

Mary Ellen Zurko leads security architecture and strategy for Lotus Workplace, Portal, and Collaboration Software at IBM. She defined the field of User-Centered Security in 1996. She is on the steering committee for New Security Paradigms Workshop and the International World Wide Web Conference series (she is co-chair for WWW2007 in Banff). She has worked in security since 1986, at The Open Group Research Institute and Digital Equipment Corporation, as well as IBM. She is a contributor to the O'Reilly book "Security and Usability: Designing Secure Systems that People Can Use." Her vita is at <http://mysite.verizon.net/resqwf60/id1.html>. She is the chair of the W3C Web Security Context Working Group.

Maritza Johnson is a Ph.D. student at Columbia. Her research interests are in human-computer interaction and human factors and how they relate to and affect the usability of security. Her current projects include creating a recommended procedure for evaluating the usability of technologies for authenticating a financial institution to the customer on the web (in collaboration with the FSTC), in addition to participation in this working group. She is a newcomer to the field, but is currently steeped in projects that heavily rely on effective user studies and their results.

Security User Studies and Standards: Creating Best Practices

Maritza L. Johnson, Columbia University
Mary Ellen Zurko, IBM

The Web Security Context (WSC) Working Group¹ is breaking ground by being the first standards effort to have usable security indicators as (one of) its primary goal(s). A goal of the WSC WG is to create a usable baseline and best practices for the display of security context information a user should be provided when expected to make a trust decision on the web. (Our other goal is recommendations on how to make that information robust against attacks.) Our work is a direct result of the rising tide of web-based social engineering attacks such as phishing, which motivated the W3C Workshop on Transparency and Usability of Web Authentication². The participants of our working group put us in an excellent position for getting the standards not only recognized but implemented as well. Group participants include a number of the major browser vendors, security specialists in industry and academia, recognized names in anti-phishing and usable security, and businesses who rely on web applications. A major challenge is that we lack the experience of a pattern for success for making the difficult tradeoffs in designing for both security and usability in the context of a standards effort.

The quality of our standards work relies on the same aspects all standards work relies on; deployment experience, architecture and design expertise, prototyping and implementation experience, and testing multiple implementations against the standard and each other. The success of our recommendations is rooted in their security and usability, which means it will be important to verify them with effective user studies and testing. While our group consists of members who are active in usability research and product design, being a standards group, the opinion and collaboration of other area researchers will serve beneficial for our purposes. As far as we know a standards group has not previously undertaken usable security even though the need for a baseline is there. We are calling on a number of established techniques in both security and usability, including user studies, to make this effort successful. User studies and testing will augment or supplant traditional interop testing in this standards context.

In addition to the issues present in creating security user studies with typical research goals in mind, user studies to motivate and validate usable security standards also include unique implications. From a process point of view, how does the lack of a pre-defined institutional control board impact ethical concerns, since a user study in academia typically has a designated board to guide and approve practices. A critical aspect for our user testing is what the impact is to the design of a user study if it is known in advance that the results are intended to support or guide the development of standards (as opposed to forming the basis of publishable research). Since standards discussions are based on group consensus, the impact and meaning of the results of any user study must be accepted by the group to be useful. And since standards recommendations are meant

¹ <http://www.w3.org/2006/WSC/W3C>

² <http://www.w3.org/2005/Security/usability-ws/>

to apply to a broad set of existing and future technology, the results of user testing must be generalizable to the standard's scope. We will inevitably face numerous questions about how to structure our user studies so that the testing on specific users and specific implementations (whether low or high fidelity) applies to our full recommendations (or the parts of the recommendations that most need to be validated with direct user testing). Those recommendations will in turn be used in implementations we can neither anticipate or control, and need to be useful and valid in those unknown future contexts.

Published results from recent user studies have given our group a starting point but we agree with the goals of the workshop that we can benefit from attempts to improve the design of user studies. For an example of a user study whose goals were different from those of established research, see *Did You Ever Have To Make Up Your Mind? What Notes Users Do When Faced With A Security Decision*³. The goal of that work was to motivate the commitment to features that would improve the deployability (and usability) of Lotus Notes Execution Control Lists. In order to be maximally compelling to product management, the entire design of the study was biased strongly in favor of disproving the need for additional features. That bias was to ensure that any data from the "in the wild" user testing that argued for additional product work would be accepted by product management as valid.

The WSC WG is producing the draft Note⁴ that will define our goals, scope, use cases, foundational software, and approach. As part of our proposed design principles, we have combined the published results from prior user studies to establish a concept for who the average user is, how they understand security on the web, and how they interpret current security indicators. Relying on the results of previous user studies (*Why Phishing Works*⁵, *Decision Strategies and Susceptibility to Phishing*⁶, *Gathering Evidence: Use of Visual Security Cues in Web Browsers*⁷, *Do Security Toolbars Actually Prevent Phishing Attacks?*⁸) the following issues come up: users do not know what to look for when they need security information, they are not familiar with security concepts like certificates, and they may not be able to easily access the security information they need. As a standards group, we feel this data is useful as a starting point, but the limited reach of the participant group for each study leaves the strength and scope of its use open to future questions. There are definitely categories of users for which we have no such data, who we will also need to consider in our recommendations (such as web power users and web developers). Discussion of what can and should be done both within the working group and outside of it to make this approach more complete includes lobbying for a large-scale user study and doing studies (similar or differently) targeted at the other user groups we define.

³ <http://www.acsa-admin.org/2002/papers/7.pdf>

⁴ <http://www.w3.org/2006/WSC/wiki/NoteIndex>

⁵ http://people.deas.harvard.edu/~rachna/papers/why_phishing_works.pdf

⁶ http://cups.cs.cmu.edu/soups/2006/proceedings/p79_downs.pdf

⁷ http://tjwhalen.googlepages.com/eye-tracking_gi.pdf

⁸ <http://www.simson.net/ref/2006/CHI-security-toolbar-final.pdf>

Another reason we have an interest in designing user studies is because the majority of the studies published that can inform our standards work are focused on one or two attack use cases, mostly phishing. Since our recommendations will be for displaying security context information in general on the web, we have a much broader set of use cases to consider. A core issue will be consideration of display of positive security context information that may be taken by the user as a confirmation of their safety. While the browser padlock model and some encrypted mail studies (Why Johnny can't encrypt⁹, Johnny 2: A User Test of Key Continuity Management with S/MIME and Outlook Express¹⁰) provide data on displaying information on encryption or authentication (signing), there are many questions on applying them to general web security context display, particularly given the ability of any web application to attempt to mimic them. Published user studies have again given us a good model for designing a study, but it may be beneficial to consider different user study techniques depending on the various use cases.

Given the limited amount of data available regarding the typical user's knowledge of security as it relates to the web, it would be helpful to conduct a user study with a larger and/or more diverse participant pool with the following objectives:

- validate or refine our understanding of who the average web user is and what she understands about security
- validate or refine our assumptions about how current security cues are understood and used, and how well that works today
- develop a model of how consistent use of positive or negative security indicators effects users over the long term (and how use of those same indicators by their friends, family, and colleagues influences behavior)
- test user reaction to proposed WSC solutions using UI prototypes of new security indicators, messages, and task flow.

Our working group time and resources are limited; our current schedule requires us to draft recommendations by the time of this workshop, and flesh them out with enough assurance to make them Candidate Recommendations by the end of the year. Certain types of studies will not fit into that sort of tactical schedule, so we must consider what we can, should, and must do now, and what we recommend be done as follow on, iterative, or maintenance work.

The results from user studies will be important at several points in our recommendations development. We will create a profile of average users with results from prior user studies in the design of our recommendations. Results that highlight user's problems with the lock icon which include: its susceptibility to spoofing attacks and confusion about the meaning lead us to ponder whether they could have been avoided by standardizing the lock in a fashion backed with results from user studies prior to its use. We also have a growing list of references to results from prior user studies and are documenting how they support or disprove common design techniques, which we will use to guide how we recommend displaying information to the user. These sources have provided a foundation for the recommendations, which we will evaluate through a combination of techniques,

⁹ http://www.cs.berkeley.edu/~tygar/papers/Why_Johnny_Cant_Encrypt/OReilly.pdf

¹⁰ <http://cups.cs.cmu.edu/soups/2005/2005proceedings/p13-garfinkel.pdf>

including traditional expert and community feedback, and our own user studies. The user studies performed will be modeled after previous user studies. The variety of our use cases may create a need to develop other methods for drawing feedback; which is something we have yet to explore, but the issue may rise later in our work.