###### Possible Security related Projects for Undergraduates

###### Net Trust

NetTrust (Tsow, Viecco et al. 2010) <http://www.ljean.com/NetTrust/> is a system that informs individual browsing and information-sharing decisions by leveraging first, second and third party information. Net Trust uses first person browsing history to create implicit ratings as well as enabling explicit ratings and comments. NetTrust similarly provides information from a user-selected social network by sharing ratings from browsing histories and annotations. This second person information is similar to social browsing. Net Trust integrates third party ratings into the display these individual ratings. NetTrust allows an individual to select their own trusted authoritative sources of information from a market of third party ratings agencies.

This project has already involved undergraduates. First, a group of undergraduate students developed an ambient interaction (rather than a toolbar) suitable for elders. Secondly, one of these students worked to implement a usability study with elders with the ambient interaction.

Related researchers: L Jean Camp

###### Health Security Informatics

The health arena has long been a domain for fraud, even the phrase “snake oil” in cryptography comes from health. Because Medicare payments are required to be processed so quickly, Harvard Professor Malcolm Sparrow uses statistical analysis of patient and provider patterns to identify millions in fraud. This is an example of how statistical methods (for example used in health and psychology) could be applied to computer security and electronic fraud prevention. Medical information is subject to phishing as well.

Medical identity theft is the fastest growing dimension of identity theft. Medical Identity Theft occurs when a thief uses stolen personal information in order to receive medical care under an assumed name. According to the World Privacy Forum in 2009, half a million Americans had been victims of medical identity theft. Certainly securing online payments, preventing fraud, and preventing medical identity theft will be areas of research for undergraduates, as shown in the Net Trust case.

Related researchers: L Jean Camp, XiaoFeng Wang

###### Genome Privacy

With rapid advancement in genome sequencing technologies,human genomic data has been increasingly collected and disseminated to facilitate scientific research. This is leading to technological revolutions in many critical domains, particularly biomedical science and engineering. Of great importance to these studies is protection of participants’ genetic information, which, once leaked to unauthorized parties, could have a disastrous consequence. Unfortunately, to date, only minimum effort has been made to investigate the privacy risks and related ethic issues involved in human genome study, which offers little support for effectively protecting the participants of these studies. We are working on multiple projects in this critical direction (R. Wang, X. Wang et al. 2009; R. Wang, Y. Li et al. 2009). As an example, we have recently discovered that the outcomes released by genome-wide association studies (GWAS) could be used to identify some participants of the studies. For another example, we are applying program specialization techniques to automatically partition a genome-computing program into the part that processes sensitive data, and the part that only uses public information. The latter can be delegated to the client to save the computing resources of the server trusted with sensitive genome data. Another under-estimated privacy risk is related to the researcher that conducted the genome experiment and generated the raw data. Their genetic materials (e.g. DNA) may become contaminants in the biological samples and thus enter into public data (e.g. (Helena Malmström, Jan Storå et al. 2005) ). As a result, an attacker can retrieve human-originated DNA sequences from a non-human research project, and deduce the genetic information of the researchers. We will study the methods towards the design of an efficient cyber-infrastructure to assess the information leaks in the raw genomic data and to filter human-originated information before the raw data is dissimilated.

Some of these research projects have already involved undergraduates. For example, one student with both biology and security backgrounds begins participating in the GWAS privacy project.

Related researchers: XiaoFeng Wang, Haixu Tang, Geoffrey Fox

###### Wireless Security @Home

Integrated multidisciplinary research in wireless security includes three approaches. From a complex system standpoint, urban areas with overlapping home-based wireless networks create a highly interconnected but unorganized ad-hoc network infrastructure. Myers and Vespignanni illustrate that the large scale implications of local insecure router installation leads to potential digital pandemics that are entirely wirelessly transmitted, i.e. without utilizing the Internet routing plane (Hao Hua, Steven Myers et al. 2009). Using econometric models, Camp lead a an examination of adoption of encryption for wireless home networks using economic theory to test predictions that increases in density (e.g., risk), income, and education would predict rational homeowners to secure routers. The resulting thorough experimental rejection of these hypotheses were both published at Cambridge, UK and widely blogged, as it examined the choice of home users to secure their wireless networks (Matthew Hottell, Drew Carter et al. 2006). Currently the experiment continues with a qualitative evaluation of changes over time with interviews of individuals in different categories by income, education, density and wireless configuration. (This is work in progress.)

These experiments have included three faculty, one post-doc, two doctoral students, two masters students, and three undergraduates.

Related researchers: Steven Myers, L Jean Camp

###### Security and Privacy in Cloud and Web

With computation moving onto the world-wide web, web security becomes increasingly important. Mitigation of the threats such as cross-site scripting, cross-site request forgery requires effective technologies for analyzing and controlling scripts. Also of great importance is the security and privacy issues related to cloud computing, particularly, the security weaknesses in computing model such as software as a service (SaaS), which have been intensively studied by the IU system security research group (<http://sysseclab.informatics.indiana.edu>).

Research on system and web security at the School of Informatics and Computing (C. Kolbitsch, P. Milani et al. 2009; K. Zhang and X. Wang 2009)has traditionally involved undergraduates. As an example, Wang regularly teaches “I-430: Security for Networked Systems”, which includes the course projects related to his research on system security.

Related researchers: XiaoFeng Wang, Geoffrey Fox

###### Smartphone Security

The move towards using popularly smartphones and PDAs to access the Internet and perform a number of personal computations has opened the door to a fleet of opportunities and new security problems (A. Kapadia and N. Triandopoulos 2008). In particular, these devices are platforms that contain a large number of sensors, the results of which from a large number of people can be combined to obtain a sensor network. However, even if the users of these devices can be trusted, the results from the sensors are not necessarily trustworthy. In particular, it is possible the devices have been absconded by an adversary who will affect the devices sensors by either changing the device's environment, or its programming. The use of traditional authentication technologies to ensure a legitimate user is in control of the smartphone is not practical, as said users cannot be queried to authenticate every time the sensor-net requests information. We are working on multiple projects in the area. An example is to enable a phone attempts to determine if a phone is no longer in the possession of its user, and when this happens, deauthenticate the phone (essentially removing it from the sensor network, or tagging its information to note potential risk in including it). This project uses machine-learning techniques and the sensors of the smartphone to estimate the likelihood that the legitimate user is in possession of the phone. In particular, the phone learns normal behaviors of the phone's sensors when it's knowingly possessed by the legitimate user, and uses that information to estimate the likelihood that it is still in the user's possession based on current sensor readings.

There will be a number of places where undergraduate students can partake in research on this project and interact with PhD and MSc students who are currently working on it. Such work would include implementing the programming on smartphones for reading the sensor data in novel ways and developing or implementing appropriate learning protocols specific to the problems; these will be based on standard learning techniques, such as Hidden Markov Models, Support Vector Machines and Neural Networks, which are within the grasp of undergraduate students.

Related researchers: Steven Myers, Apu Kapadia, XiaoFeng Wang, Geoffrey Fox

###### Risk Communication

Networked computers enable us to complete tasks remotely, instantly, and without direct interaction with other humans. As a result, the context that would normally assist us in evaluating risks (such as packaging, geography, social context, and temporal clues) is absent. In addition, mechanisms designed to counter specific, technical risks have demonstrably failed to provide useful contextual guidance, as proven by the prevalence of spam, malware, botnets, and phishing. The articulation of risks by those with the greatest security expertise has proven sorely inadequate in real time. In part, this is because networked computers (unlike toxic waste, for example) are not inherently dangerous. Computers do not themselves pose risks. Using networked computers to their fullest requires people to make choices that can create hazards to personal, psychological, and financial security. And an increasing number of computer users have little to no technical expertise to understand the existence and nature of these hazards. Examples of chosen, potentially hazardous behaviors include sharing personal information, banking on-line, sharing social network information, and (mis)managing passwords.

Our research project evaluates on-line human risk behaviors according to the guidance provided by classic works in offline risk behavior. In order to do this, the students select previously identified risks and evaluate the increase or mitigation of risk possible with trusted hardware. From that analyses and experiments, we will provide research experiences that offer not only the creation of experimentally proven risk communication for specific threats but also a broader set of heuristics to guide development of future systems. This research can be integrated with trusted computer protocols and architectural proposals that align with human risk behaviors for the more advanced technical students. Ideally, the result of our research will enable individuals to distinguish risky and non-risky behaviors in a way aligned with actual computing risk.

Related researchers: Apu Kapadia, L Jean Camp

###### Elder and Caregiver Support

The Ethical Technology for the Homes of Senior Project (described below in Facilities) takes a holistic view towards personal and data risks. Examples of prototypes are available at <http://ethos.indiana.edu>. The Portal Monitor is an example of a privacy-aware risk-reducing design as created at ETHOS (L. Jean Camp and Kay Connelly 2007). The Portal Monitor consists of three cameras: one inside facing the back of the closed door; one with a fisheye view out of door 5’5” off the ground; and one facing inward at the front of the closed door. The cameras are triggered either by the doorbell or by the opening of the door. Upon activation the Monitor cameras take three pictures, 2 seconds apart, and sends this via the cellular network to one or more cell phones selected by the elder (e.g., to caregivers). The real time event-driven photos is a design that addresses the privacy risks of currently marketed systems, which offer unencrypted video sent over the network real time. The Portal Monitor minimizes data collection, by being alert-driven. There is no central mechanism for collection of data. There is no secondary collection of data for marketing. The Portal Monitor utilizes the cellular network to decrease the expectation of wiretapping which is more common on the Internet as packet sniffing. The Portal Monitor addresses the threat to safety without creating additional risks to privacy or exposing home video on the network. All the components to implement a broader examination of the Portal Monitor, either in homes or in a more public setting exist.

ETHOS has hosted 11 summer undergraduate researchers in its first year and 12 in its second. This has been supported with a combination of REU supplements, Computing Research Association support, School support, and University undergraduate research programs. Two of these undergraduates have applied to graduate programs in informatics. Undergraduate projects have included prototype development, usability testing, threat modeling, adversarial modeling, and privacy analyses of both deployed systems and prototypes. In the case of the Portal Monitor students would install prepared prototypes, with research options being red-teaming attacks or in situ user evaluations.

Related researchers: L Jean Camp, Kay Connelly

###### Improving Safety of Web Browsing

Internet fraudsters are making extensive use of the Web in luring users to websites connected to phishing, malware, and, scam (R. Rasmussen and G. Aaron 2009). A popular technique for defending end users against such malicious websites is blacklisting. A typical blacklist today contains the URLs and IP addresses belonging to machines that host malicious websites. Web browsers and email programs make extensive use of the blacklists. Unfortunately, the current blacklist-generation process requires human intervention, which delays the blocking of malicious sites. Further, blacklists are losing ground as attackers are increasingly exploiting newer techniques to evade them. In our research group, we are exploring novel ways in which malicious websites can be automatically identified and put into blacklists. This will not only proactively identify malicious websites, it will also automate the process of blacklist generation.

The project is aimed at undergraduate students who will be working with one or more senior graduate students. Upon completion of the project, the students would have improved their programming skills. They will enhance their knowledge of the Web and the various security issues it faces. Paper reading and writing skills will be developed as students will be contributing to the writing of the technical paper as co-authors.

Related researchers: Minaxi Gupta

###### Privacy in Peer-to-Peer and Social Networks

We are currently investigating several applications of large-scale peer-to-peer networks of mobile devices such as smartphones. These applications include fully distributed social networking, structured and unstructured anonymous routing networks, and sensor data sharing (C. Cornelius, A. Kapadia et al. 2008) for health-monitoring applications. We will also explore the use of redundant routing in peer-to-peer networks to evade censorship attacks (A. Kapadia and N. Triandopoulos 2008) and in particular improve the privacy of searching for resources in such networks.

Undergraduates will assist in developing and implementing simulations in our supercomputing facilities, as well as in building research prototypes on smartphones such as the Apple iPhone and other smartphones supporting the Google Android platform.

Related researcher: Apu Kapadia

**Privacy Enhanced Online Human Subjects Data Collection**

In this electronic age, the collection of self-report data in the social sciences is increasingly moving from paper-and-pencil to online survey methods. This move to electronic media poses a number of novel challenges, in terms of data and privacy protection, for researchers and institutional review boards (IRBs) alike. While protecting paper-and-pencil surveys once may have been as simple as keeping the data locked away and limiting access to the key, these physical mechanisms do not satisfy the requirements for protecting electronic data. The problem of protecting electronic data is further complicated by the use of mobile devices, portable storage media and networked storage devices. The protection of electronic data sources becomes even more difficult when researchers who are responsible for collecting the data are not familiar with the vulnerabilities and threats to the systems that host the data.

This project proposes to design, develop and test a privacy enhancing online survey system prototype, using sexuality research as a model. The goals of the online survey system are to protect privacy and ensure confidentiality, while incorporating the following features:

1. The website should allow participants to return if they prefer to not finish all questionnaires in one session, and for this purpose should be able to identify participants and keep track of their previous responses;
2. The website should allow for longitudinal data collection and include the option to automatically notify participants when it is time to return to the website to complete questionnaires again;
3. The website should incorporate mechanisms to allow for the provision of feedback to participants based on their answers to certain questionnaires, feedback which could be based on previously published data as well as on the performance of others who completed the questionnaires;
4. The website should have the ability to link a participant’s data to the data collected from others (e.g., their partner(s) or other participants, obtained through respondent-driven recruitment--see below).

Most of the above features introduce the need for linking data without compromising identities. Therefore we propose to use pseudonyms to de-identify participants. We plan to explore a variety of techniques that have been proposed by previous research to determine which would be most appropriate for online surveys. These include but are not limited to group based credentials, traceable credentials and credentials that have blacklisting capabilities. We do not propose to create new pseudonym techniques, but to evaluate previously proposed techniques for application to our problem space (Giuseppe Ateniese, Jan Camenisch et al. 2000; Mihir Bellare, Daniele Micciancio et al. 2003; Mihir Bellare, Haixia Shi et al. 2005).

Related researcher: Raquel Hill

**References**

A. Kapadia and N. Triandopoulos (2008). Halo: High-Assurance Locate for Distributed Hash Tables. Proceedings of the 15th Annual Network & Distributed System Security Symposium (NDSS '08). San Diego, CA**:** 61-79.

C. Cornelius, A. Kapadia, et al. (2008). AnonySense: Privacy-Aware People-Centric Sensing. Proceedings of The 6th ACM/USENIX International Conference on Mobile Systems, Applications, and Services (MOBISYS '08). Breckenridge, Colorado**:** 211–224.

C. Kolbitsch, P. Milani, et al. (2009). Effective and Efficient Malware Detection at the End Host. Proceedings of the USENIX Security Symposium.

Giuseppe Ateniese, Jan Camenisch, et al. (2000). A practical and provably secure coalition-resistant group signature scheme. CRYPTO, Springer LNCS.

Hao Hua, Steven Myers, et al. (2009). "WiFi networks and malware epidemiology." PNAS **106**(5): 1318-1323.

Helena Malmström, Jan Storå, et al. (2005). "Extensive Human DNA Contamination in Extracts from Ancient Dog Bones and Teeth." Mol Biol Evol **22**: 2040-2047.

K. Zhang and X. Wang (2009). Peeping Tom in the Neighborhood: Keystroke Eavesdropping on Multi-user Systems <http://www.informatics.indiana.edu/xw7/papers/keystroke.pdf>. Proceedings of the USENIX Security Symposium.

L. Jean Camp and Kay Connelly (2007). Privacy in Ubicomp. Digital Privacy: Theory, Technologies and Practices. Alessandro Acquisti, Sabrina De Capitani di Vimercati, Stefanos Gritzalis and Costas Lambrinoudakis. New York NY, Taylor & Frances.

Matthew Hottell, Drew Carter, et al. (2006). Predictors of Home-Based Wireless Security <http://weis2006.econinfosec.org/docs/51.pdf>. Fifth Workshop on the Economics of Information Security. Cambridge, UK.

Mihir Bellare, Daniele Micciancio, et al. (2003). Foundations of group signatures: Formal definitions, simplified requirements, and a construction based on general assumptions. EUROCRYPT. E. Biham, Springer. **2656:** 614-629.

Mihir Bellare, Haixia Shi, et al. (2005). Foundations of group signatures: The case of dynamic groups. CT-RSA. A. Menezes, Springer. **3376:** 136-153.

R. Rasmussen and G. Aaron, An Anti-Phishing Working Group (APWG) Industry Advisory, October 2009. (2009). "Global Phishing Survey: Trends and Domain Name Use 1H2009." Retrieved January, 2010, from <http://www.apwg.org/reports/APWG_GlobalPhishingSurvey_1H2009.pdf>.

R. Wang, X. Wang, et al. (2009). Privacy-Preserving Genomic Computation Through Program Specialization. 16th ACM Conference on Computer and Communications Security (CCS).

R. Wang, Y. Li, et al. (2009). Learning Your Identity and Disease from Research Papers: Information Leaks in Genome Wide Association Study. 16th ACM Conference on Computer and Communications Security (CCS).

Tsow, A., C. Viecco, et al. (2010). "Privacy-Aware Architecture for Sharing Web Histories." IBM Systems Journal.