November 20, 2013 | By Tiffany Fox

Understanding Information: How We Get It, How We Use It, How to Benefit from It

Tara Javidi

For most computer users, information is only valuable when it serves a context-specific purpose, such as providing the GPS coordinates for a new restaurant or a list of search results for a query on airline flights to Fiji.

But for University of California, San Diego electrical and computer engineering professor Tara Javidi, understanding how people acquire and use information in various engineering applications is just as valuable. Her most recent grant from the National Science Foundation (NSF) — a $1 million collaborative research award to Javidi, Andrea Goldsmith of Stanford

University and Bruno Sinopoli at Carnegie Mellon University — will fund the development of a new theoretical framework for understanding how to best control information flow in large cyber-physical systems such as datacenters or smart energy grids.

The grant is the fourth NSF grant in a series of research projects Javidi has been awarded in the past 18 months, all dealing with questions of information acquisition in four contexts: cognitive networking, enhanced spectrum access, computer vision and social networking.

The main thrust of all four projects is to combine tools from control theory, information theory and statistics to jointly optimize data collection, analysis and processing.

"In most inference and learning applications, often the existence of a set of collected samples/data is assumed and the designers focus on how information can be mined and inferred from this data," said Javidi. She notes, however, that with the abundance of sensing technologies and the massive scale of many cyber-physical systems, "a decoupled process of data collection introduces a large degree of inefficiency. In particular, the problem of information collection cannot be ignored or taken for granted without introducing inherent loss of performance."

Javidi and her collaborators are taking a novel approach — trying to predict which sensing and data collection resources are best utilized, where, and when.

"With this new NSF project, we want to take that a step forward," added Javidi, who is affiliated with Calit2's Qualcomm Institute at UC San Diego, including its Information Theory and Applications Center, as well as UCSD's Center for Wireless Communications and Center for Networked Systems.

"Say I want to look at the data from a large electricity grid to predict failures (small or cascading) or even typical aspects of the network operation such as what time of day is best for doing certain energy-rich tasks," she explained. "To do so, I need to predict which piece of data is most informative, where and when, but also which piece of information can add reliability to such a prediction.

"Practically speaking this means that an efficient design must address how, when and where certain sensors in certain areas of the grid need to be turned on and which piece of information needs to be moved around in the network."

The problem, says Javidi, is that these types of large, cyber-physical systems are so ever-changing, and have so many sensors, that collecting all possible data typically results in a glut of information. This may be inefficient and overwhelming in terms of processing, and it also makes the data collection excessively expensive and resource-intensive.

"There are two sets of problems at the core of the proposed work: one is that of scale and decentralization of control, and the other one is the trade-off between cost, accuracy, and dimension of data," continued Javidi, who joined the Jacobs School of Engineering faculty in 2005. "In modern, large-scale, cyber-physical systems, it is not feasible for a central 'decision-maker' to regulate the information from the sensors. So what we need to do is control the information in a fairly localized fashion, so it's more efficient and scalable, and at the same time be careful about the nature of information exchange to ensure reasonably accurate global knowledge."

To address scalability and decentralized design, Javidi and her team will extrapolate from results of a simultaneous, three-year project awarded by NSF in July 2012, titled "Inference by Social Sampling." The motivation behind that research is to see what information about the world could be derived from a social network by asking certain questions or by looking for certain clues, such as how frequently a topic appears in online conversations.

"You look at the way your friends are interacting, and from that you try to determine what types of global behavior the network might see," she explained. "Here the critical issue is that in a social network, any individual user only has access to decentralized information held in a their local social circle. However, the users' fairly rich social connectivity allows for the information to be disseminated across the network globally, albeit at a slower time scale. In the design of cyber-physical systems we extend this line of work and reasoning to find out how we can design a mechanism through which global information sharing is achieved through local interactions."

In addition to studying these problems of network scale and decentralized control of information, Javidi is studying a second class of problems: the tradeoffs that arise when acquiring information. These tradeoffs chiefly include the effects on information acquisition speed, accuracy of the acquired information, and the cost of collecting the data.

Javidi notes that many problems in data acquisition today involve scenarios where an entity needs to collect information but also to control how much and what type of information it collects. "In fact, when gathering a limited amount of information, there is often an inherent

tradeoff between how fast information is acquired and how accurate the information is," explained Javidi. "Another challenge is you want to be smart about the measurements you're taking because the value of the information varies over time."

To answer the above question, Javidi and her former Ph.D. student, Mohammad Naghshvar, first carefully examined the challenges of active hypothesis testing and the difficulty of controlling uncertainty sequentially. Javidi says that initially these challenges "felt like a control problem, a resource allocation, sensor scheduling, problem. But then we noted that this really is an information theory/statistics problem."

They approached the problem by applying what they knew about how radios transmit knowledge over a channel (by coding and decoding the information) and applying it to the understanding of sensor scheduling, for example. They eventually cracked the problem by combining analytical tools from statistics, information theory and control theory, specifically stochastic control theory.

"We showed that while variants of the same problem have appeared in the literature for each of these fields, a combination of tools and viewpoints are necessary to fully understand the fundamental trade-offs in the problem of information acquisition," said Javidi. "We showed that an optimized process of information collection faces a fundamental tradeoff between speed and accuracy, which becomes further complicated by the cost and feasibility of data collection and sensor management." The results of this study, funded by NSF in 2010, are described in a forthcoming article, "Active Sequential Hypothesis Testing," to appear in the prestigious *Annals of Statistics*.

The study begun in 2010 and its findings have enabled Javidi and her collaborators to apply these theoretical concepts to a wide variety of practical and tangible applications of interest.

In an NSF-funded project begun in October 2012, titled "Enhanced Radio Spectrum via Information Acquisition and Learning," Javidi and her co-principal investigator at the University of Southern California are investigating the specific application of her theoretical work in the context of cognitive networking and wireless frequency spectrum management.

Similar issues arise, adds Javidi, when considering the sequential rendering and processing of images and videos. In a collaborative $1.2 million NSF research award to Maxim Roginsky, Svetlana Lazebnik from the University of Illinois at Urbana-Champaign and Javidi in July 2013, Javidi began extending and applying her work to feature-rich decision problems in computer

vision and cognition. These two ongoing projects complement and enhance the activities proposed in her most recent NSF grant in the context of cyber-physical systems, with the additional complication of dealing with scalability across large areas.

But Javidi argues that "this is only the tip of the iceberg," as the problem arises in ever increasing areas of engineering. "There has already been research into how to make accurate decisions with the data you have," Javidi continued. "That's not new. What's new with this research is now we're advocating that in a cyber-physical system – instead of mindlessly collecting all possible information and centralized processing of this information – we must, in a manner similar to what an experienced clinician often does, prioritize which information is most valuable over time and space, and how it can be collected and processed in a scalable and cost-effective manner."

UC San Diego's Studio Ten 300 offers radio and television connections for media interviews with our faculty, which can be coordinated via studio@ucsd.edu. To connect with a UC San Diego faculty expert on relevant issues and trending news stories, visit https://ucsdnews.ucsd.edu/media-resources/faculty-experts.