DARPA-BAA-13-16

VOLUME 1: Technical and Management Proposal

Technical Areas 1a and 1b

Pace University

Continual Computer-Usage Behavioral Biometric Authentication

Technical POC: Charles C. Tappert, Ph.D.

Professor, Computer Science Department
Goldstein Room 325, 861 Bedford Road, Pleasantville NY 10570-9913
Tel 914-773-3989, ctappert@pace.edu, Fax 914-773-3533

Administrative POC: Victor Goldsmith, Ph.D.

Associate Provost for Sponsored Research and Economic Development
Room 312, 163 William Street, New York NY 10038
Tel 212-346-1277, vgoldsmith@pace.edu, Fax 212-346-1116

Award Instrument: Grant for Institution of Higher Education

Places and Periods of Performance: Pace University Campuses in NYC, White Plains, and Pleasantville for duration of contract

Proposal validity period: 12 months beginning on start date

DUNS Number: 064961022
Taxpayer Identification Number: 135562314
CAGE Code: CAGE/NCAGE OKT12

Table of Contents

Executive Summary …….. 1

Goals and Impact ……... 1

Technical Plan ……..…….. 2

Management Plan …….. 8

Capabilities ……... ……...…….. 11

Statement of Work (SOW) …….. 17

Schedule and Milestones ……...….. 19

Cost Summary ……...….. 20

Appendix A

Appendix B

Executive Summary

Currently, users of computer systems are authenticated once per session by inputting difficult-to-remember passwords at the time of login. To augment and possibly replace these non-intuitive password authentication systems, this proposal is aimed at developing a behavioral biometric system to continually authenticate users of DoD desktop computers and users of mobile devices. Four research task areas will be investigated, and four corresponding biometric system components developed for integration into one robust system. The four biometric system components operate at various human cognitive levels to provide a multi-level computational behavioral cognitive "fingerprint" of the person operating the computer. The four biometric components target *keystroke, mouse activity, stylometry, and operational behavior*. The keystroke and mouse components operate at the subconscious automatic motor control level, the stylometry component operates at the higher cognitive linguistic (character, word, syntax) level, and the operational behavior component operates at the highest cognitive semantic level of intentional motivation. The proposed project will leverage our previous work – our extensive research on the keystroke biometric and our ongoing research on mouse activity, stylometry, and operational behavior. Using standard biometric performance metrics, the four biometric system components will be evaluated for viability through extensive testing, on over 500 experimental participants, using standard DoD-equipped office desktop computers and a variety of mobile devices. Given our extensive experience in the pertinent research areas, we are confident that the research tasks conducted will provide a solid foundation for a unique and more human-oriented biometric approach to the authentication process. Our cost estimate is less than $500K for one year of work. Demonstrations of system component performance will be conducted.

Goals and Impact

The proposed team will develop and evaluate four biometric system components that target keystroke, mouse activity, stylometry, and operational behavior biometrics, which can be combined into a single robust biometric system to provide a multi-level computational behavioral cognitive "fingerprint" of the person operating the computer. If successful, the combined system can essentially replace current awkward and non-intuitive password authentication systems to qualitatively improve the nature of authentication systems. Each of the four components will be innovative and unique:

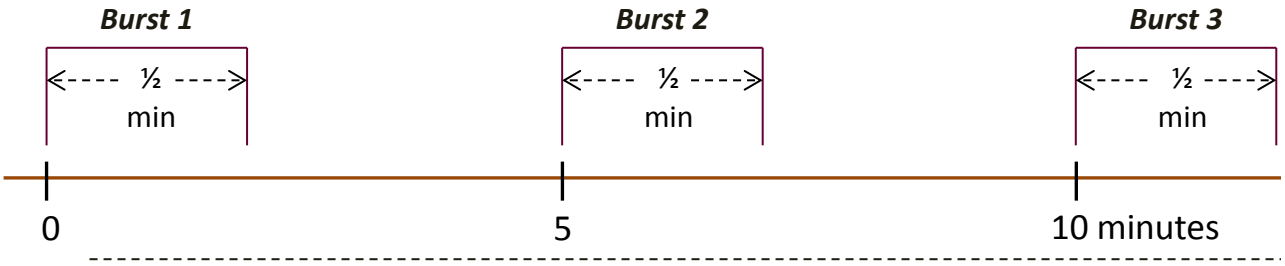
1. Keystroke - While most work in the keystroke biometric area and all the commercial keystroke dynamic systems focus on passwords, our system focuses on considerably longer-than-password text input and on arbitrary text input such as report writing, email, tweets, etc. This longer input permits the use of powerful statistical feature measurements – and the number, variety, and strength of the measurements used in the system are much greater than those used by other long-text-input systems reported in the literature, significantly rising above the state-of-the-art. Key goal: attain a ROC curve operating point of FAR = 0.001 and FRR = 0.01 for a population of 500 users inputting 200-keystroke samples (½ minute input for a fast typist or 1 minute for an average computer user).
2. Mouse movement – Key goal: attain a ROC curve operating point of FAR = 0.01 and FRR = 0.1 for a population of 500 users on samples of ½ to 1 minute duration.

3. Stylometry – Key goal: attain a ROC curve operating point of FAR = 0.1 and FRR = 0.3 for a population of 500 users on text samples of 2 or more minutes duration. Note that longer input is usually required for stylometry because it operates on words.
4. Operational behavior – Key goal: to be determined.

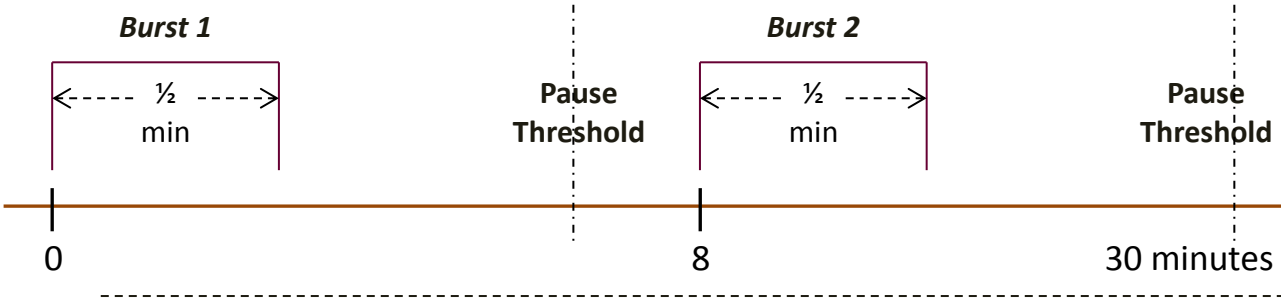
The proposed project is revolutionary in that it aims to combine four distinct and rather orthogonal behavioral computer user traits into one robust system. Although we have no plans to commercialize this work, we do plan to implement and deploy such a system to authenticate online student test takers at our university. As long as the raw data captured is not compromised which would allow mimicking of users, we see nothing to prohibit the sustainment of the technology over its entire lifecycle.

Technical Plan

For clarification we define several terms. **Active Authentication** is continual authentication which is ongoing but with possible interruptions, in contrast to continuous authentication which would mean without interruption. **Burst Authentication** is authentication on a short period of input **after a pause**. We believe this is an important concept. An obvious strategy would be to have a moving data interval window that captures, for example, a minute of computer input for an authentication check every fixed-interval-spacing of five minutes as shown in Figure 1 (a).



(a) Uniform burst authentication



(b) Burst authentication with pauses

Figure 1. Burst authentication

However, we believe it is only necessary to capture the first burst of input after each pause. Computer users often pause for a telephone call, conversation with a colleague, coffee/bathroom break, etc. Also, there would likely be a pause for the entry of an intruder. Therefore, only after a pause would re-authentication of the user be required as described in Figure 1 (b).

The primary motivation for using this concept of burst authentication is to reduce the frequency of independent authentication checks. This has the advantages of reducing the false alarm rate, avoiding the capture of unnecessarily large quantities of data and using excessive computing resources to process the data, while still providing sufficient data for continual training of the biometric system components.

There are two time periods that need to be determined. One is the **length of the pause** for burst authentication which needs to be shorter than the entry time of an intruder. Therefore, estimating plausible intruder entry times will provide the critical upper bound on the pause time. Measuring actual authentic user pauses once the system is deployed could provide useful additional data to determine the potential savings resulting from the reduced authentication frequency of the burst mode relative to the fixed-interval-spacing mode. Note that in an open office environment with computers close together and available to many users, the plausible pause time between an authentic user and an intruder may be negligible, causing the burst authentication approach to revert to the fixed-interval-window-spacing approach.

The second time period of interest is the **length of the data capture authentication window**, which is presumably a minute or less. This needs to be short enough to catch the intruder before significant harm is caused, yet long enough to make an accurate detection and reduce false alarms. We propose in this study to measure the tradeoff between the time length of the authentication window and the biometric performance, providing a plot of performance as a function of the length of the authentication window, so that a reasonable operational trade-off can be made when the system is deployed.

The occurrence of **low-volume computer input** must also be considered. For example, with a user browsing the Internet or checking email or while on a phone call, the computer input activity may not provide sufficient data for authentication in a short window. Furthermore, in situations, such as a phone call or drinking coffee, in which the user is using only one hand for keyboard input, the data may not be representative of normal user behavior. Fortunately, low-volume computer input of this nature would also not be considered likely intruder behavior. Therefore, data capture windows containing only small quantities of data can probably be safely ignored. The threshold for the quantity of data required for reasonable authentication is therefore an additional parameter to be determined.

Technologies

We envision undertaking an approach to the active authentication problem that will enable revolutionary advances in the science of behavioral biometric authentication in computer security applications. Our unique innovative approach will operate at several human cognitive levels to provide high-performance continual authentication that can be integrated into a powerful cyber-security system. Using a variety of technologies – biometrics, statistical pattern recognition, machine learning, data mining, and experimental testing procedures – behavioral biometric system components will be developed to continually authenticate computer users based on their input. For the computer environment we will target the standard DoD office environment desktop computer (the United States Government Configuration Baseline, USGCB) that specifically includes keyboard, mouse, Windows 7 operating system, network interface card, connection to a printer, and the standard DoD software product suite of Microsoft Office applications. For a broader use of the technology, users of applications like Google Apps, a cloud based suite, which is Federal Information Security Management Act certified (FISMA),

and may be (or soon be) in-use at the DoD, could also be authenticated. Additionally, we will develop corresponding biometric components for mobile devices.

The biometric system components operate at various human cognitive levels to provide a multi-level computational behavioral cognitive "fingerprint" of the person operating the computer. These biometric components will employ keystroke, mouse activity, stylometry, and operational behavior biometrics. The keystroke and mouse components operate at the subconscious automatic motor control level, the stylometry component operates at the higher cognitive linguistic level, and the operational behavior component operates at the highest cognitive semantic level of intentional motivation (Figure 2).

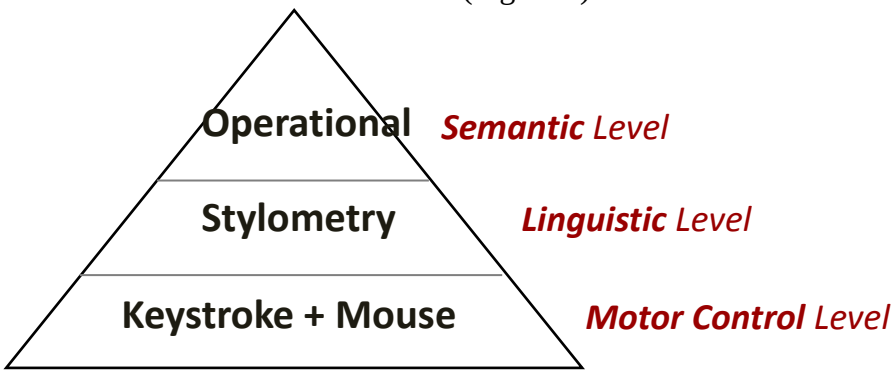


Figure 2. Behavioral biometrics and human cognitive levels.

This work will leverage our previous studies – our extensive research on the keystroke biometric and our ongoing research on mouse activity and stylometry. With our previous extensive experience in developing the Pace University text-input keystroke biometric system, we will develop a more general, innovative, and robust keystroke biometric system that will handle input from all the commonly used applications, such as text (Microsoft Word, email, etc.), spreadsheet (Microsoft Excel, etc.), and web browser input (Microsoft Internet Explorer, etc.). With our previous experience on mouse activity and stylometry biometrics we will develop unique, innovative, and robust mouse and stylometry biometric systems. We will also develop a unique, innovative, and robust biometric system that targets operational behavior at the computational linguistic and semantic levels.

We believe these technologies capture enough of the unique qualities of an individual for high-performance biometric authentication. For example, there is substantial evidence in the literature that the keystroke biometric is reasonably reliable on text input (see Tappert et al., 2010 and associated references). Therefore, with the addition of other keyboard input like spreadsheet and browser activity, together with the weaker stylometry and mouse biometrics, we strongly believe a combined biometric system will provide a highly usable authentication biometric system. While the keystroke biometric work on text input has already been done, this work will be extended to handle keyboard input other than text. The other biometric work is relatively new or has been only preliminarily explored. The limitations of the keystroke biometric, and presumably the other proposed biometrics as well, are that performance decreases over time as the individual's behavior patterns change, but this limitation can be eliminated by continually collecting new input data and updating the system. In fact, the continual updating of the system should improve the biometric system performance as the quantity of training material increases. We anticipate that the impact of this work will be substantial – for example, it will have the capability of authenticating users of computers in the workplace (in both the government and the private sector), university students taking online tests (a goal of the 2008

Higher Education Opportunity Act), and users of mobile devices performing online transactions (a goal of the 2013 NIST-NSTIC-01, National Strategy for Trusted Identities in Cyberspace, Pilots Cooperative Agreement Program).

Methods of Testing

As for expected variability, reliability, and accuracy, we will be providing Receiver Operating Characteristic (ROC) curves that portray the expected accuracy of the biometric system. There is substantial evidence in the literature that the keystroke biometric is reasonably reliable, particularly if the system is continually trained on new data (Tappert et al., 2010). However, with the addition of the weaker stylometry and mouse activity biometrics, the system should be even more robust. The intruder operational behavior biometric could be more robust than the mouse or stylometry biometrics, but with occasional false alarms when authentic users initiate activities that could be interpreted as coming from an intruder.

The system components will be evaluated separately with the aim of later combining them into a powerful overall active authentication biometric system for government desktop computer users and mobile users. We plan to use standard biometric performance metrics to evaluate and validate the various components and overall system. These standard methods include the derivation of Receiver Operating Characteristic (ROC) curves that portray the trade-off between the False Accept Rate (FAR) and the False Reject Rate (FRR), and show the Equal Error Rate (EER) where FAR = FRR. We recently conducted keystroke experiments on 14, 30, and 119 users –Figure 3 shows EER as a function of the number of keystrokes, and Figure 4 shows the ROC curves at the maximum 755 keystrokes (Monaco et al., 2013).

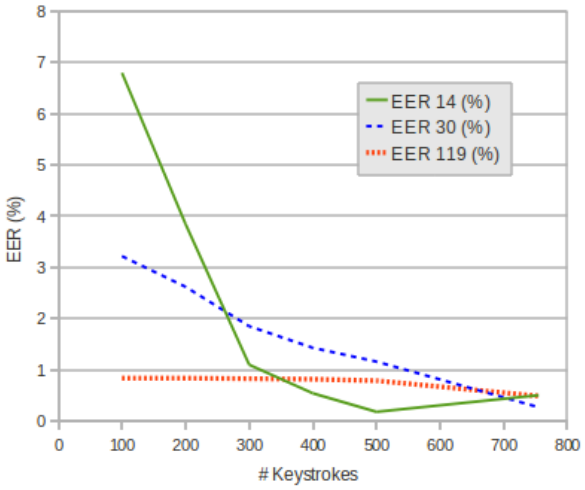


Figure 3. EER versus #Keystrokes (Monaco et al., 2013).

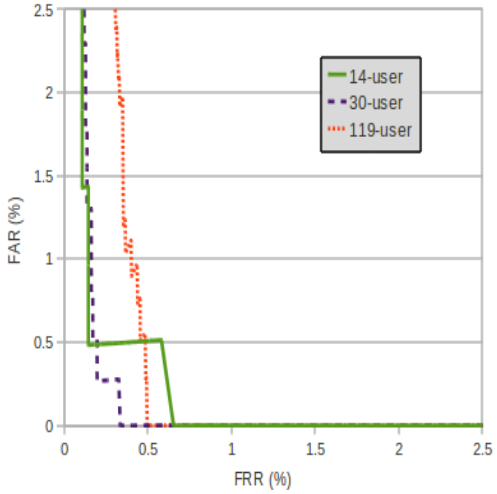


Figure 4. ROC curves (Monaco et al., 2013).

Another test we are interested in conducting, and one that is not commonly performed, is to obtain a breakdown of system performance over the user population. For most biometrics it is well known that there are subjects that are easy to authenticate because they have distinctive traits and those that are difficult to authenticate because their traits are not very distinctive. These two categories of subjects are traditionally known as sheep and goats, respectively. To obtain this breakdown of system performance over the user population we can measure each user's FAR and FRR. Knowing this information can be important in a deployed system because administrators can judge the risk of a non-authentication alarm based on knowledge of the

performance characteristics of the purported user causing the alarm. Over time, for example, it would be possible to make certain adjustments, like setting a different operating point on the ROC curve for individuals likely to trigger far more false alarms than the average.

Over 500 experimental participants will provide extensive testing data. Demonstrations of system performance will be conducted. Additional details of the testing methodology are explained later in the section on "Dichotomy vector-difference authentication model," under "Capabilities."

Data Collection

The four biometric system component evaluations will be performed through extensive testing on more than 500 (and possibly up to 1000) experimental participants (perhaps paying students an hourly wage). Smaller quantities of data will be collected in the early months of work to test the various biometric system components under development. A university is an ideal environment for obtaining a large number of experimental participants.

Computer input data types will be collected separately for text input, spreadsheet input, and browser input; and also collected for arbitrary input from a variety of the aforementioned types. A keylogger will be installed on university machines designated for data collection. To address privacy concerns the experimental participants will be warned not to input confidential information.

The data will be collected in five-minute bursts of computer input. This will allow for testing on 15-second, 30-second, one-minute, two-minute, etc. portions of the samples to measure performance as a function of computer input sample length. Instructions for the participants will be determined in the early planning stages. We may have specific scenarios for the participants to perform, as well as free unstructured computer input. For the keystroke data, for example, we will likely collect bursts of computer input from several applications – text input in Microsoft Word, email input, spreadsheet input in Microsoft Excel, browser input, etc. For the intruder operational behavior biometric we plan to collect some attack scenario data in which participants mimic conducting specific intruder behavior, such as using SQL injection to an online website, or accessing sensitive data. For the mobile devices, we will collect similar data from corresponding applications. For the various types of input, 5-10 samples will be obtained from each participant since our earlier work found such sample sizes sufficient for system evaluation.

False Alarm Rates

We will be providing ROC curves that will permit the installers of the biometric system components to set the operating point at an appropriate trade-off between FAR and FRR. For the overall system, methods can also be used to quietly handle most false alarms, such as visually checking the computer user if cameras are available. Another possible method would be to immediately alert the computer user of an anomaly and cause a pop-up window to appear asking the user to re-enter a password or other identification information. A drawback of this method is that it might scare off an intruder. For additional security, the entered user ID and password could be strengthened via keystroke password "hardening." So as not to disturb the employees unnecessarily, only as a last resort might an audible alarm be sounded and Military Police deployed to an office potentially under attack by an intruder.

For an example estimate of the false alarm rate, assume four burst authentication windows are required per user per active hour, one every 15 minutes. Assume the average user uses the computer four hours per day, which is likely an overestimate considering vacations, business

trips, meeting times, and other non-computer usage times. For a 1000 employee facility, this gives:

- 1000 users * 16 authentications/user/day * 22 work days/month = 352K authentications/month

Assume EERs for keystroke, mouse, stylometry, and intruder-like behavior biometrics:

- keystroke 0.005 * mouse 0.05 * stylometry 0.1 * operational 0.1 = 0.0000025

These assumptions result in ***less than one false alarm per 1000-person facility per month***, a seemingly tolerable false alarm rate. This false alarm rate assumes statistically independent performances of the component systems which is optimistic. However, it also assumes operating at the EER, while most biometric systems operate on the ROC curve at a considerably lower FAR than FRR. For example, according to our current keystroke performance, shown in Figure 4, operating at a FRR of 0.01 would essentially drop the FAR to zero for large keystroke samples. Although a lower FAR operating point would incur more false rejections, several authentication failures could be required before making an unauthorized-user decision. Thus, a goal of less than one false alarm per 1000-person facility per month appears attainable and we anticipate further system improvement. It will be interesting, for example, to determine whether this low false alarm rate can be maintained with a population of 500 or even 1000 users.

Privacy, Security and Resiliency, Interoperability, Cost Effectiveness and Ease of Use

The behavioral data collected (keystrokes, mouse events/movements, writing style, and semantic operational characteristics) are non-invasive personal characteristics or traits that can be reduced to template data so the raw data are never stored. Users can be told that these personal characteristic data types are collected, that only the templates are stored as long as they remain employed and deleted when a user leaves, that the templates are used to verify their identity for the security of the organization, and that security measures are taken to ensure proper retention of the data while the person is in an employee relationship. Because of the nature of these data it will not be possible for the employee to realistically review and correct the information.

In a government facility (or private facility for that matter) it is assumed that the desktop computers are owned by the government (or private company) and that it is perfectly legal and appropriate to install keyloggers on the machines. Of course, users should be warned that the machines are monitored and non-keylogger machines could be made available for employee use during non-work times such as coffee breaks, lunch, etc.

Because these are biometric identity solutions they are inherently secure. They are also resilient as long as the raw data are not stored and the template data are not compromised. For example, one possibility is to operate with limited homomorphic encryption, using a one way function to create/update the template so that personal information strictly remains on the local machine. These biometric solutions provide material security advances over the usual password mechanisms in place today. They are also reasonably secure and reliable methods of electronic authentication, and demonstrate the integration of all major aspects of the project, but as biometric solutions they are not 100% accurate. Finally, combined with other authentication methods, such as the standard password, user authenticity is ensured and strengthened.

The biometric solutions proposed will enable service providers to augment the existing variety of credentials with additional ones. Over time it is anticipated that these biometric solutions could foster the reduction and elimination of policy and technology silos. The degree of

interoperability of the biometric services will be explored for different machines, keyboards, mice, and mobile equivalents.

The biometric solutions proposed utilize existing user actions performing their work and therefore require no special user training, and as such they are simple to understand, intuitive, and easy to use. Because the equipment used to capture the biometric information is furnished to make the work possible, there are no additional equipment costs, and these solutions lower barriers for user acceptance.

Management Plan

The principal investigators on this proposal are research-oriented faculty of Pace University. They have extensive management expertise and will oversee the contract work performed.

- Dr. Tappert, while at IBM, was the principal investigator on a series of six one-year Rome Air Development Center contracts in the area of speech recognition, and later the project leader on several internal projects in the areas of handwriting recognition and pen computing. While at West Point he was the course director of a projects course in which the cadets built computer information systems for various departments at the academy, and for several summers he was the principal investigator on projects investigating wearable computing equipment at the Aberdeen Army Research Laboratory. At Pace University, he has advised the completion of 15 doctoral dissertations and teaches a master-level capstone projects course in which student teams frequently build the supporting infrastructure for doctoral student research.
- Dr. Cha has advised 7 doctoral dissertations and 9 master dissertations at Pace. Prior to joining Pace University, he was affiliated with the Center of Excellence for Document Analysis and Recognition (CEDAR) and a research team member of the individuality of handwriting biometric project funded by National Institute of Justice. This project was granted a U.S. Patent in 2009 (United States Patent No. 7580551 B1).
- Dr. Chen has extensive grant management experience. She is a principle investigator (PI) on two National Science Foundation grants, a co-PI on one Department of Defense grant, and a PI on three Verizon Foundation grants. Dr. Chen currently serves as the program director of the Scholarship for Service (SFS) program at Pace University, a program supported by the National Science Foundation. The SFS program trains Pace students to become Information Assurance professionals who are obligated to enter the federal workforce upon graduation. She is also an adviser to students in the DoD Information Assurance Scholarship Program at Pace University. Dr. Chen conducts research in the areas of web application security, usability and security compliance. Her research team consists of doctoral, masters and undergraduate students.
- Dr. Grossman, while at Columbia University's Hudson Laboratories (1966 – 1968), New York University Graduate School of Engineering and Science (1968 – 1973), and Polytechnic University (1973-1979), was co-principal investigator on research grants sponsored by the Office of Naval Research (ONR), the Advanced Research Projects Agency (ARPA), and the National Science Foundation (NSF). This research was concerned with programming language design applied to input, output and processing of mathematical text (formulas and equations) in its normal two-dimensional format; the support of automated mathematics via such a programming system; the computer verification and typesetting of

mathematical text; and pattern recognition applied to computer recognition of hand-printed mathematics. At Columbia University's Hudson Laboratories he was the Assistant Director and later Director of the Computing Department. At NYU Engineering he was a Research Scientist and Director of the Systems Science Lab. At Polytechnic he was Associate Director of the Westchester Graduate Center, Director of Computer Science and Information Systems – Westchester, and Director of the Computer Science Doctoral program. At Pace University, Dr. Grossman was Chair of Information Systems program in the Lubin Graduate School (1979-1983), Program Advisor of M.S. in Information Systems, (1985-1996). He has been the Program Chair of the Doctor of Professional Studies (DPS) in computing since 2000. He organized two companies during the period of 1975 to 1981 for the purpose of designing and developing business systems for the apparel industry. The systems provide full support for order processing, billing, inventory control, production/purchasing, manufacturing including bill-of-materials, labor and materials cost-accounting, accounts receivable, sales analysis, complete automated EDI processing. He was the co-founder and President of KGK Automated Systems, Inc. He did research and development in the automation of the programming process in scientific and mathematical applications. He developed a software system, The Automated Programmer, which was marketed. This tool permitted the user to specify solutions in normal mathematical notation and technical English and then automatically generated executable Fortran or C code. Some of his other administrative and entrepreneurial activities include: Member Steering Committee, NY City SPIN, 2004 – 2009; Co-founder and Vice President of Board of Directors, New York Software Industry Association -- 1992 – 1999; Co-founder and Member Board of Directors, New York Software Alliance -- 1992 – 1999; Academic Federal Credit Union, Member of the Board of Directors and Chair, Supervisory Committee, 1984 – present.

The research and high-level work will be conducted by the principal investigators in conjunction with research-oriented students, primarily doctoral and masters-level students supported by this contract. This research work will involve the development of the biometric component algorithms, systems, testing, and evaluation. The detailed lower-level work, such as collecting the data from the experimental participants, and running the data through the biometric system components, as well as some of the programming, will be conducted by masters-level graduate research assistants supported by this contract and by student project teams in Dr. Tappert's capstone masters-level projects course that runs each semester. The majority of the experimental participants will be paid undergraduate/graduate students. However, the capstone project students and graduate assistants will also provide data for developing and initial testing of the system components.

Masters-level Capstone Projects Course

An important aspect of this course is the interplay of student projects and research conducted by graduate students and/or faculty. One of the novel approaches we use is to support doctoral student dissertation and faculty research by creating research-supporting projects. We teach our dissertation students how to conduct research, and our student project teams how to develop real-world computer information systems. In recent years, we have experimented with the interplay of dissertation research and projects created specifically to develop the supporting software infrastructure for that research. Many of the project customers are faculty members or dissertation students who need supporting software infrastructures to conduct their research. Thus, there is an exciting and productive symbiotic interplay relationship between the project

and research activities. Between 2005 and the present, over 30 masters-level capstone projects contributed to the biometric research that relates to this proposal by creating supporting software infrastructure. Our M.S. in Computer Science students have a thesis option and there is currently one related thesis in progress. The project work has been documented in over 25 internal Pace University conference papers – 20 on the keystroke biometric, three on the mouse activity biometric, and four on the stylometry biometric (Appendix B).

Doctor of Professional Studies (DPS) in Computing

The Doctor of Professional Studies (DPS) in Computing at Pace University provides computing and information technology (IT) professionals a unique opportunity to pursue a doctoral degree while continuing to work full time. The program supports interdisciplinary study among computing areas and applied research in one or more of them, thus providing students with a background highly valued by industry. It is an innovative post-master's doctoral program structured to meet the needs of the practicing computing professional. The DPS in Computing, while advanced in content and rigorous in its demands, can be distinguished from the Doctor of Philosophy (PhD) in that it focuses on the advancement of the practice of computing through applied research and development. It is designed specifically for people who want to do research in an industrial setting.

Much of the research and development that has contributed to our biometric systems, such as the robust Pace Keystroke Biometric System, was performed by our doctoral students (five completed dissertations and one in progress), and this research area continues to provide a number of avenues for additional research.

Pace University's Seidenberg School of Computer Science and Information Systems

As one of the first schools of computing in the country, the Seidenberg School of Computer Science and Information Systems remains in the forefront of the field. Though much has changed since its founding in 1983, the School still maintains the same core values: a diverse, practical education founded in theory and based in reality. The School's goal is to become one of the leading academic institutions for research in cybersecurity. The Seidenberg School was among the first schools designated as a National Center of Academic Excellence in Information Assurance Education (CAE/IAE) by the National Security Agency (NSA) and the Department of Homeland Security (DHS).

The Seidenberg Cybersecurity Institute, launched in 2011, aims to become a forum to propagate basic and applied research and to address the acute shortage of trained cybersecurity professionals through the offering of certificates, degree programs, and ongoing education for managers and high-level executives. The Institute builds on the expertise of the Seidenberg faculty. Today, over 25 full-time Pace faculty, most of whom reside in the Seidenberg School, are involved in cybersecurity teaching and/or research. Their cybersecurity-related work over the past five years resulted in one book, four book chapters, and over 40 publications in journals and conference proceedings. Research streams include biometrics, information security management, web security, computer forensics, information technology auditing, intrusion detection, national security, IA education, and privacy.

The School also provides scholarships to students pursuing a course of study in cybersecurity, through both the NSF's Scholarship for Service Program and the DoD's Information Assurance Scholarship Program.

Pace University

Pace University provides high quality professional education and training coupled with an excellent liberal education to students for whom that education offers the opportunity to lift their lives and prospects to new levels. Pace is a university dedicated to offering a wide array of programs of education for professions in demand, framed by the perspective and independent critical thinking that comes from an excellent liberal education. In selected areas in each school or center, Pace offers professional programs that are among the best in the New York tri-state area. The University seeks to relate its programs of professional education to the most important currents in those professions, capitalizing on its location in and around New York City to offer students real-world experience through internships and co-operative work experiences, using community service as a learning tool and employing problem-solving and other teaching techniques that re-enforce the relationship between a student's university experience and professional challenges and satisfactions.

Capabilities

The project investigators have expertise in a number of technologies relevant to this work – biometrics, pattern recognition, machine learning, data mining, security, and software system development and testing. Over recent years we have developed the following technologies and innovative algorithms that we plan to use and extend in this work:

- Dichotomy vector-diff. authentication model (Cha & Srihari, 2000; Yoon et al., 2005)
- Unique kNN classifier that generates ROC curves (Zack, 2010; Tappert et al., 2010)
- Robust Pace University long-text-input keystroke biometric system (Curtin, 2006; Villani et al., 2006; Villani, 2006; Ritzmann, 2007; Tappert et al., 2010; Monaco et al., 2012; Monaco et al., 2013)
- Stylometry biometric system (Stewart et al., 2011; Stewart, 2012)
- Preliminary mouse activity biometric system (several internal publications, see below)
- Preliminary operational behavior biometric system (one internal publication, see below)

Pace University has conducted research in a number of areas related to this proposal: the dichotomy vector-difference model, the unique kNN classifier that generates ROC curves, and the individual biometrics of keystroke, mouse activity, stylometry, and operational behavior.

Raw Data Capture

We have a multi-platform behavioral biometrics logger which can run natively on all popular platforms. For less intrusive applications, we also have a JavaScript API which can run within a browser only. We also have a mobile logger which runs natively on android devices with the capability to extend the platform to tablets and other touchscreen devices.

Dichotomy Vector-Difference Authentication Model

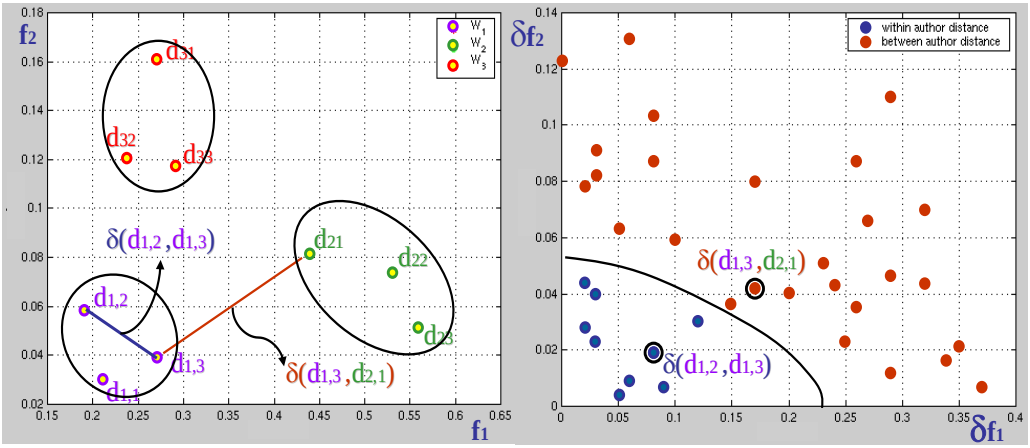
The dichotomy vector-difference model was invented by Dr. Cha during his dissertation work (Cha & Srihari, 2000). This unique and innovative model transforms a multi-class problem into a two-class problem. The resulting two classes are "within-class (intra-person), you are authenticated" and "between-class (inter-person), you are not authenticated." This is a strong inferential statistics method found to be particularly effective for multidimensional feature-space problems (Cha & Srihari, 2000; Yoon et al., 2005; Zack et al., 2010; Tappert et al., 2010; Stewart et al., 2011). All of our biometric system components use this dichotomy model.

To explain the dichotomy transformation process, take an example of three people $\{P_1,P_2,P_3\}$ where each person supplies three biometric samples. Figure 5 (a) plots the biometric sample data for these three people in two-dimensional feature space. This feature space is transformed into a feature-difference space by calculating vector distances between pairs of samples of the *same* person (*intra-person distances*, denoted by $x_{\oplus}$) and distances between pairs of samples of *different* people (*inter-person distances*, denoted by $x_{\varnothing}$). Let d_{ij} represent the individual feature vector of the i^{th} person's j^{th} biometric sample, then $x_{\oplus}$ and $x_{\varnothing}$ are calculated as follows:

$x_{\oplus} = |d_{ij} - d_{ik}|$ *where* $i=1$ *to* n, *and* $j,k=1$ *to* m, $j \neq k$

$x_{\varnothing} = |d_{ij} - d_{kl}|$ *where* $i,k=1$ *to* n, $i \neq k$ *and* $j,l=1$ *to* m

where n is the number of people, m is the number of samples per person, and the absolute value is of the elements of these vectors. Figure 5 (b) shows the transformed feature distance space for the example problem.



(a) Feature space (b) Feature distance space

Figure 5. Transformation: feature space (a) to feature distance space (b) (Yoon et al., 2005).

If n people provide m biometric samples each, the numbers of intra-person and inter-person distance samples, respectively, are:

$$n_{\oplus} = \frac{m \times (m-1) \times n}{2} \quad n_{\varnothing} = m \times m \times \frac{n \times (n-1)}{2}$$

In the authentication process, a user's keystroke sample requiring authentication is first converted into a feature vector. The difference between this feature vector and an earlier-obtained enrollment feature vector from this user is computed, and the resulting difference vector is classified as within-class (intra-person) for authentication or between-class (inter-person) for non-authentication. The *k*-nearest-neighbor method performs this classification by comparing this feature-difference vector against those in the training set. Thus, differences of difference vectors are being calculated. Because most pattern recognition systems calculate difference vectors in the matching/classification process, the fact that the dichotomy model takes differences of difference vectors is often not comprehended, even by reviewers of our papers.

To obtain system performance we simulate the authentication process of many true users trying to get authenticated and of many imposters trying to get authenticated as other users. This is done by using the numbers of the inter- and intra-person distances explained above. For example, if we have eight keystroke samples from each of 15 users, then (from the equation

above) there are 420 intra-person distances to simulate true users and 6420 inter-person distances to simulate imposters. The feature distance space is populated similarly during training.

A more accurate "engineering" procedure matches the incoming test feature vector against all the enrollment vectors from the claimed user using the leave-one-out procedure (Monaco et al., 2013). Due to the large number of possible vector-difference matches for each authentication test, and especially for the inter-person distances, it was necessary to reduce the number of training difference vectors for efficiency and performance. By evaluating several reduction methods, an innovative one was discovered that retains only the difference vectors that include samples from the purported user. The number of retained between-person distance vectors for both positive and negative authentication tests was found to be $O(n)$, a considerable reduction especially from the original procedure inter-person matches of $O(n^2)$. Most importantly, this new procedure greatly improved system performance as indicated by the new keystroke results shown in Figures 3 and 4.

Unique kNN Classifier that Generates ROC Curves

A unique method was developed to derive Receiver Operating Characteristic (ROC) curves directly from *kNN* classification results without having to estimate density functions. The derivation of an ROC curve requires a controlling parameter, usually a threshold, and the *kNN* procedure also has a parameter *k* that can be varied. Three procedures were developed that extend the *kNN* one-parameter method into a two-parameter method, and having two parameters greatly extends the operating options and tradeoffs when deploying the system (Zack et al., 2010; Tappert et al., 2010).

The Pace University Keystroke Biometric System

The robust Pace University keystroke biometric system for long-text input has been developed over the past eight years. For input of 300 or more keystrokes it is probably the best system in the world today, and 300 keystrokes is one minute of input at the typing rate of 50 words per minute. As described earlier and shown in Figures 3 and 4 (Monaco et al., 2013), for a population of 119 users, authentication performance measured by ERR was under 1% on input of 100 or more keystrokes and 0.4% on 755 keystrokes, while a more realistic operating point was FRR = 1.00% and FAR = 0.00%. It remains to be determined whether such a remarkable operating point can be obtained for larger populations or for smaller keystroke samples.

The system has user-identification and user-authentication Internet applications that are of increasing importance as the population of these application users continues to grow. An example of a user-authentication application is verifying the identity of students taking online quizzes or tests, an application that is becoming more important with the student enrollment of online classes increasing and instructors becoming concerned about evaluation security and academic integrity. Similarly, in a business setting employees can be required to take online examinations in their training/orientation programs where the companies would like the exam-takers authenticated. An example of a user-identification application in a small company environment is a closed system of known employees where there has been a problem with the circulation of inappropriate (unprofessional, offensive, or obscene) e-mail, and it is desirable to identify the perpetrator. Because the inappropriate email is being sent from company computers, there are no ethical issues in capturing users' keystrokes. In addition, as more businesses move to e-commerce, the keystroke biometric in Internet applications can provide an effective balance between high security and customer ease-of-use (Yu & Cho, 2004).

Keystroke biometric systems measure typing characteristics believed to be unique to an individual and difficult to duplicate (Bolle et al., 2004; Jin et al., 2004). The keystroke biometric is one of the less-studied behavioral biometrics and has been reviewed in several articles (Karnan et al., 2011; Revett, 2008). While most of the systems developed previously have been experimental in nature, there are about ten commercial software products for keystroke biometric authentication, primarily for password "hardening," on the market.

The keystroke biometric is appealing for several reasons. First, it is not intrusive, but rather transparent, to computer users who type frequently for both work and pleasure. Second, it is inexpensive since the only hardware required is a computer with keyboard. Third, keystrokes continue to be entered for potential repeated checking after an authentication phase has verified a user's identity (or possibly been fooled) since keystrokes exist as a mere consequence of users using computers (Gunetti & Picardi, 2005), and this continuing verification throughout a computer session is sometimes referred to as dynamic verification (Leggett et al., 1991).

While most earlier studies used passwords or short name strings (Bender & Postley, 2007; Bolle et al., 2004; Giot, El-Abed & Rosenberger, 2009; Killourhy & Maxion (2009); Li et al. 2011; Monrose, Reiter & Wetzel, 2002; Monrose & Rubin, 2000; Obaidat & Sadoun, 1999; Revett, 2008; Rodrigues et al., 2006), some used long-text input (Bergadano, Gunetti, & Picardi, 2002; Gunetti & Picardi, 2005; Leggett et al., 1991; Messerman et al., 2011; Peacock, Ke, & Wilkerson, 2004). Gunetti and Picardi (2005) focused on long free-text passages and also attempted the detection of uncharacteristic patterns due to fatigue, distraction, stress, or other factors.

Generally, a number of measurements or features are used to characterize a user's typing pattern. These measurements are typically derived from the raw data of key press times, key release times, and the identity of the keys pressed. From key-press and key-release times a feature vector, usually consisting of keystroke duration times and keystroke transition times, can be created (Woodward et al., 2002). Such measurements can be collected from all users of a system, such as a computer network or web-based system, where keystroke entry is available, and a model that attempts to distinguish an individual user from others can be established. For short input such as passwords, however, the lack of sufficient measurements presents a problem because keystrokes, unlike other biometric features, convey only a small amount of information. Moreover, this information tends to vary for different keyboards, different environmental conditions, and different entered texts (Gunetti & Picardi, 2005).

The Pace University keystroke biometric system is unique in several respects. First, it can collect raw keystroke data over the Internet as well as from a keylogger, which is desirable for Internet security applications such as those described above. Second, it focuses on long-text input where sufficient keystroke data are available to permit the use of powerful statistical feature measurements – and the number, variety, and strength of the measurements used in the system are much greater than those used by earlier systems reported in the literature. Third, it focuses on applications using arbitrary text input because copy texts are unacceptable for most applications of interest. However, because of the statistical nature of the features and the use of arbitrary text input, special statistical fallback procedures were incorporated into the system to handle the paucity of data from infrequently used keyboard keys.

The system consists of a raw keystroke data collector (a Java applet to collect raw data over the Internet or a keylogger installed on a local machine), a feature extractor, and pattern classifiers to make identification or authentication decisions. Experiments on over 100 subjects investigated two input modes – copy and free-text input – and two keyboard types – desktop and

laptop keyboards. The system can accurately identify or authenticate individuals if the same type of keyboard is used to produce the enrollment and questioned input samples. Longitudinal experiments quantified performance degradation over intervals of several weeks and over an interval of two years. Additional experiments investigated the system's hierarchical model, parameter settings, assumptions, and sufficiency of enrollment samples and input-text length. Although evaluated on input texts of several thousand keystrokes, we found that input of 300 keystrokes, roughly four lines of text easily typed in well under a minute, is sufficient for the important applications described above.

In a book chapter, journal article, and several conference papers, the Pace University keystroke biometric system has been described in detail and put into context with related work in the literature (see references below). The IJISP journal article (Tappert et al., 2010) best describes the system. Our most recent results demonstrate dramatically improved performance over earlier results (Monaco et al., 2013). Our work on the keystroke biometric began in 2005 and, in addition to external publications, has been documented over the years in 20 internal Pace University conference papers (Appendix B).

The Pace University Mouse Activity Biometric System

Beginning in 2007 we developed mouse movement biometric systems for user identification and authentication. The features of our current mouse movement biometric system are divided into the following categories: *mouse trajectory, mouse click, mouse wheel spin or scroll, applications accessed, mouse activity time*. Mouse trajectories can arise from the following actions: *system wake up, move and click, highlight, and drag and drop*. Sometimes these actions occur in combinations. For example, to highlight a piece of text in a text editor, the usual sequence is *move and click* followed by *highlight*, and the same click that ends the move begins the highlight. A sequence to move a piece of text (e.g., in Microsoft Word) could consist of *move and click* to locate the text, *highlight* to highlight the text, keying CTRL-x to cut the highlighted text, a *move and click* to locate the insertion point, and a CTRL-v to insert the text.

The trajectories produced by most of these mouse actions manifest different characteristic properties even within the same individual. In particular, the *system wake up* action produces individually characteristic trajectories within most individuals that are different from the other mouse trajectory-producing actions and important to handle separately. Also, the *highlight* and *drag and drop* trajectory actions are usually different from the *move and click* action because greater care in producing the trajectory is likely required when the mouse button is held down. The *highlight* and *drag and drop* trajectory actions, however, are rather similar.

The following nine basic trajectory measures can be calculated within any mouse trajectory: *number of trajectory points, time of the trajectory, point-to-point distance in the trajectory, length of the trajectory, point-to-point velocity in the trajectory, point-to-point acceleration in the trajectory, point-to-point direction angle change, number of inflection points in the trajectory, curviness of the trajectory*. In a user's session sample, however, there can be many trajectories and the following **biometric features** are derived from the basic measures.

1. From the **number of trajectory points** over all trajectories in a sample: mean, median, min, max, std
2. From the **time of the trajectory** over all trajectories in a sample: mean, median, min, max, std
3. From the **point-to-point distance** over all trajectories in a sample: mean, median, min, max, std
4. From the **length of the trajectory** over all trajectories in a sample: mean, median, min, max, std
5. From the **point-to-point velocities** over all trajectories in a sample: mean, median, min, max, std
6. From the **point-to-point accelerations** over all trajectories in a sample: mean, median, min, max, std

7. From the **point-to-point direction angle changes** over all trajectories in a sample: mean, median, min, max, std
8. From the **number of inflection points** over all trajectories in a sample: mean, median, min, max, std
9. From the **curviness of the trajectory** over all trajectories in a sample: mean, median, min, max, std

The above 45 (9x5) features are computed from the sample-session mouse trajectories within each of the three mouse action types that produce trajectories – *wake up; move and click; and highlight* and *drag and drop combined.* – resulting in 135 statistical mouse trajectory features. Features are also obtained from the other major mouse activity categories. Since 2007 the mouse activity biometric research resulted in three (and a fourth in progress) internal Pace University conference papers (Appendix B).

The Pace University Stylometry Biometric System

Stylometry is the study of the unique linguistic styles and writing behaviors of individuals in order to determine authorship, and has been used to attribute authorship to anonymous or disputed documents. Stylometry uses statistical analysis, pattern recognition, and artificial intelligence techniques, and typically analyzes the text by using word frequencies and identifying patterns in common parts of speech. At Pace we have been interested in the question of the authorship of email, an area of forensic linguistics. In recent years we developed an experimental system consisting of data collection, feature extraction, and classification components. In contrast to prose writing, email is less formal and characterized by shorter sentences, shorter words, different format/structure, and often chat-room or other shorthand such as 'lol', 'asap', 'btw', etc. The feature extraction component calculates such measurements as the frequency of use of each alphabet letter, punctuation, and other commonly used characters; average number of sentences per paragraph; average number of words per sentence; average word length; the percentage use of capitalization; the use of common shorthand; and other features that specifically characterize the informality of email. Because we already focus on the informality of email, we have a promising beginning toward broadening the features to the informal input expected to be analyzed under this contract.

In 2011, using keystroke and stylometry biometrics, we conducted a study on 40 university students actually taking online tests (Stewart et al., 2011; Stewart, 2012). This work was targeted at the 2008 federal Higher Education Opportunity Act that requires institutions of higher learning to make greater access control efforts to assure that student of record are those actually taking the exams. Compared with the keystroke biometric on long-text input, stylometry yields considerably weaker performance with best results so far of about 10% EER. Since 2007 the stylometry research has been documented in five internal Pace University conference papers (Appendix B), in addition to the two external publications mentioned above.

Preliminary Operational Behavior Research

We have recently initiated a study on intruder detection via semantic-level operational behavior biometrics where an unauthorized person (the intruder) uses a desktop computer posing as an authorized user. A keylogger is used to capture computer input that is completely arbitrary and independent of any application(s) running on the user system. An intruder, for instance, could impersonate an authentic user to send malicious email, modify financial documents, submit fake expense reports, search for private account codes, install malware, etc. While most of the features used in this study come from typical keystroke timings, we will also employ features based on the occurrences of specific commands or keys an intruder might enter from a command prompt. These include DOS commands (cd, dir, copy, del, systeminfo, regedit, etc.), UNIX commands (ls, cp, rm, whoami, chmod, ipconfig, etc.) and executable file extensions (exe, com,

dll, etc.). Because an intruder will likely interact with the GUI, we are including mouse information – context, clicks, trajectory, speed, and acceleration. This effort is currently focused on intrusion detection in the private sector but could easily be extended or redirected to apply to government activities.

Statement of Work (SOW)

The SOW is presented in four sections corresponding to the proposed four research task areas of keystroke, mouse activity, stylometry, and intruder operational behavior biometrics.

Keystroke Biometric Research Task 1

Objective: The objective of the keystroke biometric effort is to design, develop, test, and evaluate a keystroke biometric system targeting DoD desktop/laptop computers.

Approach: The Task 1 Research Area will provide the data gathering framework for all four research task areas, although some additional equipment and labor are required for Tasks 2, 3, and 4. The Task 1 Research Area is therefore essential for the other three task areas. Most of the equipment requested will be used to extend our laboratory facilities for the work of all four task areas. We are currently using a free open-source keylogger that captures keyboard and mouse input, although we do not currently use the mouse information. Therefore, we will appropriately recode the keylogger to provide three types of data: keystroke data appropriate for our keystroke system, mouse activity data appropriate for our to-be-developed mouse biometric system, and finalized application input appropriate for our stylometry and intruder operational behavior biometric systems. Note that finalized application-specific input is the input submitted to the application. For example, an email can be created by keying in text, pasting text from other sources, making corrections (deleting, backspacing, cutting, pasting, etc.) before finally sending the email, and by finalized application input we mean the email actually submitted though the email client. We consider it essential to gather the various user input data simultaneously because the various component biometric systems are to be combined later to provide a multi-level computational behavioral cognitive "fingerprint" of the person operating the computer, and the input data collected during this phase would not be useful to other phases unless collected simultaneously.

For the keystroke biometric work will leverage our previous work in developing the Pace University long-text-input keystroke biometric system. We are currently using a free open-source keylogger and converting its keystroke output into the format required for our Pace University long-text-input keystroke biometric system. The keylogger records the identity of the application "window" in use. A typical user will usually have several windows open – for example, an email client, a browser, a programming environment window, a command-line window, etc. Because our current keystroke biometric system was designed to handle only text input, we will extend it to handle input from the commonly used applications – text input (Microsoft Word, email, etc.), spreadsheet input (Microsoft Excel, etc.), web browser input (Microsoft Internet Explorer, etc.), command line input, as well as mixtures of arbitrary input from several applications. We will explore several ideas with regard to handling the various types of application input. For example, we could explore having individual keystroke biometric subsystems for each type of application input versus one system that handles all the types.

Primary organization responsible: The primary organization responsible for the execution of this research task area is Pace University. Dr. Tappert will be the key responsible team member and he will coordinate this effort with the co-investigators, Drs. Cha, Chen, and Grossman.

Exit criteria: The milestone that defines the completion of this task will be the final test and evaluation of the developed system that shows satisfactory biometric authentication performance – that is, acceptance of authorized users and rejection of unauthorized users (imposters).

Deliverables: The following deliverables will be provided to the Government in support of this task: monthly progress reports, quarterly presentations at a Government designated facility, final technical report, final presentation together with a demonstration of the biometric system, and all related keystroke biometric software.

Mouse Activity Biometric Research Task 2

Objective: The objective of the mouse activity biometric effort is to design, develop, test, and evaluate a mouse activity biometric system targeting DoD desktop/laptop computers.

Approach: This work will leverage our previous preliminary work in this area. As with the keystroke biometric system, the dichotomy vector-difference model and the kNN classifier that generates ROC curves will be used in the classification back-end. Reuse of these methods across the various biometric will greatly simplify the development and the later integration of the components.

Primary organization responsible: The primary organization responsible for the execution of this research task area is Pace University. Drs. Cha and Tappert will be the key responsible team members.

Exit criteria: The milestone that defines the completion of this task will be the final test and evaluation of the developed system that shows satisfactory biometric authentication performance – that is, acceptance of authorized users and rejection of unauthorized users (imposters).

Deliverables: The following deliverables will be provided to the Government in support of this task: monthly progress reports, quarterly presentations at a Government designated facility, final technical report, final presentation together with a demonstration of the biometric system, and all related mouse activity biometric software.

Stylometry Biometric Research Task 3

Objective: The objective of the stylometry biometric effort is to design, develop, test, and evaluate a stylometry biometric system targeting DoD desktop/laptop computers.

Approach: This work will leverage our previous preliminary work in this area. As with the keystroke biometric system, the dichotomy vector-difference model and the kNN classifier that generates ROC curves will be used in the classification back-end. Reuse of these methods across the various biometric will greatly simplify the development and the later integration of the components.

Primary organization responsible: The primary organization responsible for the execution of this research task area is Pace University. Drs. Chen and Tappert will be the key responsible team members.

Exit criteria: The milestone that defines the completion of this task will be the final test and evaluation of the developed system that shows satisfactory biometric authentication performance – that is, acceptance of authorized users and rejection of unauthorized users (imposters).

Deliverables: The following deliverables will be provided to the Government in support of this task: monthly progress reports, quarterly presentations at a Government designated facility, final technical report, final presentation together with a demonstration of the biometric system, and all related stylometry biometric software.

Operational Behavior Biometric Research Task 4

Objective: The objective of the intruder operational behavior biometric effort is to design, develop, test, and evaluate an intruder behavior biometric system targeting DoD desktop/laptop computers.

Approach: We have done very little prior work in this area. However, if possible, we will again use the dichotomy vector-difference model and the kNN classifier that generates ROC curves in the classification back-end. Otherwise, we will develop different methods and algorithms.

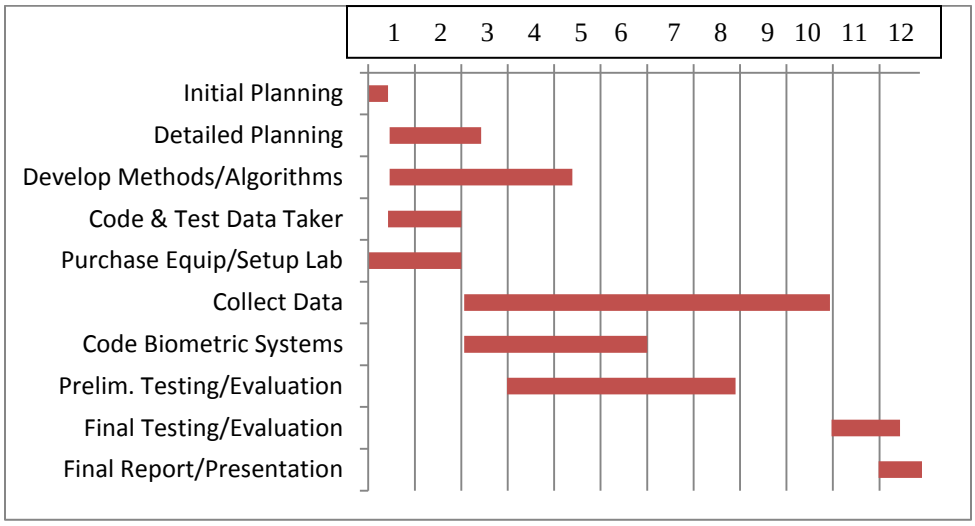
Primary organization responsible: The primary organization responsible for the execution of this research task area is Pace University. Drs. Grossman and Tappert will be the key responsible team members.

Exit criteria: The milestone that defines the completion of this task will be the final test and evaluation of the developed system that shows satisfactory biometric authentication performance – that is, acceptance of authorized users and rejection of unauthorized users (imposters).

Deliverables: The following deliverables will be provided to the Government in support of this task: monthly progress reports, quarterly presentations at a Government designated facility, final technical report, final presentation together with a demonstration of the biometric system, and all related intruder operational behavior biometric software.

Schedule and Milestones

The contract work schedule is shown in the following Gantt chart, months 1 through 12.



Milestones

End of month 2: completion of the biometric data collection system and the lab setup
End of month 6: completion of the coding of the biometric systems
End of month 6: preliminary performance results on the biometric systems
End of month 12: completion of contract period, final performance results and presentations

Cost Summary

Pace University Seidenberg School of Computer Science and Information Systems

DARPA-BAA-13-16: Active Authentication - Task Breakdown

Charles C. Tappert, Principal Investigator for Pace University

Task Breakdown	TOTAL YEAR
Keystroke Biometric Research - Task 1	193,543
Mouse Biometric Research - Task 2	98,433
Stylometry Biometric Research - Task 3	99,511
Operational Behavior Biometric Research - Task 4	97,018
	488,505

Task Breakdown by Category	
Total Salaries & Fringe Benefits	125,328
Staff Travel (Project related + conferences)	41,400
Programming and Testing Support	74,000
Supplies & Equipment: computer hardware & software	37,000
Student Tuition Remissions for 1 year:	147,832
Indirect @ 67.0% of direct salaries and wages (excluding all fringe benefits)	62,945
Total	488,505

Appendix A

(i) Team Member Identification – all are US citizens:

Individual Name	Role	Organization	Non-US	FFRDC/Govt
Dr. Charles C. Tappert (PI)	Prime	Pace University	no	no
Dr. Sung-Hyuk Cha (PI)	Prime	Pace University	no	no
Dr. Li-Chiou Chen (PI)	Prime	Pace University	no	no
Dr. Fred Grossman (PI)	Prime	Pace University	no	no

Dr. Charles C. Tappert (PI) has a Ph.D. in Electrical Engineering from Cornell University. He worked on speech and handwriting recognition at IBM for over two decades, secured and was the principal investigator on six government contracts in speech recognition, and holds nine patents. After IBM, he taught at the U.S. Military Academy at West Point for seven years and has been a professor of computer science at Pace University since 2000 where he is Associate Program Chair of the Doctor of Professional Studies in Computing. He has over 100 publications and his research interests include pattern recognition, biometrics, pen computing and voice applications, human-computer interaction, and artificial intelligence.

Dr. Sung-Hyuk Cha (PI) joined the Pace faculty in 2001 and is an Associate Professor of Computer Science with over 100 publications in pattern recognition and related fields. During his PhD years at the Center of Excellence for Document Analysis and Recognition (CEDAR) at SUNY Buffalo his major contributions included a dichotomy model to establish the individuality of handwriting, distance measures on histograms and strings, and a nearest neighbor search algorithm. He has over 100 publications and his main interests include biometrics, computer vision, data mining, and pattern matching and recognition. He is a member of AAAI, IEEE and its Computer Society, and IS&T.

Dr. Li-Chiou Chen (PI) has a Ph.D. in Engineering and Public Policy from Carnegie Mellon University (CMU) and was affiliated with the Center for Computational Analysis of Social and Organizational Systems (CASOS). While at CMU, she built a computational simulation tool to investigate policy decisions in countering distributed denial-of-service attacks. In addition, she participated in a DARPA-funded research project, BioWar, where she analyzed policy decisions to respond against the spread of biological agents. Her current research has been focused on the impact of policy and managerial decisions on countering computer attacks on the Internet. Specific topics include web application vulnerability testing, web security usability and security regulatory compliance. She has published papers in top academic journals, including Computer & Security, Decision Support Systems and IEEE Transactions. Li-Chiou has taught numerous courses on the theory and practice of information security and developed hands-on information security laboratory modules. She is also the principal investigator for the NSF Scholarship for Service Program in Information Assurance at Pace University.

Dr. Fred Grossman (PI) has been a professor of Computer Science and Information Systems for more than 30 years, and has been involved in software development for 40 years. He is a professor and Program Chair of the Doctor of Professional Studies in Computing. His principal research interests are in software engineering, agile methodologies and processes, automated

systems development, automated programming, very-high-level language design, and the integration of information systems and organization strategies. He has trained and coached agile teams in academic and industrial settings, and is a certified ScrumMaster. He has been active in the software industry as a founder of the New York Software Industry Association, has started several software companies, and has extensive consulting experience in computing and information technology. Dr. Grossman has a B.S. in Mathematics from Polytechnic University, a M.S. in Mathematics from New York University Courant Institute, and a Ph.D. in Computer Science from New York University Graduate School of Engineering.

(ii) Government or FFRDC Team Member Authority to Propose to this BAA: NONE

(iii) Government or FFRDC Team Member Statement of Unique Capability: NONE

(iv) Organizational Conflict of Interest Affirmations and Disclosure: NONE

(v) Intellectual Property: NONE

(vi) Human Use: It is expected that this research will be an exempt research project from IRB review in accordance with the Common Rule subsection 101.b (2) "Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior." We will submit applications of exempt proposal to Internal Review Board (IRB) at Pace University.

(vi) Animal Use: NONE

(viii) Representations Regarding Unpaid Delinquent Tax Liability or a Felony Conviction under Any Federal Law:

(1) The proposer represents that it is [ ] is not [X] a corporation that has any unpaid Federal tax liability that has been assessed, for which all judicial and administrative remedies have been exhausted or have lapsed, and that is not being paid in a timely manner pursuant to an agreement with the authority responsible for collecting the tax liability.

(2) The proposer represents that it is [ ] is not [X] a corporation that was convicted of a felony criminal violated under Federal law within the preceding 24 months.

(ix) Cost Accounting Standards (CAS) Notices and Certification: NONE

(x) Subcontractor Plan: NONE

Appendix B

Literature Context: Non-Pace-University Keystroke Biometric Publications Cited Above

Bender, S.S. & Postley, H.J. (2007). Key sequence rhythm recognition system and method. U.S. Patent 7,206,938.

Bergadano, F., Gunetti, D., & Picardi, C. (2002). User authentication through keystroke dynamics. *ACM Trans. Information & System Security, 5(4),* 367-397.

Bolle, R., Connell, J., Pankanti, S., Ratha, N., & Senior, A. (2004). *Guide to biometrics*. New York: Springer.

Giot, R., El-Abed, M., & Rosenberger, C. (2009). Keystroke dynamics with low constraints svm based passphrase enrollment. *IEEE Int. Conf. Biometrics: Theory, Applications, and Systems (BTAS 2009)*.

Gunetti, D. & Picardi, C. (2005). Keystroke analysis of free text. *ACM Trans. Information & System Security, 8(3),* 312-347.

Jin, L., Ke, X., Manuel, R., & Wilkerson, M. (2004). Keystroke dynamics: A software based biometric solution. *Proc. 13th USENIX Security Symposium*.

Karnan, M., Akila, M., & Krishnaraj, N. (2011). Biometric personal authentication using keystroke dynamics: A review. *Applied Soft Computing Journal, 11(2),* Elsevier.

Killourhy, K.S. & Maxion, R.A (2009). Comparing Anomaly-Detection Algorithms for Keystroke Dynamics. *Int. Conf. Dependable Systems & Networks (DSN-09),* 125-134, Lisbon, Portugal, June 2009. IEEE Computer Society Press, Los Alamitos, California, 2009.

Leggett, J., Williams, G., Usnick, M., & Longnecker, M. (1991). Dynamic identity verification via keystroke characteristics. *Int. J. Man Machine Studies, 35(6),* 859-870.

Li, Y., Zhang, B., Cao, Y., Zhao, S., Gao, Y. & Liu, J. (2011). Study on the BeiHang Keystroke Dynamics Database. Proc. *Int. Joint Conf. Biometrics (IJCB 2011),* Wash. D.C., October 2011.

Messerman, A., Mustafi´c, T., Camtepe, S.A. & Albayrak, S. (2011). Continuous and Non-intrusive Identity Verification in Real-time Environments based on Free-Text Keystroke Dynamics. Proc. *Int. Joint Conf. Biometrics (IJCB 2011),* Wash. D.C., October 2011.

Monrose, F., Reiter, M.K., & Wetzel, S. (2002). Password hardening based on keystroke dynamics. *Int. J. Information Security,* 1(2), 69-83.

Monrose, F. & Rubin, A.D. (2000). Keystroke dynamics as a biometric for authentication. *Future Generation Computer Systems, 16(4),* 351-359.

Obaidat, M.S. & Sadoun, B. (1999). Keystroke dynamics based authentication. In *Biometrics: Personal Identification in Networked Society* by Jain, et al. (New York: Springer), 213-230.

Peacock, A., Ke, X., & Wilkerson, M. (2004). Typing patterns: A key to user identification. *IEEE Security & Privacy, 2(5),* 40-47.

Revett, K. (2008). Chapter 4: Keystroke dynamics, pp. 73-136. *Behavioral biometrics: A remote access approach*, Wiley.

Rodrigues, R.N., Yared, G.F.G., Costa, C.R., Yabu-Uti, J.B.T., Violaro, F., & Ling, L.L. (2006). Biometric access control through numerical keyboards based on keystroke dynamics. Lecture notes in computer science, 3832, pp. 640-646.

Woodward, J.D. Jr., Orlans, N.M., & Higgins, P.T. (2002). *Biometrics*, NY: McGraw-Hill, 107.

Yu, E. & Cho, S. (2004). Keystroke dynamics identity verification – Its problems and practical solutions. *Computers & Security, 23(5)*, 428-440.

Related Pace University External Publications Cited Above (in chronological order)

Cha, S. & Srihari, S.N. (2000). Writer Identification: Statistical Analysis and Dichotomizer. *Proc. SPR and SSPR 2000, LNCS - Advances in Pattern Recognition, v. 1876*, 123-132.

Yoon, S., Choi, S-S., Cha, S-H., Lee, Y., & Tappert, C.C. (2005). On the individuality of the iris biometric. *Proc. Int. J. Graphics, Vision & Image Processing, 5(5)*, 63-70.

Curtin, M. (2006). Long-Text Keystroke Biometric Applications over the Internet. Doctoral dissertation, Pace University, New York.

Villani, M., Tappert, C., Ngo, G., Simone, J., St. Fort, H., & Cha, H-S. (2006). Keystroke biometric recognition studies on long-text input under ideal and application-oriented conditions. *Proc. Computer Vision & Pattern Recognition Workshop on Biometrics,* New York.

Villani, M. (2006). Keystroke Biometric Identification Studies on Long Text Input. Doctoral dissertation, Pace University, New York.

Ritzmann, M. (2007). Strategies for Managing Missing and Incomplete Information with Applications to Keystroke Biometric Data and a Business Analytical Application. Doctoral dissertation, Pace University, New York.

Tappert, C.C, Villani, M., & Cha, S. (2010). Chapter 16, Keystroke Biometric Identification and Authentication on Long-Text Input, pp. 342-367. *Behavioral Biometrics for Human Identification: Intelligent Applications*, Ed. Liang Wang & Xin Geng. IGI Global, 2010.

Zack, R.S. (2010). An Improved k-NN Classification Method with Application to Keystroke Biometric Authentication. Doctoral dissertation, Pace University, New York.

R.S. Zack, C.C. Tappert & S.-H. Cha (2010). Performance of a Long-Text-Input Keystroke Biometric Authentication System Using an Improved k-Nearest-Neighbor Classification Method. Proc. *IEEE 4th Int Conf Biometrics: Theory, Apps, and Systems (BTAS 2010)*, Wash. D.C.

Tappert, C.C, Cha, S.-H., Villani, M., & Zack, R.S. (2010). A Keystroke Biometric System for Long-Text Input, invited paper, *Int. J. Information Security and Privacy (IJISP)*, Vol 4, No 1, 2010, pp 32-60, http://www.csis.pace.edu/~ctappert/dps/d891b-12/keystroke-tappert-IJISP.pdf .

J.C. Stewart, J.V. Monaco, S. Cha, and C.C. Tappert (2011). An Investigation of Keystroke and Stylometry Traits. Proc. *Int. Joint Conf. Biometrics (IJCB 2011)*, Wash. D.C., October 2011.

Stewart, J.C. (2012). An Evaluation of Stylometry and Keystroke Biometrics in the Identity Verification of Online Test-Takers. Doctoral dissertation, Pace University, New York.

Monaco, J.V., Bakelman, N., Cha, S. and Tappert, C.C. (2012). "Developing a Keystroke Biometric System for Continual Authentication of Computer Users," *Proc. 2012 European Intelligence and Security Informatics Conference*, Denmark, August 2012.

Monaco, J.V., Bakelman, N., Cha, S. and Tappert, C.C. (2013, submitted to conference). "Recent Advances in the Development of a Keystroke Biometric Authentication System for Long-Text Input"

Bakelman, N. (in progress). Development and Evaluation of a Keystroke Biometric Authentication System for Text, Spreadsheet, Browser, and Number Pad Input. Doctoral dissertation, Pace University, New York.

Monaco, J.V. (in progress). Development and Evaluation of a Dichotomy-Model Authentication System for Keystroke, Mouse, and Stylometry Biometrics. M.S. Computer Science Thesis, Pace University, New York.

Related Pace University Internal Publications (in chronological order, with links to PDF)

G. Bartolacci, M. Curtin, M. Katzenberg, N. Nwana, S. Cha & C. Tappert (2005). Long-Text Keystroke Biometric Applications over the Internet. *Proc. Research Day*, Pace Univ., 2005.

G. Ngo, J. Simone, & H. St. Fort (2006). Developing a Java-Based Keystroke Biometric System for Long-Text Input, *Proc. Research Day*, Pace Univ., 2006.

M. Curtin , C. Tappert, M. Villani, G. Ngo, J. Simone, H. St. Fort, and S.-H. Cha (2006). Keystroke Biometric Recognition on Long-Text Input: A Feasibility Study, *Proc. Research Day*, Pace Univ., 2006.

M. Villani, C. Tappert, G. Ngo, J. Simone, H. St. Fort, & S. Cha (2006). Keystroke Biometric Recognition Studies on Long-Text Input under Ideal and Application-Oriented Conditions. *Proc. Research Day*, Pace Univ., 2006.

M. Ritzmann & L. Weinrich (2007). Strategies for Managing Missing or Incomplete Data in Biometric and Business Applications. *Proc. Research Day*, Pace Univ., 2007.

R. Goodman, M. Hahn, M. Marella, C. Ojar, & S. Westcott (2007). The Use of Stylometry for Email Author Identification: A Feasibility Study. *Proc. Research Day*, Pace Univ., 2007.

A. Weiss, A. Ramapanicker, P. Shah, S. Noble, & L. Immohr (2007). Mouse Movements Biometric Identification: A Feasibility Study. *Proc. Research Day*, Pace Univ., 2007.

C. Eusebi, C. Gliga, D. John, & A. Maisonave (2008). A Data Mining Study of Mouse Movement, Stylometry, and Keystroke Biometric Data. *Proc. Research Day*, Pace Univ., 2008.

N. Ajufor, A. Amalraj, R. Diaz, M. Islam, & M. Lampe (2008). Refinement of a Mouse Movement Biometric System. *Proc. Research Day*, Pace Univ., 2008.

K. Calix, M. Connors, D. Levy, H. Manzar, G. McCabe, & S. Westcott (2008). Stylometry for E-Mail Author Identification and Authentication. *Proc. Research Day*, Pace Univ., 2008.

S. Bharati, R. Haseem, R. Khan, M. Ritzmann, & A. Wong (2008). Biometric Authentication System Using the Dichotomy Model. *Proc. Research Day*, Pace Univ., 2008.

T. Buch, A. Cotoranu, E. Jeskey, F. Tihon, & M. Villani (2008). An Enhanced Keystroke Biometric System and Associated Studies. *Proc. Research Day*, Pace Univ., 2008.

E. Wood, J. Zelaya, E. Saari, K. King, M. Gupta, N. Howard, S. Ismat, M.A. Kane, M. Naumowicz, D. Varela, & M. Villani (2008). Longitudinal Keystroke Biometric Studies on Long-Text Input. *Proc. Research Day*, Pace Univ., 2008.

M. Wuench, M. Bi, E. Urbaez, S.M. Varghese, M. Tevnan, M. Villani, & C. Tappert (2009). Keystroke Biometric Test-Taker Authentication System. *Proc. Research Day*, Pace Univ., 2009.

A. Amatya, J. Aliperti, T. Mariutto, A. Shah, M. Warren, R. Zack, & C. Tappert (2009). Keystroke Biometric Authentication System Experimentation. *Proc. Research Day*, Pace Univ., 2009.

G. Shalhoub, R. Simon, R. Iyer, J. Tailor, & S. Westcott (2010). Stylometry System - Use Cases and Feasibility Study. *Proc. Research Day*, Pace Univ., 2010.

S. Janapala, S. Roy, J. John, L. Columbu, J. Carrozza, R. Zack, & C. Tappert (2010). Refactoring a Keystroke Biometric System. *Proc. Research Day*, Pace Univ., 2010.

K. Doller, S. Chebiyam, S. Ranjan, E. Little-Torres, & R. Zack (2010). Keystroke Biometric System Test Taker Setup and Data Collection. *Proc. Research Day*, Pace Univ., 2010.

M. Lam, U. Patel, M. Schepp, T. Taylor, & R. Zack (2010). Keystroke Biometric: Data Capture Resolution Accuracy. *Proc. Research Day*, Pace Univ., 2010.

R. Zack, A. Kanchan, P. Ranadive, S. Desai, P. Mahotra, N. Wang, & C. Tappert (2010). Keystroke Biometric: Receiver Operating Characteristics (ROC) Experiments. *Proc. Research Day*, Pace Univ., 2010.

A.C. Caicedo, K. Chan, D.A. Germosen, S. Indukuri, M.N. Malik, D. Tulasi, M.C. Wagner, R.S. Zack, & C. Tappert (2010). Keystroke Biometric: Data/Feature Experiments. *Proc. Research Day*, Pace Univ., 2010.

J.C. Stewart & K.M. Anne (2010). The Scalability of Keystroke Biometrics in Access Control and Identity Management Systems. *Proc. Research Day*, Pace Univ., 2010.

J. Deluca, D.R. Worley, H. Henry, P. Folkes, & N. Bakelman (2011). A System-wide Keystroke Biometric System. *Proc. Research Day*, Pace Univ., 2011.

V. Monaco, E. Zych, O. Canales, T. Murphy, J. Stewart, C. Tappert, A. Castro, O. Sotoye, L. Torres, & G. Truley (2011). A Stylometry System for Authenticating Students Taking Online Tests. *Proc. Research Day*, Pace Univ., 2011.

V. Monaco, E. Zych, T. Allman, M. Lamrabat, M. Manohar, J. Stewart, C. Tappert, H. Poorshatery, G. Garcia, E. Teracino, X. Zhao, & R. Zack (2011). A Keystroke Biometric System Test Taker Setup and Data Collection. *Proc. Research Day*, Pace Univ., 2011.

N. Bakelman, J.V. Monaco, S. Cha, & C.C. Tappert (2012). Continual Keystroke Biometric Authentication on Short Bursts of Keyboard Input. *Proc. Research Day*, Pace Univ., 2012.

C. Funk, S. Hanchar, & N. Bakelman (2012). Analysis of the Fimbel Keylogger and Pace University Converter. *Proc. Research Day*, Pace Univ., 2012.

B. Tschinkel, B.d Esantsi, D.k Iacovelli, P. Nagesar, R. Walz, M. Guglielmo, A. Weisman, E. Prekelezaj, D. Camilo, V. Monaco, & N. Bakelman (2012). Keystroke Biometric Intrusion Detection. *Proc. Research Day*, Pace Univ., 2012.

I. Schulstad, M. Boga, C. Jordan, K. Pally, J. Monaco, R. DeStefano, J. Stewart, & C. Tappert (2012). Evaluation of a Stylometry System on Various Length Portions of Books. *Proc. Research Day*, Pace Univ., 2012.