, 2002b). Just as a standard VoiceXML format that facilitates the development of voice applications has been accepted, so the acceptance of a standard InkML format to facilitate pen applications is anticipated once it is developed. The student team developed a multimodal interface architecture that combines standardized voice and ink formats to facilitate the creation of robust and efficient multimodal systems, particularly for noisy mobile environments. By providing mutual disambiguation of input signals and superior error handling, this architecture should broaden the spectrum of users to the general population, including permanently and temporarily disabled users. Integration of VoiceXML and InkML provides a standard data format to facilitate Web-based development and content delivery. Diverse applications ranging from complex data entry and text editing applications to Web transactions can be implemented on this system, and a prototype platform and sample dialogues were developed.

Spam Detection System (Stuart, Cha, Tappert 2004). Most e-mail readers spend a non-trivial amount of time regularly deleting junk e-mail (spam) messages, even as an expanding volume of such e-mail occupies server storage space and consumes network bandwidth. An ongoing challenge, therefore, rests within the development and refinement of automatic classifiers that can distinguish legitimate e-mail from spam. A few published studies have examined spam detectors using Naïve Bayesian approaches and large feature sets of binary attributes that determine the existence of common keywords in spam, and many commercial applications also use Naïve Bayesian techniques. Spammers recognize these attempts to thwart their messages and have developed tactics to circumvent these filters, but these evasive tactics are themselves patterns that human readers can often identify quickly. This preliminary study tests an alternative

approach using a neural network (NN) classifier on a corpus of e-mail messages from one user. The feature set uses descriptive characteristics of words and messages similar to those that a human reader would use to identify spam. The results of this study are compared to previous spam detectors that have used Naïve Bayesian classifiers.

Automatic Language Detection of Text Files. As distance barriers collapse and modern technology unites the world, it is desirable to develop a method to determine the language that information is expressed in. The language classifier application was designed to classify eleven languages: English, Spanish, French, Italian, German, Swedish, Polish, Dutch, Romanian, Portuguese, and Danish. The application has five main steps. First it asks the user for a text sample. Second, the text sample is cleaned up and prepared for processing. Third, the cleaned sample is sent to the feature extractor, processed, and put into a vector. The vector would be sent to the classifiers, which would then finally output the answer. The language classifier started out as a simple idea, but as with many ideas, when put into action, their execution is not so simple.

User Verification based on Keystroke and Mouse Movement. It has been hypothesized that each person types on a keyboard in a characteristic way. In order to validate this hypothesis, a simple experimental project was conducted. The program records the mouse movement, clicking speeds, keystroke speed, etc. An artificial neural network was trained and 84% accuracy was reached. Although further experiments are necessary, this project has a potential to verify the users based on their mouse and keyboard usages.

4. DIRECTING THE RESEARCH AND PROJECTS

In this section we discuss the activities involved in directing the research and project activities. For dissertation work at the doctoral and master's levels, and also for our research seminar, we introduce the student to the research process. Students are first given specific research papers to read and are expected to participate in discussions about the essence and context of the research area. Then, students are introduced

to the literature search process and begin the individual progression toward a specific problem by finding several articles that relate to a research area of their interest. The steps of the research process are typically to find a general topic of interest and then a specific problem in that area, to review the literature to investigate previous approaches, to describe the problem and approach in a research proposal, to conduct the research (obtain data, develop system, run experiments, analyze results, etc.), and finally to document the research by writing a dissertation or research paper.

For the projects, on the other hand, the instructors solicit and interact with customers to set up new projects, work with the computer support personnel to create the project development infrastructure, and monitor the systems development process. Projects came from faculty interested in developing systems to further their research (sometimes in collaboration with other universities or with technology companies), from other departments or schools of the university needing computer information systems, from non-profit community institutions such as local hospitals, and from interests of the students. The instructor sizes and shapes each project to be an appropriate systems development experience for the students, and posts project descriptions on the course Website. The inducement for the customers' involvement with the students throughout the year is the anticipation of receiving a useful system, although they are warned that the primary purpose of the projects is to provide the students with a good educational experience and that not all projects are successfully completed. The instructor then forms the student teams and assigns each team to a project.

The development of the computer information systems requires a systematic approach by the students and proceeds through an evolutionary process model consisting of analysis, design, build, and test phases. The first phase requires the students to work closely with their customer to understand what is desired, usually by developing use cases and use scenarios about how the system will be used; to plan the project; and to perform a risk analysis, all culminating in a written requirements document. The second phase involves designing the system using

various analysis and design tools, and this work varies greatly depending on the type of system being developed. For example, a database backend requires database design methodologies, whereas a Web interface requires interface and Web design engineering. The third phase is the actual construction of the system and, depending on the project, usually requires the use of one or more computer software languages and/or database implementation. The fourth phase is the testing of the system to ensure that it meets the customer's requirements and that all of its functions operate correctly.

The project infrastructure is substantial. This infrastructure is also used, although to a lesser extent, for research activities, mostly for storing and managing images, speech data, and other information in databases. Since most of the projects use the client/server architecture, two NT servers and two UNIX based servers with Solaris-Oracle configuration, having access only within the university, provide development platforms for the student systems. The students are allowed to install software on these hosting platforms as required by their projects. After the development phase, the students move their application systems to a staging server having access from outside the university. The development and staging servers are independent and separate from the CSIS production servers so that students can not corrupt data or interfere with operations on those servers. The application systems use a variety of database-related software, including different scripting languages, such as Cold Fusion, PHP, Perl, and JSP, to communicate with backend databases in Microsoft Access, MySQL, MS-SQL Server, and Oracle. Some projects also use Java-related software, such as Java servlets, and Tomcat is installed to handle the processing of that code. A quality assurance team checks the quality of the project systems on the staging server using metrics meeting the quality standards of the software industry. Finally, those systems that we want to continue using, and that meet the quality standards, are migrated to production servers. For the project systems that are not web related other equipment is set up as required.

5. BENEFITS OF RESEARCH AND PROJECTS

In this section we describe the benefits of having our students conduct research studies and develop computer information systems for real customers in the security area.

The research students learn the required individual skills necessary to conduct a research study. They learn how to perform literature searches to gain general knowledge about an area and to determine what previous work has been done on a specific problem. They learn organizational and critical thinking skills, how to be innovative and creative, and how to structure and perform their research studies. Finally, in writing their dissertations they learn how to set their research in a proper context, to describe their methodology and findings, and to estimate the potential impact of the work.

The students doing project work obtain a stellar real-world learning experience as individual technologists, as team members, and as maturing professionals in the computing discipline. Individually, the students learn the technology skills necessary to develop real-world computer information systems. Through project reviews and team presentations, the students also learn about the various technologies used in other projects, and they especially appreciate the exposure to projects involving cutting-edge technology and research. Working in teams, the students learn fair-mindedness, intellectual humility, intellectual integrity, and the ability to work with others to produce useful systems and to take responsibility for them. Because most of the students are employed full time in various areas of computing, they bring their knowledge and expertise to bear in their project work, and by exchanging information, they learn from each other in this student-centered learning environment. As maturing professionals, the students learn how to act in the computing field not only as technologists but also as value providers. By working with real customers in developing their project systems and focusing on human-centered computing, the students learn important value skills (Denning and Dunham 2001).

These research and project development learning paradigms foster lifelong habits for

learning and the application of critical thinking and value skills. These activities also appear to be initiating what may become a significant impact on our undergraduate program. Bringing together undergraduate and graduate students in the research seminar gave the undergraduates a taste of research, and introducing project activities in several undergraduate courses gave the students a taste of industry-like team work and associated benefits. These experiences have motivated several of the undergraduates to continue their studies at the graduate level.

Finally, we found an interesting interplay between the projects and the research. We create some projects, either specifically or secondarily, to provide appropriate data or infrastructure for research studies. For example, the students that developed the forgery quiz system were required to construct a database of authentic and forgery handwriting samples for their interactive quiz system, and this database was subsequently used in the forgery research studies. Not only can appropriate structuring of the project activities can be beneficial to the research work, but the research activities often lead to interesting projects.

6. CONCLUSIONS

Overall, the security-related research studies and projects result in a beneficial outcome for all concerned. The research students learn the research methodologies and the joy of making new discoveries. The project students learn the technological skills of the computing discipline, team-related skills, and value skills by following a human-centered development process. Projects also provides a vehicle for fostering interdisciplinary collaboration, encouraging student involvement in the university and local communities, furthering student and faculty research, enhancing relationships between the university and local technology companies, and increasing national recognition of the university. Finally, it is interesting to note that some undergraduates, having been introduced to research and/or project activities, were motivated to continue their studies at the graduate level.

7. REFERENCES

- [1] Baker, W., Evans, A., Jordan, L. and Pethe, S., 2002, "User verification system." *Proc. MASPLAS*, <http://csis.pace.edu/cs/masplas/>
- [2] CAM – Center for Advanced Media, 2004, School of CSIS, Pace University, <http://csis.pace.edu/~cam/> (accessed May).
- [3] Cha, Sung-Hyuk and Srihari, Sargur N., 2000, "Writer Identification: Statistical Analysis and Dichotomizer." *Proc. SPR & SSPR, LNCS-Advances in Pattern Recognition*, vol. 1876, p 123-132.
- [4] Cha, S.-H., 2001, "Use of Distance Measures in Handwriting Analysis." PhD dissertation, SUNY Buffalo, CSE, March.
- [5] Cha, S.-H. and Tappert, C.C. 2002, "Automatic detection of handwriting forgery." *Proc. IWFHR-8*, Niagara, Canada, pp. 264-267.
- [6] Cha, S.-H., Chee, Y.-M., and Tappert, C.C., 2004, "Automatic Detection of Handwriting Forgery using a Fractal Number Estimate of Wrinkliness." To appear in *Int. J. Pattern Recognition and Artificial Intelligence*.
- [7] Chen, H.-C. 2003, "Forged Handwriting Detection." M.S. Dissertation, CSIS, Pace University, May.
- [8] Chen, H.-C. , Cha, S.-H., Chee, Y.-M. and Tappert, C.C., 2003, "The Detection of Forged Handwriting Using a Fractal Number Estimate of Wrinkliness." *Proc. 11th Int. Graphonomics Soc. Conf.*, Scottsdale, AZ, November, pp. 312-315.
- [9] Choi, S.-S., Yoon, S., Cha, S.-K., and Tappert, C.C., 2004, "Use of histogram distances in iris authentication." *Proc. MCSCE 2004 MLMTA*, Las Vegas, NV, June.
- [10] Denning, P.J. and Dunham, R., 2001, "The core of the third-wave professional." *Communications of the ACM*, Vol. 44, No. 11, pp. 21-25.
- [11] Evans, A., Sikorski, J., Thomas, P., Zou, J., Nagy, G., Cha, S.-H., and Tappert, C.C., 2003, "Interactive Visual System." *Pace CSIS Tech. Report 196*.
- [12] Gallivan, P., Hong, Q., Jordan, L., Li, E., Mathew, G., Mulyani, Y., and Visokey, P., 2002, "VoiceXML Absentee System." *Proc. MASPLAS*, <http://csis.pace.edu/cs/masplas/>.
- [13] Hart, E., Cha, S.-H., and Tappert, C., 2004, "Interactive Flag Identification using Image Retrieval Techniques." *Proc. MCSCE 2004 CISST*, Las Vegas, NV, June.

- [14] HVCET – Hudson Valley Center for Emerging Technologies, 2004, <http://www.csis.pace.edu/hvcet/> (accessed May).
- [15] IAERC – Information Assurance Education and Research Center, 2004, School of CSIS, Pace University, <http://csis.pace.edu/csis/cqi-front/sec/security.pl> (accessed May).
- [16] Kalia, S.K., 2002, "A Pervasive Computing Solution to Asset, Problem and Knowledge Management." Doctoral Dissertation, School of CSIS, Pace University, online <http://www.pace.edu/library/pages/theses/>
- [17] Kalia, S.K., Tappert, C.C., Stix, A., and Grossman, F., 2002, "A Pervasive Computing Solution to Asset, Problem and Knowledge Management." *Proc. E-Learn 2002 World Conf. on E-Learning in Corporate, Government, Healthcare, and Higher Ed.*, Montreal, October.
- [18] Law, J., 2002, "An efficient first pass of a two-stage approach for automatic language identification of telephone speech." Doctoral Dissertation, School of CSIS, Pace University, available online at <http://www.pace.edu/library/pages/theses/>
- [19] Merritt, S.M., Grossman, F., Tappert, C., Bergin, J., Blum, H., Frank, R., Sachs, D., Stix, A., and Varden, S., 2001, "The Doctor of Professional Studies in Computing: An Innovative Professional Doctoral Program." *Proceedings of ISECON 2001*, v 18 (Cincinnati): paper 17b, October, <http://isedj.org/isecon/2001/17b/index.html>
- [20] Merritt, S.M., Stix, A., and Sullivan, J.E., 2004, "Information Assurance Across the Curriculum: A Paradigm Shift in Information Systems Education." *CSIS Tech. Report*, Pace University, in preparation.
- [21] Nagy, G. and Zou, J., 2002, "Interactive Visual Pattern Recognition." *Proc. Int. Conf. Pattern Recognition*, vol. III, pp. 478-481.
- [22] NSA National IA Education and Training Program, 2004, <http://www.nsa.gov/ia/academia/acad00001.cfm> (accessed May).
- [23] Palmer, S., Panchee, N., Sullivan, J., Thabet, K. and Westgard, S., 2002, "Migrating an application to Java2 Micro Edition." *Proc. MASPLAS*, <http://csis.pace.edu/csis/masplas/>
- [24] Park, H.G., Pastore, Jr., J.M. and Tappert, C.C., 2002, "Wireless technologies in pre-hospital communications." *CSIS Tech. Rep. 176*, Pace Univ.
- [25] Pervasive Computing Laboratory, 2004, School of CSIS, Pace University, <http://www.csis.pace.edu/~ctappert/pervasive/> (accessed May).
- [26] Srihari, S. N., Cha, S.-H., Arora, H. and Lee, S., 2002, "Individuality of Handwriting." *Journal of Forensic Sciences*, vol. 47, no. 4, pp 856-872.
- [27] Stuart, Ian, Cha, Sung-Hyuk, and Tappert, Charles, 2004, "A Neural Network Classifier for Junk E-Mail." *Proc. DAS 2004, 6th IAPR Int. Workshop on Document Analysis Systems*, Florence, Italy, Sept.
- [28] Tappert, C.C. et al., 2001, "Military applications of wearable computers and augmented reality." Ch. 20, pp. 625-647, in *Fundamentals of Wearable Computers and Augmented Reality*, ed. Barfield, W. and Caudell, T., Lawrence Erlbaum.
- [29] Trabelsi, Z., Cha, S.-H., Desai, D., and Tappert, C.C., 2002, "Multimodal Integration of Voice and Ink for Pervasive Computing," *Proc. IEEE 4th MSE*, Newport Beach, CA, December.
- [30] Trabelsi, Z., Cha, S.-H., Desai, D., and Tappert, C.C., 2002, "A Voice and Ink XML Multimodal Architecture for Mobile e-Commerce Systems." *Proc. 2nd ACM Int. Workshop on Mobile Commerce*, Atlanta, GA, September.
- [31] Trilok, P.N., 2004, "Assessing the Discriminative Power of Voice." M.S. Dissertation, School of CSIS, Pace University, January.
- [32] Trilok, P.N., Cha, S.-H., and Tappert, C.C., 2004, "Establishing the Uniqueness of the Human Voice for Security Applications," *Proc. CSIS Research Day*, Pace University, NY, May.
- [33] U.S. Supreme Court ruling, 1993, "Daubert vs. Merrell Dow Pharmaceuticals." 509 U.S. 579.
- [34] USC Center for Software Engineering, 2004. <http://sunset.usc.edu/cse/> (accessed May).
- [35] VivoMetrics Website, 2002, <http://www.vivometrics.com/ce/> (accessed 2002).