SYMPOSIUM ON USABLE PRIVACY AND SECURITY CONFERENCE REPORT

edited by Fahd Arshad and Rob Reeder

Carnegie Mellon University, Pittsburgh, PA, USA. July 2005

TABLE OF CONTENTS

Introduction - SOUPS 2005 3
Tutorials 3
Introduction to Computer Security and Privacy 3
User Interface Design, Prototyping, and Evaluation 4
Opening Session 5
Welcome and Opening Remarks, Best Paper Award 5
Invited Talk 5
Refereed Paper Sessions 6
Session I: Usable Security 6
Authentication Using Graphical Passwords: Effects of Tolerance and Image Choice 6
Johnny 2: A User Test of Key Continuity Management with S/MIME and Outlook Express 6
Two Experiences Designing for Effective Security 7
Session II: "Usable Privacy" 7
Usable Security and Privacy: A Case Study of Developing Privacy Management Tools 7
Stopping Spyware at the Gate: A User Study of Privacy, Notice and Spyware 7
Making PRIME usable 7
Developing Privacy Guidelines for Social Location Disclosure Applications and Services 7
Session III: "Visualizing Security" 8
The Battle Against Phishing: Dynamic Security Skins 8
Attacking Information Visualization System Usability Overloading and Deceiving the Human 8
Social Navigation as a Model for Usable Security 8
Poster Session 8
Panels 9
Usability of Security Administration vs. Usability of End-user Security 9
When User Studies Attack: Evaluating Security by Intentionally Attacking Users 10
Discussion Sessions 11
When User Studies Attack: Evaluating Security by Intentionally Attacking Users 11
Usability and Acceptance of Biometrics 11
Valuation and Context 12
Usable Interfaces for Anonymous Communication 12
Commentary on SOUPS 2005 12
Threats Addressed at SOUPS 12
Evaluation: A Big Issue 12
Bringing HCI and Security Together 13
Who's Looking Out for the Administrators? 13

INTRODUCTION - SOUPS 2005

The first Symposium On Usable Privacy and Security (SOUPS 2005) took place July 6-8, 2005 on the Carnegie Mellon University campus in Pittsburgh, PA. The symposium was organized by CMU's CyLab and the CMU Usable Privacy and Security (CUPS) Lab, and was chaired by CMU's Lorrie Faith Cranor. SOUPS 2005 brought together academic and industry researchers from the HCISEC (human-computer interaction and security) field with a variety of backgrounds, including computer security, privacy, human-computer interaction, cognitive science, and public policy. The symposium was a response to a need, voiced at the 2004 DIMACS Workshop on Usable Privacy and Security Software held at Rutgers University, for a refereed forum for researchers in HCISEC to present their work.

This document summarizes all of the SOUPS 2005 tutorial, opening, paper, panel, and discussion sessions. The end of the document contains a brief commentary on some of the themes and issues that were addressed or raised by SOUPS 2005.

The editors would like to thank student members of the CUPS Lab who contributed summaries of SOUPS sessions to this report: Steve Sheng, Elaine Newton, Ponnurangam Kumaraguru, and Serge Egelman. Thanks also to Cynthia Kuo for her detailed notes on the various SOUPS sessions.

Fahd Arshad <fahd@cmu.edu>
Rob Reeder <reeder@cs.cmu.edu>
Carnegie Mellon University

TUTORIALS

INTRODUCTION TO COMPUTER SECURITY AND PRIVACY

Instructor: Simson Garfinkel, MIT

Summary contributed by Fahd Arshad

Simson Garfinkel, ex-journalist and current security expert and privacy advocate, gave a three-part tutorial on Computer Security and Privacy. This was meant to be a primer for attendees who were not experts in computer security and would benefit from an overview of the basic principles involved.

Garfinkel started with defining security in terms of availability, confidentiality, data integrity, control, and audit. Different environments have different priorities, so in some, the confidentiality or integrity of the data may be the most important aspects, whereas in others, availability may trump other considerations such as confidentiality. Hence whenever we ask whether data is secure, it must be in some context. Security policies must be constructed to fit needs within this context. Choosing "Best Practice" policy templates only obfuscates the particular security concerns of each organization, according to Garfinkel. Furthermore, those who have the responsibility for creating the policies must also have the power to enforce them. Finally, it must be realized that risk can not be removed entirely, but good security policies, based on properties such as fail-safe defaults, separation of privilege, open design, etc. can be an effective safeguard.

Next, Garfinkel tackled privacy, the definition of which, like security, is contextual. Privacy can mean freedom from intrusion, control of person information, or misappropriation of one's name or image. The threat to privacy may come from the government, businesses, or the media. Garfinkel chronicled the historical threat to personal privacy that came with the rising use of photography and an aggressive press corps in the United States at the beginning of the 20th century. Credit reporting was born through the need to extend credit to anonymous customers around the same time. It was not until 1970 that the Fair Credit Reporting Act was passed by the US Congress to try and rein in some gross abuses in the credit reporting industry. The European governments, under the aegis of OECD and EU, accelerated the public debate on privacy and the Fair Information Practice Principles were drafted. In contrast with EU, in the US legislation was enacted sector-by-sector, using laws such as HIPAA, COPPA, Gramm-Leach-Bliley Act and most recently, the Sarbanes-Oxley Act. According to Garfinkel, policy rather than technology is the best tool to tackle privacy issues because policy can be technology-neutral and can address the human element (via lawsuits, etc.). However, policy can be influenced by special-interest groups and doesn't apply across national boundaries, whereas technology can shortcut some of these hurdles.

In the second hour, Garfinkel gave a quick overview of cryptography techniques. He explained that message digests or hashes were special functions that created a "fingerprint" for a block of data. Digest algorithms such as MD5 or SHA can be computed easily, and since any change in the data will result in a different digest output, such algorithms can be used to ensure integrity of data exchange. Digests also allow safe storage and communication of secret data such as passwords, and hence are commonly used in authentication mechanisms. Then he took a brief trip down the memory lane of encryption technologies, recalling the Nazi Enigma machine of WWII and the development of DES and AES. Next, he quickly discussed public key encryption technologies.

Garfinkel then showcased a number of privacy-protecting technologies. To block the increasingly annoying pop-up

ads on the Web, users can use a client-based proxy such as AdSubtract that re-writes each visited page's HTML to screen out advertising elements. The Firefox browser comes with a handy extension called AdBlock that allows users to black-list ads based on their URLs. Bugnosis was another showcased tool. It can be used to identify web bugs, invisible images that websites use to track user behavior.

Shifting to email privacy, Garfinkel reported that S/MIME, an email encryption standard, is in fact built into most email clients available today, but due to the apathy of vendors it was never widely adopted in the field. Hush Mail provides a Web-based encrypted email service, using a Java client that the users run locally. In addition to encryption in transit, there is interest in messages that are non-perpetual and may expire after a certain period. Omniva offers such a service to parties that consent to message expiration. It holds the decryption key on its servers and deletes them after a specified period, thereby rendering any copies stored on the clients' machines as well as intermediate servers and backup media, unreadable. Anonymous remailers and Web-browsing are also available, though all such services require the client to trust the server operator.

For more general protection against surveillance of network traffic, Garfinkel reviewed mix nets and onion-routing networks. The basic idea behind mix nets is that a message is handed off from one machine on the network to another, with some random chance at each step that it will be routed to the destination. Hence, each machine only knows the identity of the one before it, and no more. In onion-routing networks, the general idea is still that the trust relationship is spread out over multiple machines, so that a snoop would have to compromise multiple machines to compromise privacy, not just one, as is the case with anonymous proxies or remailers. Finally, various anonymous, distributed document publishing systems have been proposed and built, offering varying amounts of privacy to the publishers and viewers.

USER INTERFACE DESIGN, PROTOTYPING, AND EVALUATION

Instructor: Jason I. Hong, CMU

Summary contributed by Serge Egelman

Jason Hong, a professor in the Human-Computer Interaction Institute (HCII) at CMU, ran a tutorial on the basics of user interface design and evaluation. He started by giving an overview of why HCI is important. For programmers, the interface for a program often represents the majority of the design, development, and testing. Failure to create a good interface can lead to many problems including loss of revenue, reputation, time, and even lives. Potential costs aside, creating a good interface can be very difficult as the designers often do not represent those who will be using the program and therefore have little idea of the usability problems that most users will encounter. Because of this, interface design considerations need to be made all the way through the development process, iterating from prototyping to user testing at every step.

After giving an overview, Hong gave some examples of good and bad designs as well as methodologies for evaluating designs. One example mentioned was the Bay Area Rapid Transit (BART) ticket machines. The machines accept cash, ATM cards, and credit cards, yet have a different set of instructions for each payment method. This confuses users who have never used the machine before which results in frustration for the user as well as frustration for the people waiting in line behind the user. It was clear that the interface for this system was designed without the users in mind. Thus, when creating new systems, developers should conduct a task analysis to determine who the users are and what tasks they need to perform. Next, they can observe existing systems for inspiration and then create scenarios to test new ideas with potential users before actually building the new system. This gets rid of problems earlier in the design process where they are cheaper to fix.

Designing new systems with potential users in mind is a difficult task of its own. Problems arise because users often do not really know what they want and therefore designers cannot simply ask them. Users might not be familiar with all that technology has to offer. They might not understand all of the design constraints (e.g. budget, time, etc.). They also may not be familiar with good design practices or they simply just don't know what they want. One way around these problems is through contextual inquiry, which is the practice of observing the potential users in their natural environment. This way designers can gain a better understanding of how their systems will fit into the user's world, how the user might use the system, and how the system will meet the user's needs. Contextual inquiries are executed by documenting how the user accomplishes very specific tasks. Designers can then review the interview as needed to determine what features will be important to the user. Designers can also use this data to create different personae which can be used to visualize how a user might use the system. This is accomplished by creating specific tasks for each persona and then documenting in great detail what the persona would need to do to accomplish the tasks.

Prototyping is a very important step in designing a user interface because it allows designers to quickly experiment with alternative designs and get feedback on each of them. The first type of prototyping that Hong mentioned was low-fidelity prototyping. Low-fidelity prototypes look very little like the finished product; they are constructed out of materials that are easy to reconfigure as well as throw away when finished. But most importantly they are cheap and easy to rapidly reconfigure. One such prototype might

be made out of construction paper cut-outs or even sketches on a whiteboard. Storyboards are another example of low-fidelity prototypes. They are used to map out specific steps for completing a specific task which allows the designer to concentrate on interactions with the user. High-fidelity prototypes on the other hand look very much like a finished product. While they cost more and take more time to construct, their attention to detail allows the user to have a better perception of the finished product.

In creating new interfaces, designers must make them intuitive for the user. This is often accomplished through the use of conceptual models and interface metaphors. A conceptual model is something that easily creates a mental representation of how an object works and how one might control it. An affordance is one such example; an affordance gives the user clues to its operation. The use of an image link on a web page that looks like a button is an affordance: the image looks like you can push it, which then gives the user the idea that they can push the button (click the link). Additionally, metaphors can be used to remind the user of existing conceptual models. Hong used operating systems as an example: files, folders, and the desktop are all metaphors that give users clues to their use by conjuring up conceptual models of their physical world counterparts. Interfaces often fail because they have no clues or they have misleading clues. These lead to user errors, slow performance, and frustration.

Finally, once an interface is designed, it needs to be properly tested. The first thing that must happen is for the designer to decide what needs to be tested. This is accomplished by creating a report that describes the objective of the test, a description of the system, the environment, the participants, and the specific tasks that are to be completed. The tasks should be similar to tasks that the designers believe the finished product will be used to accomplish. Next, the designers need to determine what data to collect, and how it is to be collected. There are two basic types of data, process data and bottom-line data. Process data covers what the user is doing and thinking, while bottom-line data consists of a summary of what the user did. Process data is usually collected by recording the user as they attempt each task, either video or audio (or both). To record what the user is thinking, they must "think aloud" by continuously saying exactly what they are thinking while they complete the task. Once the data has been collected it needs to be reviewed in order to draw results. Knowledge of statistics is required to determine the significance of the quantitative results as they could vary greatly. Some things to look for in the results are how well the users liked the system, how easy it was for them to navigate through it, whether they preferred it to another interface, and how it might be improved. Paying attention to detail when taking the test results and improving the design will maximize the design effort, the time spent on future user tests, as well as the usability of the final product.

OPENING SESSION

Summaries contributed by Fahd Arshad

WELCOME AND OPENING REMARKS, BEST PAPER AWARD

Lorrie Cranor, SOUPS Chair, kicked off the conference by reviewing a number of current security issues that are due to usability problems. These problems range from patching to phishing, from spyware and spam epidemics to weak passwords and loss of sensitive information to crackers and courier services. We don't make matters any easier, she said, by using confusing, non-intuitive metaphors such as "spam" and "cookies" when referring to security and privacy-related concepts. Cranor noted that there was a lack of interaction between the computer security professionals and system administrators on one side and human-computer interaction experts and privacy advocates on the other. Only together do these two hold the key to usable security and privacy solutions, and the goal of SOUPS is to bring them together.

At the end of her opening remarks, Cranor and Refereed Papers Chair Mary Ellen Zurko presented the Best Paper Award to Giovanni Iachello, Ian Smith, Sunny Consolvo, Mike Chen, and Gregory D. Abowd for their paper "Developing Privacy Guidelines for Social Location Disclosure Applications and Services."

Cranor was followed by Pradeep Khosla, the co-director of CMU CyLab, the main sponsor of SOUPS. He also expressed the hope that collaborative efforts between security and usability experts, such as the SOUPS conference itself, would result in the elimination of the "un-usability soup" we find ourselves in today. He spoke briefly about CyLab and its global reach, and then introduced the keynote speaker, Bill Cheswick.

INVITED TALK

Speaker: Bill Cheswick, Lumeta
Title: My Dad's Computer, Microsoft, and the Future of Internet Security

Bill Cheswick wrote the book on firewalls. Literally! So it was a bit surprising when he declared that he "skinny-dipped" on the Internet, that is, navigated the Internet without a firewall. Cheswick's machine runs OpenBSD with almost all ports turned off and the only public service being SSH, so he doesn't look too vulnerable. However, as a yardstick for end-user security, he wondered whether an ordinary user could skinny-dip on the Internet. His test case was his own father, a gentleman who runs a Web browser, an email client, an IM client, and a stock-tracking application on, gasp, Windows! Can the elder Mr. Cheswick "skinny-dip"?

To most attendees, it came as no surprise that the Cheswick found his father's Windows machine chock-full of adware and spyware. Also unsurprising was the fact that even after a full cleanup, the machine was infected again within weeks (when the speaker visited his father next). Here's the punch-line: the father was adamant that none of the security "fixes" or "solutions" break his machine. After all, explicit and annoying pop-up ads notwithstanding, he was still getting his work done, wasn't he? Why fix something that ain't broke?

Cheswick hence illustrated a lesson that most HCI experts know, but which hasn't filtered down to the trenches yet: the damage due to risky online behavior (such as downloading "free" screensavers) must be made visible to the user. However, warned Cheswick, users and especially home users are always going to be the weak link in the chain, and trusting them to keep their machines secure is an exercise in futility. As long as there is a lack of separation of privileges and sandboxes in OSes like Windows, virus writers and malicious hackers will ultimately prevail. Cheswick borrows his solution from the even-keeled karate maestro, Mr. Miyagi: "The best block is not to be there". Translated to geek-speak, courtesy of Microsoft, "a feature you don't use should not be a security problem for you." Cheswick proposed a simple count of the number of services listening on ports on a machine as a measure of OS security. He reported that his measure progressively worsened between Windows 95 and XP, but that SP2 has gone some way towards addressing these potential entry points. Given that security by design is impossible with the millions of lines of code Microsoft now sells as Windows, the task of cleaning it up is Herculean. However, Cheswick believes there is hope in Longhorn, the next client coming out of Microsoft, IF it is re-written from scratch with security in mind. Microsoft is finally getting it, a few years too late. Of course, others such as SGI and other OS vendors made the same mistakes, and didn't live to recover from them.

Cheswick then proposed "Windows OK", a stripped down version of Windows that would suit a huge slice of the consumers, such as the senior Mr. Cheswick, who need basic Internet access and not much more. Portable code attack vectors, such as executable code in Office applications, would be disabled, security-related settings including ActiveX controls would be placed in a single, conveniently accessible security panel, and network services listening on ports would be minimal, if any. Cheswick said Microsoft may or may not achieve this, but they are certainly heading in the right direction with SP2. He believes that other OSes offer some hope, especially the Mac OS X; however, they are currently not an option for all users, given the lack of application support for them. His final message was optimistic: we seem to be converging on a more secure computing environment, with a safer OS under the hood, more clear controls, and hopefully virus prevention, instead of detection (which we can't keep up with).

In the Q&A session, Cheswick clarified that he had not written Linux off, but it needed to do more to become a true replacement for Windows for the average home user. He accepted that Windows may already have some of the tools he wants, such as data execution protection and local security policies, but administering them is a pain, and in desperate need of improved user interfaces. The final point of discussion was whether safety and security of software could be enforced by legislation, the direction in which Europe may be moving. Cheswick was not averse to the idea of liability for software, but wondered what would happen to Open Source products and individual application developers if such a scheme were to be adopted.

REFEREED PAPER SESSIONS

Summaries contributed by Rob Reeder

SESSION I: USABLE SECURITY

Chair: Mary Ellen Zurko (IBM Software Group)

Authentication Using Graphical Passwords: Effects of Tolerance and Image Choice

Susan Wiedenbeck, Jim Waters (Drexel University), Jean-Camille Birget (Rutgers University), Alex Broditskiy, and Nasir Memon (Polytechnic University)

Susan Wiedenbeck started the first paper session with a talk on authentication using graphical passwords. She first reviewed her PassPoints system, previously introduced by the author, in which users authenticate by clicking on a sequence of reference points within a photographic image. Prior work showed that such a system is comparable in terms of security (i.e., resistance to being cracked by a brute-force attack) and human memorability to textual passwords. This talk focused on whether the "tolerance" - the pixel size of the box within which a user must click for a point to be recognized as the same as the reference point - and choice of image made a difference in users' ability to remember their sequence of reference points. Wiedenbeck found that while a larger tolerance did lead to a greater reference-points retention rate, as might be expected, image choice made no significant difference in users' ability to remember their reference points.

Johnny 2: A User Test of Key Continuity Management with S/MIME and Outlook Express

Simson L. Garfinkel and Robert C. Miller (MIT)

The omnipresent Simson Garfinkel presented the second paper in the session. Garfinkel spoke about his work on a

user interface add-on, called CoPilot, to the Outlook Express email client for implementing email encryption through Key Continuity Management (KCM). KCM addresses the problems users have looking up and authenticating correspondents' public keys, and generating their own key pairs. KCM obviates the need for users to carry out these steps, and instead requires the email client to keep track of known correspondents. In CoPilot, Outlook presents a colored border that is green for signed messages from known correspondents, and red, yellow, or gray for other types of (potentially suspicious) messages. Garfinkel and coauthor Rob Miller carried out a laboratory user study in which users were involved in a scenario that required sharing secret information with trusted correspondents, and were "attacked" by virtual enemies trying to obtain the secret. The authors found that KCM with CoPilot did significantly reduce users' susceptibility to some forms of attack compared to the ordinary Outlook interface, but that users remained quite vulnerable to attack. Under the best conditions, only 33% of users consistently resisted attacks.

Two Experiences Designing for Effective Security

Rogerio DePaula, Xianghua Ding, Paul Dourish, Kari Nies, Ben Pillet, David Redmiles, Jie Ren, Jennifer Rode, and Roberto Silva Filho (University of California, Irvine)

Finally, David Redmiles spoke about some of his research group's design experiences. The highlight of this talk was Impromptu, a user interface for visualizing shared resources in a ubiquitous computing environment. Impromptu was presented as a design prototype with no substantive evaluation.

SESSION II: "USABLE PRIVACY"

Chair: John Karat (IBM T.J. Watson Research Center)

Usable Security and Privacy: A Case Study of Developing Privacy Management Tools

Carolyn Brodie, Clare-Marie Karat, John Karat (IBM T. J. Watson Research Center), and Jinjuan Feng (University of Maryland Baltimore County)

Carolyn Brodie of IBM Watson started the session with a talk on SPARCLE, a tool for authoring and managing privacy policies within an organization. The SPARCLE tool supports the construction of formal privacy policies through natural language, template-based, and structured-text interfaces. Brodie and coauthors have been evaluating SPARCLE with real users who are responsible for privacy within their organizations.

Stopping Spyware at the Gate: A User Study of Privacy, Notice and Spyware

Nathaniel S. Good, Rachna Dhamija, Jens Grossklags, David Thaw, Steven Aronowitz, Deirdre Mulligan, and Joseph Konstan (UC Berkeley)

Nathan Good from UC Berkeley followed Brodie. Good's talk focused on a user study which demonstrated what anyone might guess - that users don't read EULAs (End-User Licensing Agreements) when installing software. This can be a serious problem when failure to read a EULA leads a user to install spyware inadvertently. Good et al. found that even when EULAs were short and written for clarity, users still didn't read them. Good concluded that putting more resources into developing shorter or better-written EULAs will likely be wasted.

Making PRIME usable

John Soren Pettersson, Simone Fischer-Huebner, Ninni Danielsson, Jenny Nilsson (Karlstad University), Mike Bergmann, Sebastin Clauss, Thomas Kriegelstein (TU Dresden), and Henry Krasemann (Independent Centre for Privacy Protection)

Next up was John Soren Pettersson, who spoke about ways for users to specify privacy preferences. He discussed three user interface paradigms for specifying privacy preferences: role-based, relationship-based, and TownMap-based. He went on to show some interface designs using the TownMap-based paradigm.

Developing Privacy Guidelines for Social Location Disclosure Applications and Services

Giovanni Iachello (Georgia Institute of Technology), Ian Smith, Sunny Consolvo, Mike Chen (Intel Research), and Gregory D. Abowd (Georgia Institute of Technology)

Closing the session was Giovanni Iachello of Georgia Tech presenting the conference's best paper on a mobile-phone-based application for setting preferences for sharing location information. Iachello and coauthors used data from in situ observation of mobile phone users to develop Reno and Boise, mobile phone applications for sharing location information with others. A two-week evaluation period with kids during the Christmas season showed that few used the privacy-preserving features of these systems. The authors believe the systems' usefulness would be better demonstrated in a study over the summer.

SESSION III: "VISUALIZING SECURITY"

Chair: Diana Smetters (Palo Alto Research Center)

The Battle Against Phishing: Dynamic Security Skins

Rachna Dhamija and J.D. Tygar (University of California, Berkeley)

Rachna Dhamija of UC Berkeley presented a solution for securely authenticating users to webservers and webservers to users, with the goal of preventing phishing attacks. The solution uses a customized background photograph in login windows to indicate to users that the window is not spoofed, and uses random patterns in web browser borders to indicate secure connections to websites. The authors have yet to perform an evaluation to determine whether their solution will help prevent phishing attacks.

Attacking Information Visualization System Usability Overloading and Deceiving the Human

Gregory Conti, Mustaque Ahamad, and John Stasko (Georgia Institute of Technology)

Greg Conti of Georgia Tech illustrated a series of hypothetical attacks on information visualization programs. The upshot of his talk was that a smart attacker could send cover traffic to obscure an attack packet from the view of system administrators who monitor their network traffic with information visualization software. Conti and coauthors have developed a taxonomy of possible attacks on information visualization tools. While Conti presented no evidence that such attacks actually have taken place, he did show examples of attacks he and coauthors generated in the laboratory.

Social Navigation as a Model for Usable Security

Paul DiGioia and Paul Dourish (University of California, Irvine)

Paul DiGioia of UC Irvine suggested that users might be led to make better security decisions by showing them what others have done before. DiGioia believes that showing decisions made by others could help in a variety of security-related domains, but focused especially on a prototype interface for users to specify shared folders in the Kazaa peer-to-peer network. The interface has not yet been evaluated.

POSTER SESSION

Summary contributed by Ponnurangam Kumaraguru

Participants from both academia and industry presented work at the SOUPS poster session. Posters presented in the session covered many interesting topics related to security and privacy. The topics for the posters covered the complete spectrum of usable security and privacy research; topics included Understanding User Attitudes Towards Personal Information on the Web, Improving the Usability of Web Browser Security and Patterns for Aligning Security and Usability.

The posters presented could be grouped into 4 broad categories:

1. Systems research (example: "A Software Composition Flaw in Google Desktop Search", which presented the idea of combining two different policies and forming a global policy)
2. User or survey studies (example: "Private Lives: User Attitudes Towards Personal Information on the Web")
3. Applications research (examples: "POLARIS: Usable Virus Protection for Windows" and "Design and Evaluation Method for Secure 802.11 Network Configuration")
4. Usability research (examples: "Patterns for Aligning Security and Usability" and "A Dependable User Interface for Setting XP File Permissions").

A few of the posters had practical demonstrations of the solutions proposed ("Still Searching for Privacy" and "Peripheral Privacy Notifications for Wireless Networks", for example). A few other posters had a very well formatted story behind their presentation (such as "FAMILYNET: A Tangible Interface for Managing Intimate Social Networks"). Here the researchers presented various situations to the audience with the sample tags which were used in their research; they highlighted and provided the solutions keeping the audience in the discussion. This stimulated lively discussion among the attendees of the poster session.

Overall the poster session created a good platform for various researchers to present their ideas and get feedback on their research. The audience enjoyed good food and good discussions.

PANELS

Summaries contributed by Fahd Arshad

USABILITY OF SECURITY ADMINISTRATION VS. USABILITY OF END-USER SECURITY

Moderator: Konstantin Beznosov, University of British Columbia
Mary Ellen Zurko, IBM Software Group
Stephen Chan, SIMS Dept, UC Berkeley
Gregory Conti, Georgia Institute of Technology

The energetic moderator, Konstantin Beznosov, started this panel by laying down three major questions:

I. Is the notion of usable security the same for end users and security administrators, given the differences in their backgrounds, training, goals, constraints, and available tools?
II. How much of what we learn from each group about usable security is applicable to the other?
III. Since many people today are not full-time security administrators, what distinguishes users from administrators?

Mary Ellen Zurko, who has been working on security issues for almost two decades, suggested that we add the notion of "power users" to our binary model, placing these users midway between end users and administrators. Responsibility for security can be planned in terms of the software lifecycle; since developers are the earliest link in the software lifecycle, they hold the most responsibility for designing with security in mind. Similarly, administrators have more responsibility than end users. Someone has to make tough decisions about security. Since responsibility is shared, however, overrides down the line must be possible. Security policies set the defaults administrators choose, and preferences are user-level overrides.

Stephen Chan brought to the podium his extensive experience in the trenches of system administration. He noted the importance of a partnership between the designers and the security administrators to design good tools for usable security, and between different groups within administrators and users to create safe and practical security policies. He stressed the importance of discovering the work practices of security administrators when designing interfaces. He gave three main reasons why admins prefer to work with raw text (Emacs or vi). First, GUIs don't scale, especially when managing clusters or large storage sites. Second, work practices differ across security admins, based on their own background, and typical GUI abstractions don't match the admin's needs. Finally, in operational security, the routine is dynamic; the game is constantly changing, and interfaces often aren't flexible enough to keep up, forcing admins to revert to scripting.

Gregory Conti related his experience speaking on interface design at a recent DEFCON (a hacker conference). The muted response to his talk reinforced his opinion that there was a clash of cultures between security experts and admins on one hand and HCI experts and designers on the other. This is exemplified by the tools used by the two groups: man pages vs. Clippy, or iptables vs. Zone Alarm. Even within the "expert" community, there is stratification: note the contempt with which Perl programmers hold VB programmers. Hopefully SOUPS can help reduce this culture gap.

A vibrant discussion session followed. The first question was about the partition of security responsibility between users and administrators. Can admins take responsibility of all security issues? Why do end users need security interfaces at all? Chan's response was that security cannot get in the way of end users' primary goals. To achieve this, the responsibility must be delegated to and shared with users. A member of the audience provided an anecdote about a security admin who wisely left a job where he was responsible for security policies, but not enforcement.

Another set of questions dealt with interfaces for administrators. Is there a need for usable security for administrators, given their text-central approach? Why haven't administrators built their own tools? Why haven't any of the many proposed data visualization systems caught on? Will security tools ever become "services" to lessen the load on end users? The panelists noted that there was definitely a need for better interfaces. Administrators don't use grep because they are masochistic; it is because none of the existing tools offer the scalability and flexibility they need. Many tools have been built by administrators to help with their tasks, such as Tripwire, Snort, various firewalls, etc. Raw text remains the most flexible and shareable medium for these tools. Current data visualization systems often require training and don't scale very well. There has been thin task analysis and very little evaluation of such systems. Also, by their very nature, data visualization systems provide abstractions which attackers can manipulate and exploit. Finally in the service-based model of security, a great deal of trust needs to be placed in the service provider. Garfinkel pointed out that Apple already provides a comparable service called .Mac.

Another important issue was what it is that distinguishes usability in security from usability in other task domains. Three important differences were identified: first, the fact that an active human adversary is tries to exploit holes in security interfaces; second, the critical nature of security systems and the high cost of their failure; and finally, the need to assimilate large amounts of data in real-time.

WHEN USER STUDIES ATTACK: EVALUATING SECURITY BY INTENTIONALLY ATTACKING USERS

Moderator: Robert Miller, MIT
Simson Garfinkel, MIT
Filippo Menczer, Indiana University Bloomington
Robert Kraut, Carnegie Mellon University

Robert Miller started the discussion by pointing out the differences between how usability and security are evaluated: usability studies are usually carried out in the lab under controlled environments, whereas security experts tend to use analysis and attacks as their primary tools. How to we measure the security of the entire system, including the user? We need to attack the user, but we can't do so as blithely as we attack software. Designing user studies to measure security is also difficult because security is almost always a secondary task for the user. How do we motivate users to protect their security in an artificial scenario?

Fillipio Menczer presented some work he had done with his colleagues and students at Indiana University Bloomington (IUB) on studying the social context of phishing attacks. His team used data publicly available from social networking sites such as Orkut and the Web to establish potential social connections between their participants. Then they sent forged emails to the participants from other students they knew, directing them to a non-intranet Website that asked the students for their intranet login and password. The control group received emails from people they did not know. In spite of the obvious clues in both the spoofed email and the spoofed password interface, one in seven participants in the control group gave up their credentials, but even more strikingly, almost three of every four students who received emails purportedly from other students they knew gave up their credentials!

Menczer described in detail the difficulties faced in designing the study, getting approval from the Institutional Review Board (IRB) at IUB, and responding to participants afterwards. The study required a waiver of consent from the participants, and this sets a very high threshold for the study designers. The IRB engaged in a dialogue with Menczer's team to help them fulfill the conditions for obtaining the waiver by modifying the study design and adding elements, such as an anonymous blog for participants to offer feedback. He also shared some lessons learned from participant response after the study, as well as the publicity their work and methods attracted, positive and negative, from the participants, the IUB community as well as the online community at large.

Garfinkel reached back into his training as a chemist and pulled out the example of titration as an analogy to measuring human response in stimuli. In this process, a known agent is carefully added to a solution with an unknown amount of another agent (such as a base to an acid solution). At a very narrow window, the color of the solution changes, and then changes again as more of the agent is added. Just as the amount of the titration agent is closely monitored to determine its affect, argued Garfinkel, so must we measure the results of attacks on our participants, and the effectiveness of our solutions. He pointed out various design decisions we must make in designing a controlled security study: how to remove bias due to attack selection and ordering, how to handle the issue of prior consent vs. motivation, and what are the ethical obligations of teaching the participant how to avoid harm in the real world.

Robert Kraut is a member of the Carnegie Mellon University IRB and talked about the rules and ethics of IRBs, which seem to be such an enigma for most computer scientists. He gave an overview of the historical tragedies, namely the medical trials on Nazi concentration camp inmates and the Tuskegee syphilis study. These led to the Belmont Report in 1979 which established three main principles for human subject research: respect for persons, beneficence, and justice. By law, any university obtaining Federal funding must abide by these using the IRB mechanism. IRBs consider informed consent to be one of the baselines for respect for persons. Beneficence requires that risks be minimized and benefits maximized. Justice requires equal risk sharing and overall fair treatment of the participants.

In the context of Menczer et all's study, the social network harvesting did not need informed consent because the data was collected without interaction with the participants. However, consent would be required in the actual phishing contact. The IRB looked at a number of issues, including whether the participants had a reasonable expectation of privacy, what was the risk or harm to the participant vs. the risk to society at large due to this research, etc. To obtain a waiver of consent from the IRB, the study designers had to ensure that the harm to participants was minimal. They did this by not storing the credentials submitted, only verifying them against a secure server for validity and storing this information instead. The IRB asked them to beef up the post-study briefing by adding an anonymous blog as a feedback mechanism. Finally, the IRB weighed the benefits to society from the research, and having evaluated all these to its satisfaction, granted its approval.

Kraut concluded that the IRB process is necessarily thorough and involves enough ambiguities and tradeoff decisions to necessitate a case-by-case analysis. He noted that the burden was often on the petitioner to present their case strongly before the IRB and to educate them about the pertinent principles. He offered his help to anyone who wished to engage in this process.

DISCUSSION SESSIONS

WHEN USER STUDIES ATTACK: EVALUATING SECURITY BY INTENTIONALLY ATTACKING USERS

Summary contributed by Fahd Arshad

Moderator: Robert Miller, MIT
Simson Garfinkel, MIT
Filippo Menczer, Indiana University Bloomington
Robert Kraut, Carnegie Mellon University

During the panel session on this topic, the speakers didn't have time to engage in a discussion so the discussion session was welcomed enthusiastically by many of the panel audience. In response to a set of questions on what constitutes collected data not covered by IRB rules, the panelists pointed out that images collected off the Internet are often governed by copyrights, and while these guidelines apply only to research at institutions which accept Federal money (and not corporations, for example), the guidelines for publication may in some cases be stricter than those of the IRB. Of course, the existence of rules should not blind researchers to the ethics of their methods. In response to a question about the additional need for debriefing in situations where consent was not obtained, Kraut pointed out that the follow-up allows participants to withdraw from the study after the fact, even if they could not withhold consent prior to it. Also, follow-up briefings fulfill the ethical goal of educating the participants.

Other questions dealt with what is acceptable risk, given that the standard is what a participant may face in daily life. Kraut clarified that risk is a product of probability and magnitude of harm. So while the low frequency of attacks in the real world is often turned up in lab settings, a low potential for harm in HCI and security scenarios should keep risk levels reasonable. That said, "harm" is notoriously prickly to measure and almost always involves subjective measures. Comparative risk is often used as a yardstick.

Wendy Mackay pointed out that whereas psychology PhDs receive training on human participant research in graduate school, Computer Science programs usually don't include such instruction. A member of the audience suggested that consulting colleagues from other fields who have such experience, such as sociologists, can prove immensely helpful. Lorrie Cranor recalled that when she shifted from industry to academia, she didn't have any idea of the rules governing IRB. This meant that the first time she and a student needed to run a privacy-related study, the initial plans had to be scaled back vastly to fit the perceived constraints of the IRB. IRB procedures are often biased towards the fields of medicine and psychology, and their standards may even drift upwards with time. Hence a number of attendees stressed the need for a set of guidelines for human participant research that are aimed towards the usability, privacy, and security fields and cover some of the tricky issues such as what is governed by IRB and what is not, under what conditions is informed consent needed, what is acceptable risk, etc. Cranor offered to collect any guidelines the attendees could find and make them available off the SOUPS website.

USABILITY AND ACCEPTANCE OF BIOMETRICS

Summary contributed by Elaine Newton

Moderator: Andrew S. Patrick, NRC Canada

A group of six people, including moderator Andrew Patrick, convened to discuss "Usability and Acceptance of Biometrics." After each attendee introduced themselves, Patrick led a discussion on previous and possible future studies on user acceptance as well as government promotion of biometrics for various functions despite low performance and sometimes low acceptance. Previous user tests have studied ergonomics, accuracy, speed, reliability, learnability and feedback, and ability to integrate with associated systems. These studies show that many users have a lack of understanding about biometrics. Some are concerned about privacy, function creep, and/or risk to life and limb (such as touching one's eyes or fingers). Examples of large-scale government deployments with little or negative findings in research include the U.S. VISIT program and the UK enrollment trial, which found that quality of fingerprint images and enrollment to be major factors negatively affecting performance.

Patrick proposed performing a large-scale survey and made a pitch for collaborators. He drafted a survey instrument with 34 questions (available at <http://www.andrewpatrick.ca/BioSurvey/>) to gauge knowledge and acceptance of biometric security systems and cross-cultural differences and presented it to his organization's IRB (at NRC). His IRB and the group convened at SOUPS both wanted to hear more about how data about users' attitudes would be used. Would it be used to influence policy makers, and what would be right or wrong directions for use of the data? Other areas that may be covered in a survey include if/how media influences on privacy attitudes, if/how different scenarios of deployment effect attitudes, and if/how peoples' beliefs change based on which biometric is being used (e.g., fingerprint vs. iris vs. face).

The group also discussed a need for longitudinal studies to gauge changes in attitudes; codified threat models, purposes, and utility of biometrics systems; and an understanding of beliefs of the policy makers that propose use of biometric systems.

VALUATION AND CONTEXT

Summary contributed by Steve Sheng

Moderators: Kimberly Perzel and Seth Proctor, Sun Microsystems

This session brought together participants from various backgrounds, including security experts, HCI practitioners, and social science researchers from both academia and industry to discuss questions about how people place value on security and privacy in different contexts. The moderators began by asking what are some of the contexts that we generally engage ourselves in, but this question turned out to be a bit too general. So we engaged ourselves in giving a working definition of context. The next question was what are some of the general values in security? What are some of the tradeoffs? How are these tradeoffs affected by switching to different contexts?

This discussion illustrated the difficulty in trying to design values into computer systems. Articulating contexts and values is very difficult. The discussion did not go into details about how to translate these tradeoffs into computer systems.

USABLE INTERFACES FOR ANONYMOUS COMMUNICATION

Summary contributed by Fahd Arshad

Moderator: Roger Dingledine, The Free Haven Project

The moderator gave an overview of Tor, an onion-routing-based anonymous routing network. The network has a large number of nodes and has invited interest from a diverse circle of clients, including the Department of Defense, journalists, and of course, private citizens. Tor currently has a pure command-line interface, and building a graphical interface presents a few design challenges. One of the problems that the maintainers have already identified is how to make the system state visible to the user. Users often don't realize whether their packets are being routed through Tor or not. They also attribute network connectivity problems to Tor when in fact the network link itself may be down. A good network hop visualization system would help in two ways: first, it would allow novice users to get a good mental model of the multi-hop system. Currently, many users don't understand that their information is being routed through a number of nodes, instead of a single one, as with proxy-based systems. Secondly, a visualization system would allow users to pick their exit points from the network, in order to adapt to bypass local censorship. The interface for managing Tor servers also needs some thought. Finally, currently Tor is a single-user application, but given interest from various organizations, it may morph into a multi-user application. How would the interface scale to adapt to this?

The Electronic Frontier Foundation, which Dingledine works for, will soon establish a two-part GUI contest to design interfaces for Tor. In the first part prototype sketches for the Tor interface can be submitted, and implementations will be due in the second part.

COMMENTARY ON SOUPS 2005

THREATS ADDRESSED AT SOUPS

Commentary contributed by Rob Reeder

One useful way to classify the work presented at the conference is by the threats to computer security and privacy that it addressed. By compiling a comprehensive list of such threats, gleaned from SOUPS and similar forums, the HCISEC community will be better able to understand the scope of the problems it needs to address.

Table 1 shows a list of security- and privacy-related threats and the SOUPS 2005 papers that addressed each threat. Amongst the threats are two related to authentication, both users authenticating themselves to systems and systems (specifically, websites) authenticating themselves to users. Three threats stem from the inherent complexity of tasks, namely encrypting email, setting access rights, and managing privacy policies. The last two threats are user unawareness of software functionality (which can lead to malware and spyware installation) and obfuscation of network administrative tools (specifically, information visualizations).

EVALUATION: A BIG ISSUE

Commentary contributed by Rob Reeder

It was clear from the conference that one of the greatest challenges facing the HCISEC community is how to evaluate the design ideas and systems it generates. Eight of the ten talks from the paper sessions presented new interfaces for improving system security and/or privacy, but it was not clear in all cases whether the proposed interfaces were an improvement over existing technologies. Only four of the eight proposed interfaces had been evaluated in a manner that was clear from the talks: PassPoints, CoPilot, SPARCLE, and Reno, and the how-to-evaluate question was commonly asked of speakers who did not present evaluations of their systems. Evaluations that were presented varied in metrics and methodology. Some of these metrics included retention rate of passwords, attack resistance rate, and subjective user satisfaction. Methods included lab user studies, interviews, and in-situ field studies.

BROAD CATEGORY	SPECIFIC THREAT TO SECURITY/PRIVACY	SOUPS 2005 PAPERS ADDRESSING THREAT (SYSTEM NAMES IN PARENTHESES)
Authentication	User forgetting password or choosing password that is too easily guessed	Wiedenbeck et al. (PassPoints)
	Spoofed websites (phishing)	Dhamija and Tygar (Dynamic Security Skins)
Complex security and/or privacy-related tasks	User incorrectly encrypting email	Garfinkel and Miller (KCM/CoPilot)
	User setting unintended privacy or security access rights	DePaula et al. (Impromptu) Iachello et al. (Reno/Boise) DiGioia and Dourish (new Kazaa prototype)
	User or organization mismanaging or poorly specifying complex privacy policies	Brodie et al. (SPARCLE) Pettersson et al. (PRIME)
User awareness	User unknowingly installing spyware or malware	Good et al.
Attacks on administrative tools	Attacker obfuscating diagnostic information visualizations	Conti et al.

Table 1. Threats to system security and privacy and the SOUPS 2005 papers that addressed them.

The variety in evaluation methods may be necessary and even desirable, but the variety (and in some cases, subjectivity) of summative metrics makes it difficult to compare systems to each other and to determine when progress has been made solving a given problem. An entire SOUPS panel and a discussion session were devoted to the topic of evaluating systems for both usability and security, with a focus on the ethics of intentionally attacking users. But, ethical concerns aside, a greater question was: what kinds of evaluation are needed to provide convincing evidence that a given system will actually help prevent attacks? Determining a meaningful set of metrics for evaluating the actual effectiveness of usable security systems remains an important open problem.

BRINGING HCI AND SECURITY TOGETHER

Commentary contributed by Fahd Arshad

As noted earlier, this symposium was organized because at prior events a need was felt for a forum where researchers interested in the fields of usability, privacy, and security could come together. The security community seemed a bit under-represented at SOUPS 2005, though. Of course there were exceptions such as Mary Ellen Zurko. Yet mostly we heard from HCI experts talking about the usability of security. Other than Stephen Chen and Gregory Conti on the last day, we rarely heard from a security researcher posing usability and privacy questions. Conti's talk presented some of the drawbacks that arise when usability methods are applied to real-world security issues. Chen's input was invaluable to a number of HCI experts because he tried to explain what the needs of security administrators are and why many of the current proposed solutions don't cut it in the trenches. The next SOUPS would be well-served to invite more security researchers and take this dialog further.

WHO'S LOOKING OUT FOR THE ADMINISTRATORS?

Commentary contributed by Fahd Arshad

Most of the work presented at SOUPS was about the usability of security and privacy solutions at the end-user level. Perhaps this is because the end user is the weakest link, the "bozo" in Cheswick's terminology. However, most members of the HCISEC community would agree that a lot of work remains to be done at the administrator level also, and perhaps at more stages in the spectrum between the end user and the administrator where most real-world users fall. What are the usability problems of a user group that isn't distinctly identifiable? How can we provide flexibility, scalability, real-time response, and lack of obscurity (just to name a few requirements noted at SOUPS) to those responsible for security of large networks, instead of individual machines? Who is looking out for the administrators?

THE FUTURE OF SOUPS

Commentary contributed by Rob Reeder

Indications in Lorrie Cranor's closing remarks were that SOUPS will be held again in 2006. SOUPS 2005 was a successful effort in bringing together an emerging and expanding HCISEC community. SOUPS 2006 can build on this success by concentrating its vision for usable secure and privacy-protecting systems.

In talks about evaluating interfaces for security and privacy, I wondered aloud to a few people over lunch how we will evaluate SOUPS itself. Four metrics were mentioned: quality of submissions, citations of SOUPS publications, collaborations that grow out of SOUPS, and attendance. These seem like the right metrics by which to measure success, but we must ask how to improve performance along these metrics. The greatest improvement is likely to come from a more focused vision of the questions and problems that the conference is meant to address. As the HCISEC/SOUPS community is relatively new, its nascent vision is still murky. A clearer vision of the problems the SOUPS community is trying to solve and how to measure progress toward solving those problems will necessarily lead to higher-quality publications and more citations, and will likely encourage greater collaboration and attendance by providing a better intellectual product to conference participants. Future SOUPS conference participants must strive to clarify this vision.