I am pursuing an academic career, because I have a passion for teaching and research. I am a software requirements engineering researcher with applications in security. My work is in the intersection between software engineering, security, and psychology. I am interested in studying expert decision-making in the design and post-deployment phases of software development, where I try to model human knowledge to create human-centric technical solutions that conform to the real-world reasoning. My research investigates how do design secure software when: security decisions involve multiple factors (e.g. risk analysis, attacks, vulnerabilities); uncertainty is always present with varying degrees in human decisions; and the number of experts in domains like security is scarce, which limits the volume of data collected. To address these challenges, I use mixed quantitative and qualitative methods from multiple disciplines. I collect data using interviews, surveys, and user studies by creating new non-traditional computer science approaches that are well-known in social sciences, such as factorial vignettes and mixed methods designs. For data analysis, I have used the grounded analysis used in social sciences, the theory of situation awareness known in psychology, and the advanced statistical multi-level modelling. I model the analysis results using fuzzy logic, which is a formal method used in computational intelligence.

Below I will explain my past research and my future research directions.

Motivating a Hard Problem: Security Expert Decision-Making at Scale

Organizations rely on security experts to evaluate their systems and provide recommendations. Despite the abundance of well-documented security best practices, such as the NIST Special publication 800-53 that lists 256 security controls, security experts rely on their own expertise and tacit knowledge to assess the security risk and provide recommendations. The analyst often must reason over potentially millions of scenarios that account for various permutations of network type, authentication requirements, services offered, threat type, etc. When requirements change by adding new components and features such as multifactor authentication, these risk calculations must be updated. In addition, security experts in the world are scarce. There are 82,900 information security analysts in the U.S. in 2014 according to the U.S. Bureau of Labor statistics, and there is an expected 58% growth in demand by 2018[1]. The scarcity of experts and the need for cyber security as number of information security incidents keep increasing, makes the provision of intelligent decision support and semi-automated solutions a necessity. I will explain below different phases of this research that illustrates my approach to tackle this hard problem.

Technical Challenges

Exploring Challenges in Security Decision-Making. To reflect on the slow, deliberative decision-making process of security analysts, I have obtained qualitative data by interviewing security experts (some with over 10 years of experience) and used grounded analysis on interview transcripts to discover patterns of situation awareness (SA), a decision-making theory from cognitive psychology that decomposes decision making into four states: perception, comprehension, projections, and decision. In my work, I validated the decision-making model of situation awareness and the resulting patterns show that analysts try to handle uncertainty using assumptions and prior knowledge. In addition, the analysis shows that even when presented with a checklist of requirements, experts in practice tend to put security requirements in a context by creating their own scenarios and analyzing possible vulnerabilities and attacks. This work revealed the opportunity for further studies that measure how changes in certain factors, such as network configuration, and password requirements can increase or decrease the experts rating of security. Our work also reveals that experts who combine hands-on industry experience with academic knowledge exhibit different patterns of situation awareness compared to novices who rely on their academic background to evaluate security. The experts made assumptions when faced with uncertainty while novices asked the interviewer for more details. The experts also demonstrated patterns where they think from an attacker perspective, but novices failed at demonstrating an attacker. These patterns may be useful to design tests to measure whether novices can be effectively trained to reach expert SA.

Establishing the scientific validity of Ad-hoc Security Measures. To ask experts about security decision-making, we must first decide on the measure to be used for security. Our initial expert interviews show that experts were hesitant to describe a feature or a component as secure or insecure, and they preferred to say *it*

[1] A. Setalvad, "Demand to fill cybersecurity jobs booming," Peninsula Press, 31-Mar-2015.

depends. Decision scientists investigate how to measure decisions, and this requires researchers to design and validate new scales. The field of *Psychometrics* establishes the research methods for scale development. In psychometrics, we distinguish between face validity and construct validity. In contrast to construct validity, the face validity is subjective: if a test or measure looks like it will measure what it is supposed to measure, then it is said to have face validity. We focus grouped a list of possible scales to describe security, and concluded by choosing *adequacy* of security requirement(s) as a metric. To follow methodological rigor, we empirically evaluated the adequacy scale by selecting 17 words that are synonyms to: inadequate, adequate and excessive, and we invited 205 participants to rank the words. Because this ranking study depends upon human comprehension of natural language, we also screened participants for English proficiency using the Nelson-Denny English proficiency test. The empirical evaluation shows that participants can reliably use a scale of adequacy to rate security. To align our semantic scale with intervals we invited 38 security experts where we asked each expert to provide an interval for the word on a scale from 1-10 while imagining that the word is describing a security scenario. The collected intervals show that the labels: adequate, inadequate and excessive covers the entire scale from 1-10 when modeled using type-2 fuzzy sets.

Capturing the Effect and Priorities of Composable Security Requirements. Asking security experts about decision-making introduces five challenges: 1) *composition*, risk assessment of a system must consider the system context in which the requirements apply, and the composition of requirements with components of a system; 2) *priorities*, some requirements have higher priorities than others, depending on their strength in mitigating threats; 3) *ambiguity* in abstract terms that could lead two experts to interpret a requirement differently; 4) *Stove-piped knowledge*, security expertise crosses different domains of knowledge, such as hardware, software, cryptography, and operating systems; and 5) the *scarcity* of security experts. Hence, I developed the Multifactor Quality Measurement method (MQM), which models dependencies among requirements, and estimates how these requirements affect a perceived level of quality in a requirements specification, called a scenario. The MQM process starts with ad hoc bootstrapping of scenarios using *factorial vignettes*, a social sciences method where scenarios are constructed using a template consisting of factors of interest. By treating the factors as variables and manipulating the variables and their levels, we generate different instantiations of the template. Generating multiple vignettes allows us to elicit more information from a smaller number of experts. This creative and new study design adds more statistical power (increasing power reduces the probability of errors) especially that the data is analyzed later with *multi-level modelling*, which also increases power. The manipulation of factors/levels allows researchers to study the effect of changing security requirements on adequacy ratings, to identify dependencies, and to prioritize requirements based on the factor contribution to overall effect. For example, results of the multi-level modeling suggests that, although experts realize that displaying detailed error messages to end-users is an insecure approach that exposes internal vulnerabilities to hackers, their overall security ratings were slightly improved when the scenario had a stronger logging and monitoring mechanism. To close the security knowledge gap, the MQM approach also elicits new requirements from experts that has been experimentally shown to monotonically increase security.

Recruiting Real-Experts. Hoffman defines a *novice* to be the newcomer, someone with minimum exposure to the domain; and an *expert* to be the distinguished person who can handle tough cases and who has deeper experience in subdomains, which makes their judgments highly regarded by peers[2]. Based on this definition, and the results of our own analyses using patterns of situation awareness on expert interviews, we classify novices, and experts based on their number of years of industry experience. College students are often novices, because they have fewer years of experience, but they afford the benefits of a convenience sample that help a researcher discover new theory. Hence, we ran our earlier evaluation of the technique on 174 security students. In our later studies, where we build scenarios specific to professional security analysts, I recruited 69 experts with average 10 years of experience by attending a security conference. The scenarios for the expert study combines factors that require understanding of different security sub-domains: networking, databases, operating systems, and web applications.

Data-driven Approach to Modeling Expert Knowledge. Formal modelling of security knowledge is necessary to build a decision-support system (or an intelligent system in general). Security expert judgments will

[2]K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman, *The Cambridge handbook of expertise and expert performance*. Cambridge University Press, 2006.

always contain a degree of interpersonal uncertainty (two experts providing different judgments) and intrapersonal uncertainty (the same experts provides different judgments over different times). The uncertainty in the data is a characteristic that cannot be ignored assuming precision of expert judgment, but rather needs to be modeled, because it represents the diversity of opinions of experts. Because I want to model expert knowledge, with its interpersonal and intrapersonal uncertainty, I use type-2 fuzzy logic that can handle both. I also propose new approaches to build a type-2 fuzzy logic ruleset with reduced size, as I use the expert data to only keep realistic permutations of input/output, and exclude unrealistic permutations. My novel contributions to the fuzzy logic community is in the rigorous approach to eliciting and modelling real expert data, and in the application area of cyber security.

Funding and Service

My research has been funded by the NSA Science of Security (award #141333), the Army research office (Award #W911NF-09-1-0273), the Office of Naval Research (award # #N00244-16-1-0006, and #NOO244-17-0012), which illustrates the interest from different government agencies to fund this research. I also collaborated with Dr. Christian Wagner of University of Nottingham, and together with my advisor Dr. Travis D. Breaux, Dr. Stephen Broomell from CMU, and a group of researchers from the University of Nottingham; we received an award from the UK's Engineering and Physical Sciences Research Council (EPSRC) and the National Cyber Security Centre (formerly CESG) based on our proposal for a new data-driven cyber security system. The project's aim is to help organizations maintain adequate levels of cyber security through a semi-automatic, regularly updated, organization-tailored security assessment of their digital infrastructures[3]. I also co-authored a recent proposal with Dr. Breaux, to the Office of Naval Research that had just granted approval.

In addition, I served as a PC member for the 2016 International Workshop on Privacy Engineering (IWPE'16) collocated with 37thIEEE Symposium on Security and Privacy, and I was invited to review a paper for the IEEE Transactions on Software Engineering journal. I am also enthusiastically participating in the Requirements Engineering community and offered to serve in in the review committee of future RE conferences.

Future Research

In general, my research revolves around understanding and modeling human expertise in decision-making in software design. This problem has two main methodological aspects: empiricism and modelling.

For the empiricism aspect, I would like to advance empirical research in computer science. I would like to move away from traditional A/B testing and full factorial designs and develop innovative ways to conduct empirical research that sets foundations to design smarter tools. In addition, I would like to re-think our metrics in computer science empirical research. In modern day software, it is not sufficient to rely on designing a tool and then testing it on a convenience sample (such as college students) using metrics, such as task correctness and time performance. Completing a task correctly in a specific amount of time, could mean that the tool is simple to learn. However, by assuming that the completion time by itself is a surrogate measure of usability, one might raise concerns about construct validity. Interdisciplinary computer science research that uses methods and theories well-established in psychology and social sciences will have more benefit in the long run and advances our current software design practices. In the long run, such research will produce foundational theories and principles of software design rather than experimental observations specific to certain system implementations.

For the modelling aspect, I would like to explore data-driven modelling approaches that can fit data collected from experts in a domain, where the experts are scarce and the amount of data collected is not sufficient for machine learning approaches. I have used fuzzy logic, so far, and my research would benefit from examining other approaches, such as Markov Decision Processes (MDPs). I also like to explore possible modeling approaches that can handle uncertainty such as description logic and its fuzzy or probabilistic extensions. In addition, I would like to explore modelling differences between novices and experts and study how to support both groups in a decision-support system.

I believe that I am uniquely capable to handle the above challenges, based on my expertise in mixed qualitative and quantitative research methods.

[3] New Research Tackles Adaptive Security Decision Support: http://www.isri.cmu.edu/news/2017/0712-adaptsec.html