Information Sciences Institute

DELIVERING THE FUTURE

2018 ANNUAL REPORT

INNOVATION *with an impact on* SOCIETY, EDUCATION, SCIENCE *and* TECHNOLOGY

USC Viterbi
School of Engineering

INFORMATION SCIENCES INSTITUTE

is a world leader in research and development of advanced information processing, computing and communications technologies.

CALIFORNIA
4676 Admiralty Way #1001
Marina del Rey, CA 90292
310.822.1511

MASSACHUSETTS
890 Winter Street
Waltham, MA 02451
781.622.9790

VIRGINIA
3811 Fairfax Drive #200
Arlington, VA 22203
703.243.9422

Report Preparation Janice Wheeler (Editor); Craig Knoblock | Design Russo Design, Los Angeles | Printing: Typecraft, Inc.

USC Viterbi School of Engineering

Information Sciences Institute

WELCOME TO ISI

The USC Information Sciences Institute (ISI) is driven by the talent and drive of its people. In this year's annual report, we strive to capture the energy found across ISI, where researchers are working towards solving some of the world's biggest challenges using technology in efficient and creative ways.

One of the things that distinguishes ISI from a more traditional academic department is our emphasis on solving real-world problems. In this issue, you will read about ISI researchers who developed data-mining tools to help law enforcement agencies identify and prosecute human traffickers, methods for identifying fake news, an approach for predicting geopolitical events, techniques for detecting crypto-currency manipulation in social media, tools to defeat distributed denial-of-service attacks and techniques for modeling human behavior for cybersecurity and much more.

ISI is a unique research organization. Founded as part of USC, the institute was launched as a spinoff from the Rand Corporation in 1972. Starting with just a handful of researchers, ISI has grown to include more than 400 faculty, staff, post-docs and students across three locations in Marina del Rey, California; Arlington, Virginia; and Waltham Massachusetts.

ISI is an integral part of USC's research culture with faculty, staff and students of all levels working on a wide range of research problems. We currently have around 30 researchers with faculty appointments and 85 PhD students from a range of departments, including computer science, electrical engineering, industrial and systems engineering, astronautical engineering and spatial sciences.

I assumed the role of interim executive director of ISI in June 2018, when Dr. Prem Natarajan embarked on a two-year leave of absence to pursue other ventures.

After 28 years of working at ISI, most recently as director of the AI division, I am deeply honored to have the opportunity to lead this prestigious organization. I am thankful to Dr. Natarajan for his successful efforts in building a vibrant and thriving research organization.

During the five years that Dr. Natarajan led the institute, he grew full-time staff and PhD students by 75 people and significantly boosted research funding. In 2018 alone, ISI was awarded 42 new projects with a total award value of $75M.

As a result, my current focus is to continue to provide a stimulating research environment and recruit outstanding researchers and students. As you will read in the following pages, we are already well on our way.

Craig A. Knoblock
Michael Keston Interim Executive Director
Director, Data Science Program
Research Professor of Computer Science and Spatial Sciences

MARINA DEL REY, CALIFORNIA

With its main campus located in Southern California's Marina del Rey, Information Sciences Institute is in strategic proximity to the area's new Silicon Beach—home to components of key technology giants like Facebook, YouTube, Yahoo, Electronic Arts, and Google. Visitors to the Institute are also treated to a panoramic view of the Pacific Ocean and the sweeping Santa Monica mountains.

BOSTON, MASSACHUSETTS

Information Sciences Institute's Waltham Massachusetts campus is located close to the city of Boston and its array of highly ranked universities that include MIT, Harvard, Tufts, and Brandeis. It is also home to many high-tech companies involved in new, state-of-the-art research.

ARLINGTON, VIRGINIA

Information Sciences Institute's Arlington campus is located close to Washington D.C. and the campuses of major universities such as Cornell, George Mason, and Virginia Tech. It is also in close proximity to some of Northern Virginia's acclaimed centers of technology.

2018 NEW DIRECTORS

As we work towards the vision of the Information Sciences Institute as an ever-more influential and high-impact home for research in diverse areas across the information and computer sciences, it is appropriate to revisit our organizational structure and align it with emerging institutional needs and opportunities. To that end, during 2018 the Institute announced new organizational roles and associated personnel changes that will strengthen both the long-range perspective and the operational capabilities of the Institute, addressing key challenges to advancing ISI as a world-class research organization and positioning ourselves to achieve even greater success in the future.

CRAIG KNOBLOCK | INTERIM EXECUTIVE DIRECTOR, ISI

Craig has been with USC/ISI for twenty-eight years. Prior to serving in the capacity of interim executive director during 2018, Craig was the director of ISI's Artificial Intelligence Division, one of the United States' largest and most respected AI laboratories. In that appointment, Craig led a team of 100+ highly regarded researchers working on leading-edge projects—including machine learning, biomedical data integration, machine translation, knowledge graphs, and social networks. Craig holds additional positions at USC—that of research professor with the Department of Computer Science and the Spatial Sciences Institute. He is an AAAI Fellow, an ACM Fellow and Distinguished Scientist, and a senior member of IEEE.

ARAM GALSTYAN | DIRECTOR OF ARTIFICIAL INTELLIGENCE DIVISION

Aram has been with USC/ISI for nearly twenty years. This year he was appointed as ISI's director of the Artificial Intelligence Division. In this role, Aram will guide the research of ISI's largest division (over 100 scientists and researchers) in multiple research efforts related to AI. He also leads the Machine Intelligence and Data Science Group here at ISI and is a research associate professor with the Viterbi School of Engineering's Department of Computer Science. His current research focuses on different aspects of machine learning, with applications in computational social sciences, forecasting, bioinformatics, multimedia integration analysis, natural language processing, and statistical physics of learning and inference.

NEW ISI TEAM MEMBERS

2018 was another exceptional year for the Information Sciences Institute in many ways: Nearly $75M in new grants and collaborations from key federal agencies and from industry and academia; ISI's continuing sponsorship of the USC Computing Forum; ISI's hosting of the 9th IEEE International Conference on Biometrics—Theory Applications, and Systems; multiple awards and honors for ISI researchers including an IEEE Fellowship; and thirty new highly qualified directors, computer scientists, physicists, engineers, researchers and staff at ISI's facilities in Marina del Rey CA, Arlington VA, and Boston MA. We are pleased to announce the following new members of the ISI team.

ARLINGTON, VIRGINIA

TEAM MEMBER	TITLE	TEAM MEMBER	TITLE
Alexei Colin	Computer Scientist	Devanand Shenoy	Director
George Cooper	Research Programmer		

BOSTON, MASSACHUSETTS

TEAM MEMBER	TITLE	TEAM MEMBER	TITLE
Eric Berquist	Research Programmer	Oskar Singer	Research Programmer
Arun Jagannatan	Experimental Physicist	Claire Sun	Programmer Analyst I
Chester Palen-Michel	Research Programmer		

MARINA DEL REY, CALIFORNIA

TEAM MEMBER	TITLE	TEAM MEMBER	TITLE
Daniel Benjamin	Postdoctoral Fellow & Research Associate	Dongyu Li	Research Programmer
James Buff	Senior Software Engineer	Damon MacDougall	Project Manager
Keith Burghardt	Computer Scientist	Iacopo Masi	Research Computer Scientist
Rene Butcher	Programmer Analyst IV	Goran Muric	Postdoctoral Scholar & Research Associate
David DeAngelis	Senior Computer Scientist	Yesenia Ornelas	Financial Analyst
Thayer Diab	VLSI Engineering Technician	Lucy Panossian	Financial Analyst II
Amy Feng	Administrative Assistant II	Gracie Pisaric	Accountant I
Wendel Fleming	Systems Admin / DevOPS Engineer	Brendan Reid	Postdoc Research Associate
Michelle Hasan	Postdoctoral Research Associate	Gleb Satyukov	Research Programmer
Jamani King	Administrative Assistant II	Lincoln Thurlow	Research Programmer II
Jamie Kohls	Director of Security	Alexey Tregubov	Research Programmer II
Geoff Lawler	Research Programmer		

2018 RESEARCH GRANTS

During 2018, USC's Information Sciences Institute received over 40 new funded research awards. Sponsored by the federal government, academia and industry, these awards range in dollar size from thousands to millions and cover a myriad of topics from funding for student undergraduate research to large-scale projects that analyze images, speech, and text, and technology that accesses Internet outage measurements to help systems recover from natural disasters.

ISI PRINCIPAL INVESTIGATOR AND AWARD SPONSOR	2018 FUNDED RESEARCH AWARDS
Jose-Luis Ambite | NSF/Rutgers U.	PAGE III: Population Architecture Using Genomics and Epidemiology
David Barnhart | DARPA	Consortium for Execution of Rendezvous and Servicing Operations (CONFERS)
David Barnhart | NASA/JPL	JPL Student Undergraduate Research Proposal
David Barnhart | Private company	Short Duration Cubesat Mission for Communication Dynamics Research (Dodona)
David Barnhart | NASA/Parabilis	Parabilis Propulsion for STTR
David Barnhart | DARPA	Reactive Electro-Adhesive Capture (ClotH)
Terry Benzel | DHS	Cybersecurity Experimentation Testbed of the Future
Terry Benzel | NSF	DETER: Research Education & Operations Mission Sustainment (DREAMS)
James Blythe | DARPA	LAG-MD: Deep Learning Agents & Game Theory for Military Decision Making
Hans Chalupsky | DARPA/Carnegie Mellon U.	Operations-Oriented Probabilistic Extraction with Reasoning and Analysis (OPERA)
Stephen Crago | US Gov't-other	Lifelike Digital Human Replicas
John Damoulakis | USAF	Establishing a Venture Capital Program for the Micro-Electronic Innovation for National Security & Economic Competitiveness (MINSEC) Effort
Ewa Deelman | DOE/LLNL	Workflow Management for Integrating HPC with Big Data
Ewa Deelman | NSF	EAGER: Exploring & Advancing the State of the Art in Robust Science
Ewa Deelman | NSF	Cyberinfrastructure Center of Excellence Pilot Study
Ewa Deelman | NSF/U. No. Carolina	CICI: SSC: Integrity Introspection for Scientific Workflows (IRIS)
Ewa Deelman | NSF/U. No. Carolina	Integration: Delivering a Dynamic Network-Centric Platform For Data-Driven Science
Marjorie Freedman | DARPA/Raytheon BBN	Etiologist: General Knowledge from Heterogeneous Sources
Marjorie Freedman | DARPA/AFRL	PIRANHA: Preventing Information Removal and Nabbing Harmful Actors
Matthew French | DARPA/USAF	Mirage: A Security Metrics-Driven Obfuscation Design Environment
Aram Galstyan | DARPA	Discerning Group Biases in Online Communities via Linguistics Analysis
Aram Galstyan | US Gov't-other	Workshop on Sensemaking Technologies for Intelligence Analysis
Yolanda Gil | NIH	High Resolution Mapping of the Genetic Risk for Disease in the Aging Brain
Jonathan Habif | AF Office Scientific Research	Technical Forecasting Study: Quantum Information Sciences and Systems
Jonatham Habif | DoD	Demonstration of a Long-Wave IR Communication Link
Wes Hansford | DARPA	Automated Analog Mixed-Signals (AMS) Intellectual Property Generator for Complementary Metal Oxide Semiconductor (CMOS) Technologies
John Heidemann | DHS/AFRL	Planning to Anycast as Anti-DDoS
John Heidemann | NSF	RAPID: Interactive Internet Outages Visualization to Access Disaster Recovery (HOVADR)
John Heidemann | DARPA	Global Analysis of Weak Signals for Enterprise Event Detection (GAWSEED)
John Heidemann | DHS	Detecting, Interpreting, and Validating from Outside, In, and Control Disruptive Events (DIVOICE)
Itay Hen | DOE/Oak Ridge Nat'l Lab	Software Stack and Algorithms for Automating Quantum-Classical Computing
Itay Hen | AFRL	Solving Verification & Validation Problems with Quantum Samplers
Itay Hen | AFRL	Power of Non-Stoquastic Quantum Annealing Optimization
Craig Knoblock | USAF	SpaceAware: Multi-Source Data Fusion for Space Situational Awareness
Kristina Lerman | DARPA/Soar Technology	C-BOCA: Bias in Online Communication Activity
Jelena Mirkovic | NSF	Elements: Software: Distributed Workflows for Cyber-experimentation
Jelena Mirkovic | NSF	SaTC CORE: Hardening Systems Against Low-Rate DDoS Attacks
Prem Natarajan | DARPA/AFRL	Generating Alternatives for Interpretation & Analysis (GAIA)
Prem Natarajan | US Gov't/Johns Hopkins U.	JHU SCALE: Multilingual Text Search in Image Collections
Prem Natarajan | IARPA	SIGINT-Based Anticipation of Future Events
Shrivatsan Ravi | AF Office Scientific Research	Future of Autonomous Decision Making in Safety-Critical Cyber Environments
Stephen Schwab | DARPA	Dispersed Computing DETER Testbed Expansion and Experiment Support
Rafael Ferreira da Silva | NSF/UC Santa Barbara	Coordinating Curricula & User Preferences to Maximize the Participation of Women & Students of Color in Engineering
Ralph Weischedel | DARPA/AFRL	Hawkeye

REACHING OUT. . . .

Since its inception in 1972, USC's Information Sciences Institute has driven revolutionary advances in a wide variety of technological fields: machine learning and artificial intelligence, cybersecurity and high-performance computing architectures, quantum computing and heterogenous computing environments, data science and scientific workflows, health informatics and forecasting of societal and cyber events. However, the Institute also involves itself in numerous outreach activities that benefit society in general. These activities include hosting symposiums, workshops, and lecture series for the scientific community, and sponsoring undergraduate students from across the country, thereby providing these young men and women with an opportunity to participate in state-of-the-art research here at ISI.

USC/ISI COMPUTING FORUM INAUGURAL SYMPOSIUM: AI IN LIFE

Sponsored by Information Sciences Institute, the USC Computing Forum Inaugural Symposium was held on 17 May 2018 with the theme of AI in Life. ISI's Michael Keston Executive Director, Prem Natarajan, introduced the forum as a community for like-minded individuals to congregate, network, and exchange ideas in LA's technology epicenter of Silicon Beach.

An audience of more than 100 researchers, scientists, venture capitalists, startup founders and business leaders experienced a full day of thought-provoking discussions focused on the profound impact of AI in our lives and its potential to radically shape the future. Our keynote speakers and panelists explored four critically important topics:

AI for a better and safer society
AI in the media
AI in the enterprise
AI everywhere

The forum's first panel discussion, AI for a Better and Safer Society. (left to right) Christopher White, principal researcher, Microsoft; Aleksandra Korolova, USC Department of Computer Science; Jeff Alstott, program manager, IARPA; Milind Tambe, USC Center of Artificial Intelligence in Society.

NSF-FUNDED RESEARCH EXPERIENCES FOR UNDERGRADUATES (REU)

ISI's inaugural REU program participants hailing from across the U.S., pictured with their supervisors and ISI's REU leader, Jelena Mirkovic Photo/Caitlin Dawson.

The National Science Foundation funds a large number of research opportunities for undergraduate students through its REU Sites program. An REU Site consists of a group of undergraduates who are selected to participate in research programs within a host institution. Students are associated with a specific research project, where they work closely with faculty and other researchers.

In the summer of 2018, ISI hosted such a group for an eight-week program targeting research projects with the theme of "human communication in a connected world." While some of the students had a passion for cybersecurity, others focused on data analysis, natural language processing, and Internet protocols, This year's participating students were pursuing degrees at Illinois Institute of Technology, University of Puerto Rico, Wartburg College, Macalester College, Oberlin College and UCLA.

Quoting ISI's site leader Jelena Mirkovic, "Working on real-world research projects gives undergraduates a feel for the day-to-day experience of being a researcher and a leg-up in graduate school. Through this program, students from a range of academic and geographic backgrounds have the opportunity to make lasting connections with faculty, researchers, and other students, helping them to develop networks that could influence their future educational pursuits and careers in technology."

2018 AWARDS & HONORS

AWARDS

YINING CHEN, SORCHA GILROY, ANDREAS MALETTI, **JOHATHAN MAY** AND **KEVIN KNIGHT**
NAACL 2018 Outstanding Paper Award | "Recurrent Neural Networks as Weighted Language Recognizers"

AMIR HASHMABNI, JERRY LI, AND **KEN JOHNSON**
ISI Meritorious Service Award | to Computing and Information Services team members for their extraordinary efforts in executing an important government-mandated cybersecurity compliance project

ULF HERMJAKOB, JONATHAN MAY, AND **KEVIN KNIGHT**
ACL Best Demo Paper Award | Out-of-the-Box Universal Romanization Tool

CRAIG KNOBLOCK AND **PEDRO SZEKELY**
USC Viterbi School of Engineering Use-Inspired Research Award | for their joint contributions in the area of semantic data integration

CRAIG KNOBLOCK
IJCAI Donald E. Walker Distinguished Service Award | for recognizing individuals who have selflessly contributed to the advancement of the AI community through sustained, exceptional service.

XIANG REN
2018 Amazon Research Award | for his work in neural-symbolic deep learning for natural language processing

GREG VER STEEG
ISI Outstanding Achievement Award | for developing scalable machine-learning methods that are both theoretically grounded and useful to practitioners in a variety of domains, such as cancer biology, neuroimaging, and social sciences

2018 Amazon Research Award | for his work in invariant representation learning

HONORS

TERRY BENZEL
Appointed to NSF's Directorate for Computer, Information Science and Engineering Advisory Committee

EWA DEELMAN
Elected Fellow of the Institute of Electrical and Electronics Engineering (IEEE) in recognition of her contributions to scientific workflow management

STEPHEN SCHWAB
Senior Fellow Applied Computer Security Associates (ACSA)

STUDENTS | POSTDOCTORAL & PH.D GRADUATES

POSTDOCTORAL SCHOLARS	
Daniel Benjamin	Judgment and decision-making focusing on human forecasting
Benjamin Girault	Signal processing on graphs
Michelle Hassan	Wellness and productivity in the workplace
K. S. M. Tozammel Hossain	Data mining and machine learning techniques
Homa Hosseinmardi	Computational methods to uncover hidden trends in human behavior
Goran Muric	Machine learning for complex networks
Brendan Reid	Quantum heat engines, quantum non-locality and reason, state preparation in complex optical lattice structures

PH.D. GRADUATES	ADVISOR	RESEARCH AREA	CURRENTLY AT
Shyyang Gao	Aram Galstyan	Machine learning	Amazon
Marjan Ghazvininejad	Kevin Knight	Machine translation	Facebook AI Research
David Kale	Greg Ver Steeg	Machine learning	Netflix
Kuan Liu	Prem Natarajan	Algorithms and applications of machine learning	Google
Nima Pourdamghani	Kevin Knight	Machine translation	Saykara
Xing Shi	Kevin Knight	Machine translation	Didi Labs
Liang Zhu	John Heidemann	Machine translation	Microsoft

DOCTORAL STUDENT	RESEARCH AREA	ADVISOR
Sami Abu-El-Haija	Machine learning	Aram Galstyan & Greg Ver Steeg
Nada Aldarrab	Machine translation	Jonathan May
Nazanin Alipourfard	Machine learning/data science	Kristina Lerman
Calvin Ardi	Network measurement and data sharing	John Heidemann
Victor Ardulov	Human-centered machine learning	Shrikanth Narayanan
Shushan Arkaelyan	Machine learning/data science	Aram Galstyan
Adam Badawy	Machine learning & political science	Emilio Ferrara
Guillermo Baltra	Network management	John Heidemann
Nathan Bartley	Artificial intelligence	Kristina Lerman
Karel Bogomir Mundnich Batic	Signal processing/human-centered machine learning	Shrikanth Narayanan
Brandon Booth	Human-centered machine learning	Shrikanth Narayanan
Robert Brekelmans	Machine learning/representation learning	Aram Galstyan
Yu-An Chen	Dynamic resource allocation in real-time systems	Stephen Crago
Jiaxin Cheng	Computer vision	Prem Natarajan
Ashok Deb	Data science & cybersecurity	Emilio Ferrara
Xiyue Deng	Malware analysis	Jelena Mirkovic
Tu Mai Anh Do	Workflow systems in HPC environments	Ewa Deelman
Mehrnoosh Mirtaheri Feijani	Machine learning/data science	Aram Galstyan
Tiantian Feng	Human-centered machine learning	Shrikanth Narayanan
Sahil Garg	Machine learning/machine extraction	Aram Galstyan
Sarik Ghazarian	Machine learning & dialogue modeling	Aram Galstyan
Mahboobeh Ghorbani	Knowledge graphs	Pedro Szekely
Majid Ghazemi Gol	Knowledge graphs	Pedro Szekely
Thamme Gowda	Machine translation	Jonathan May
Palash Goyal	Machine learning	Emilio Ferrara
Hang Guo	Network measurement & security	John Heidemann
Lalit Gupta	Physics/computational computing	Itay Hen
Tom Halverson	Physics/quantum computing	Itay Hen
Rujan Han	Artificial intelligence	Aram Galstyan
Hrayr Harutyunyan	Machine learning	Aram Galstyan & Greg Ver Steeg
I-Hung Hsu	Speech recognition	Prem Natarajan
Che-Wei Huang	Affective computing	Shrikanth Narayanan
Di Huang	Artificial intelligence	Emilio Ferrara
Bisilael Imana	Network measurement & privacy	John Heidemann & Aleksandra Korolova
Zoe Gonzalez Izquierdo	Physics/quantum computing	Federico Spedalieri
Aysuh Jaiswal	Representational learning/adversarial learning/ Multimedia integrity analysis	Prem Natarajan
Wuxuan Jiang	Machine learning/differential privacy	Aram Galstyan

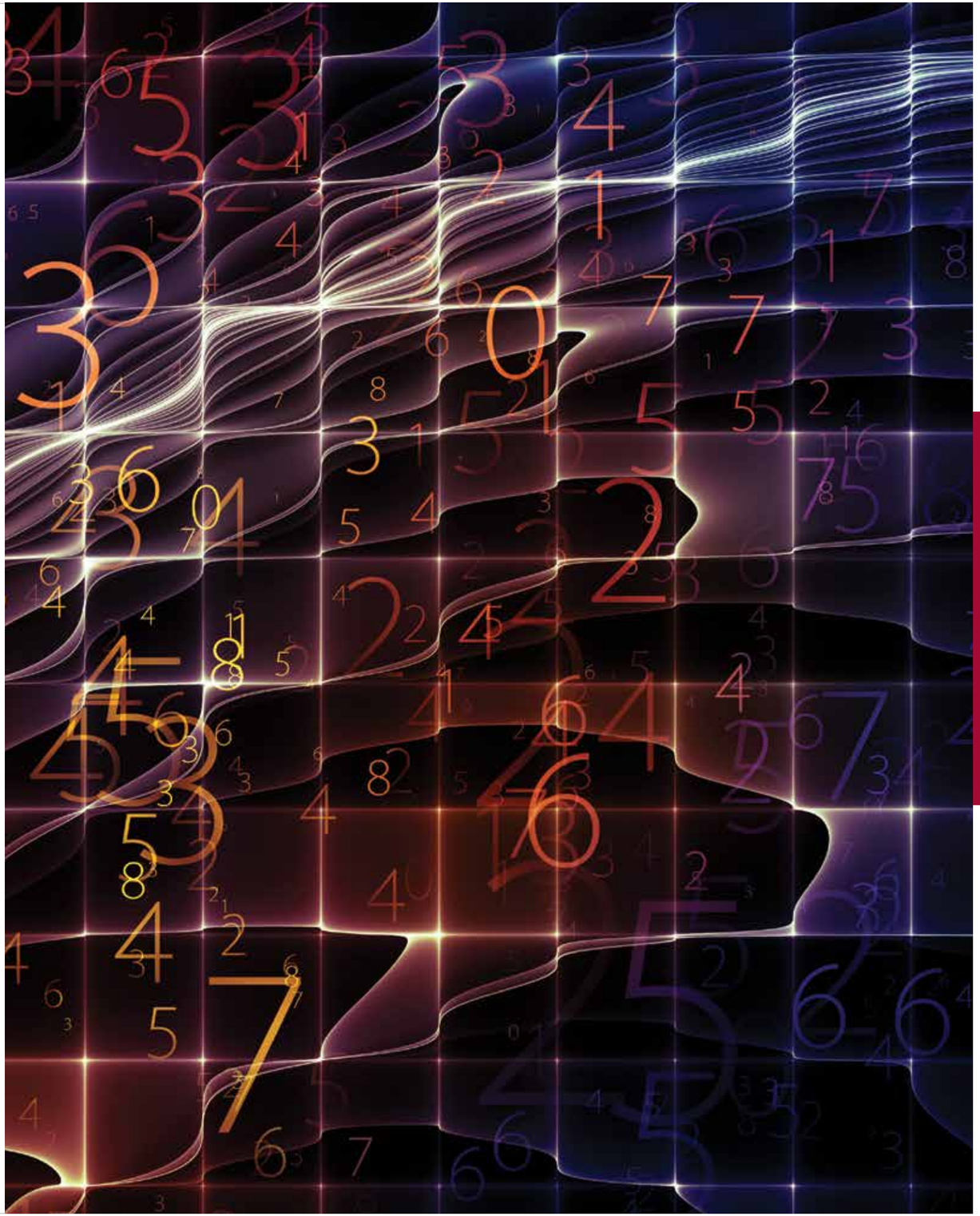
DOCTORAL STUDENT	RESEARCH AREA	ADVISOR
Hsien-Te Kao	Deep learning	Emilio Ferrara
Neal Lawton	Artificial intelligence	Greg Ver Steeg & Federico Spedalieri
Hannes Leipold	Quantum computing	Federico Spedalieri
Wenzhe Li	Machine learning/graph analytics	Aram Galstyan
Zekun Li	Artificial intelligence	Prem Natarajan
Jeremy Liu	Machine learning/quantum computing	Ke-Thia Yao
Nikolaos Malandrakias	Low resource natural language processing	Shrikanth Narayanan
Akira Matsui	Artificial intelligence	Emilio Ferrara
Ninareh Mehrabi	Artificial intelligence	Aram Galstyan
Negar Mokhberian	Artificial intelligence	Kristina Lerman
Daniel Moyer	Neuroscience & machine learning	Greg Ver Steeg
Amrutha Nadarajan	Signal processing/audio/biosignals	Shrikanth Narayanan
Soumyaroop Nandi	Computer vision	Prem Natarajan
Victor Martinez Palacios	Natural language processing	Shrikanth Narayanan
George Papadimitrious	Execution & anomaly detection	Ewa Deelman
Pavlos Papadopoulos	Human-centered machine learning	Shrikanth Narayanan
Raghuveer Peri	Signal processing/audio/speech	Shrikanth Narayanan
Minh Pham	Data transformation & knowledge graphs	Craig Knoblock
Nima Pouradamghani	Machine translation	Jonathan May
Abdul Qadeer	Network measurement & large-scale data processing	John Heidemann
Ehsan Qasemi	Knowledge graphs	Pedro Szekely
Sivaramakrishnan Ramanathan	Network monitoring & management leveraging	Jelena Mirkovic
Peri Raghuveer	Artificial intelligence	Shrikanth Narayanan
Kyle Reing	Machine learning	Aram Galstyan
A.S.M Rizvi	Network measurement & security	John Heidemann
Rahul Rughani	Satellite swarms	David Barnhart
Ekraam Sabir	Multimedia integration analysis	Prem Natarajan
Nripsuta Saxena	Artificial intelligence	Kristina Lerman
Basel Shbita	Knowledge graphs and the semantic web	Craig Knoblock
Ruhollah Shemirani	Artificial intelligence	Jose-Luis Ambite
Emily Sheng	NLP for scientific literature analysis	Prem Natarajan
Yuan Shi	Learning to adapt to sensor failures	Craig Knoblock
Jason Slepicka	Knowledge graphs	Pedro Szekely
Serban Stan	Artificial intelligence	Aram Galstyan
Dimitrios Stripelis	Large-scale data integration	Jose-Luis Ambite
Nandi Soumyaroop	Artificial intelligence	Prem Natarajan
Kexuan Sun	Artificial intelligence	Satish Thittamaranahalli
Rajat Tandon	Distributed denial of service	Jelena Mirkovic

STUDENTS | DOCTORAL (CONTINUED)

DOCTORAL STUDENT	RESEARCH AREA	ADVISOR
Nazgol Tavabi	Artificial intelligence	Kristina Lerman
Geoffrey Tran	Fault tolerance in cloud-based analytics	Stephen Crago
Ruchi Travadi	Signal processing/audo/speech	Shrikanth Narayanan
Colin Vaz	Signal processing/audio/speech	Shrikanth Narayanan
Binh Vu	Data representation & data integration	Craig Knoblock
Lan Wei	Network measurement & anycast	John Heidemann
Nicolaas Weidemann	Binary analysis	Christophe Hauser & Jelena Mirkovic
Xin-Zeng Wu	Network science/physics	Kristina Lerman
Hong Xu	Artificial intelligence	Craig Knoblock
Shen Yan	Machine learning	Emilio Ferrara
Xusen Yin	Dialogue and interaction	Jonathan May
Yilei Zeng	Artificial intelligence	Emilio Ferrara
Xiaofan Zhang	Physics/quantum computing	Itay Hen

VISITING SCHOLARS AND RESEARCHERS

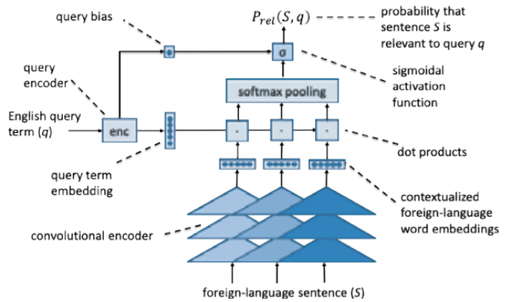
SCHOLARS AND RESEARCHERS	RESEARCH AREA	ISI HOSTS
Manuel Ciosci | Aarhus Univ., Denmark	*Information encoded in word representations*	Kevin Knight
Elizabeth Crosson | Univ. of New Mexico	*Quantum computing*	Itay Hen
Shreya Goyal | Indian Inst. of Technology, India	*Natural language processing*	Prem Natarajan
Michael Jarrett | Perimeter College, Georgia State Univ.	*Quantum computing*	Itay Hen
Hrishikesh Kale | Viterbi India summer program	*Machine learning*	Prem Natarajan
Amir Kalev | Univ. of Maryland	*Quantum computing*	Itay Hen
Devang Khamar | Virginia Polytechnic Institute	*FPGA hardware security*	Travis Haroldsen
Gaurav Kolhe, Ph.D student | George Mason Univ.	*Hardware security*	Vivek Venugopalan
Kasra Koorehdavoudi | Washington State Univ.	*Networking & cybersecurity*	Shrivatsan Ravi
Milad Marvian | Massachusetts Institute of Technology	*Quantum computing*	Itay Hen
Shreyas Mishra | Univ. of Massachusetts Amherst	*Networking & cybersecurity*	Srivastan Ravi
Subhee Rawal | Univ. of Maryland College Park	*Website design*	Matthew French
Jack Raymond | D-Wave Systems	*Quantum computing*	Itay Hen
Yiwen Shen | Columbia Univ. NY	*Optically disaggregated computing systems*	Ryan Goodfellow
Shri Shruthi Shridhar | Viterbi India summer program	*Machine learning*	Prem Natarajan
Dhanush Srinivasa | Univ. of Southern California	*Hardware security*	Joshua Monson
Peter Young | UC Santa Cruz	*Quantum computing*	Itay Hen

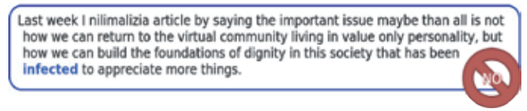


CENTER FOR VISION, IMAGE, SPEECH AND TEXT ANALYTICS (VISTA)

FOUNDING DIRECTOR **PREM NATARAJAN,** CO-DIRECTORS **WAEL ABD-ALMAGEED** AND **SCOTT MILLER**

The Information Sciences Institute's Center for Vision, Image, Speech and Text Analytics (VISTA) is an internationally recognized leader in areas such as natural language processing, computer vision, biometrics, optical character recognition, face recognition, speech and text analytics, and multimedia forensics. Our researchers contribute to several nationally influential research programs addressing some of today's most pressing challenges in these areas—from improving security to unmasking fake news, and optimizing knowledge extraction. VISTA researchers are located across the Institute's three sites—Marina del Rey CA, Arlington VA, and Boston MA—contributing to the following key programs:





Cross-Lingual Information Retrieval and Summarization

With increasing democratization of electronic media, vast information resources are available in less frequently taught languages such as Swahili, Tagalog and Somali. That information, which may be crucially important and not available elsewhere, can be difficult for monolingual English speakers to effectively access. Under IARPA's MATERIAL program, the ISI-led SARAL team is developing technologies which 1) search foreign language repositories of text and audio using English queries, 2) summarize the retrieved documents in English with respect to a particular information need, and 3) provide complete transcriptions and translations as needed.

Shared Embedding Architecture for Effective Retrieval (SEARCHER)

ISI's SEARCHER project is a key element of the SARAL approach, performing cross-lingual information retrieval and returning relevant foreign-language documents in response to English queries. Unlike traditional methods that rely on translating queries into the foreign language or, alternatively, translating the documents into English, SEARCHER translates neither. Instead, it utilizes neural networks that map both the queries and documents into a shared embedding space and performs retrieval there.

Cross-Lingual Query- and Domain-Focused Summarization

Critical to effective cross-lingual information retrieval is presenting the results to the monolingual users in their own language. To this end, ISI has developed a query- and domain-focused summarization system that allows English speakers to quickly judge a foreign-language document's relevance to a query (e.g., "*British colonies*" or "*bank* as in *financial institution*") or a domain (e.g., *Government* or *Health*).

Language and nationality like rice and pilau **bananas*** and river
**bananas, banana, plantain, plantains*

SPACE ENGINEERING RESEARCH CENTER (SERC) | DIRECTOR **DAVID BARNHART**

2018 saw a large growth in the Astronautics Department's Space Engineering Research Center (SERC) at ISI, with multiple projects across a gamut of disciplines, enabling over 30 new students to be involved in a variety of exciting space application projects.

The second year of technical support to CONFERS continued—this time increasing the scope of evaluation to look at all functions/attributes that encompass on-orbit servicing (OOS). CONFERS is the premier commercial consortium looking to set the standards and guidelines for on-orbit servicing worldwide. The SERC project team also evaluated the large constellations planned for deployment in the next 5-10 years to update the orbital spatial density plots to present to the commercial consortium members, while advising them of safe or less-populated orbits for early OOS demonstration missions.

The third year of the Jet Propulsion Laboratory's (JPL) Space Swarm project was funded, with SERC continuing to support the docking interface design and supporting rendezvous and docking for a trio of three CubeSats to build an aperture "on orbit."

The SERC received an award with a small commercial firm, Parabilis Space Technologies, under a NASA STTR effort. The SERC team began working on electronic, mechanical and control inputs for design considerations to support the integration and design input on a unique nanosatellite orbit transfer vehicle propulsion system.

The SERC received two other major space projects in 2018—REACCH and Dodona. The Reactive Electro Adhesive Capture ClotH (REACCH) project is creating a unique "octopus" robotic end affector-based gripper with multiple tendrils that use the unique properties of electro-adhesion (EA) as the "contact" points. The project is in conjunction with JPL who is creating the small EA "tiles."

The third USC-based nanosatellite, named Dodona (after the second-largest Oracle in Greek mythology), was awarded by Vector Corporation to support a payload qualification flight of their new technology. The SERC faculty and student team took a "sister satellite" from the original Aeneas CubeSat built and launched in 2012, modified its software, and upgraded various components from a nonfunctional to a functional status.

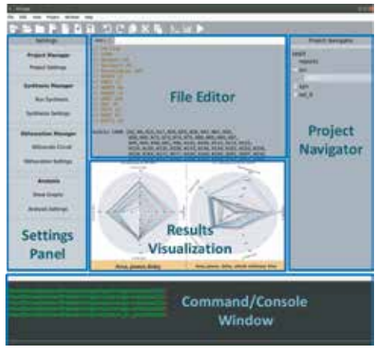
As a part of the Dodona effort and the ongoing graduate class on campus, the SERC led an upgrade to the USC Satellite Ground Communications station's UHF/VHF antennas. These specific antennas had not been fully operational, and are planned to be part of the Dodona satellite ground operations with Vector upon launch.

Members of the SERC participated as subject matter experts in ISI's Computational Systems and Technology Division's project "SpaceAware," a unique effort that couples data fusion capabilities of the ISI-developed KARMA model to the space domain. The entire team project is meant to create a unique set of standalone software that is able to pull in the latest information on every satellite identified in open depositories (like SpaceTrack and Celestrack) and identify their unique characteristics for heuristics and evaluation of current and future actions.

At the end of the year, SERC was awarded an internal ISI Keston Research Grant entitled "Satbotics Control: How to Merge Biologically Inspired Spacecraft Together." The grant will provide support to multiple graduate students to develop a new computational architecture applied to cellular space morphology, and to demonstrate this on the SERC's internal 3-DOF air-bearing testbed, beginning in 2019.

THE SECURE AND ROBUST ELECTRONICS CENTER (SURE) | DIRECTOR MATTHEW FRENCH

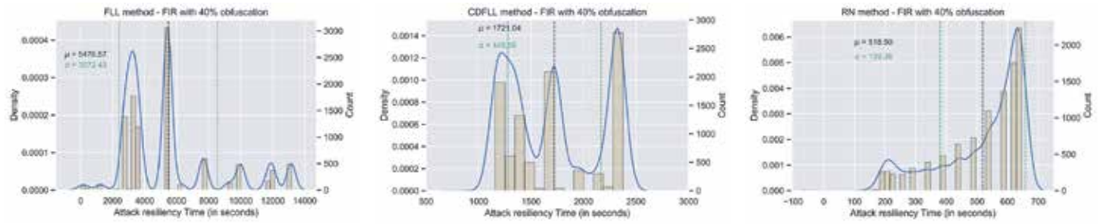
Hardware security is an emerging research theme that has risen out of the complexity of the developmental process and supply chain associated with microelectronics. Entire multi-billion-dollar ecosystems have evolved around the development of appliciation-specific integrated circuits (ASICs)—including third-party intellectual property hardware cores, CAD tools, verification tools, etc. These examine whether or not the device operates as intended, but do not provide trust or 'beyond functional verification' to determine if additional, unwanted functionality, such as hardware Trojans, are present in the device. If a device is produced in a trusted manner, it increasingly encounters more complex attacks in the field, where counterfeiting and reverse engineering are becoming more prevalent for stealing intellectual property. Hardware security now often overlaps with resiliency and reliability, where the now-famous Row Hammer attack illustrated how a reliability vulnerability could be exploited to circumvent security protections.



The SURE Center recognizes that modern-day effective hardware needs to comprehensively address trust, security, resiliency, and reliability, and is unique in its ability to address all four areas. In 2018, SURE had major developments in the areas of trust and counterfeit detection.

The SURE team, in conjunction with faculty from USC's Department of Electrical and Computer Engineering, developed a toolset called *Mirage* under the DARPA Obfuscation Manufacturing for GPS program to address IP leakage and hardware Trojan insertion during fabrication at a potentially untrusted foundry. Mirage is developing a CAD toolset which makes obfuscation a first- class design parameter. Mirage leverages state-of-the-art circuit obfuscation techniques developed in the community, determines the optimal places to position the obfuscation on full-sized, real-world designs, provides unified overhead metrics, and includes a protection time security metric in terms understandable to an end user. Mirage also uses Monte Carlo-based modeling to 'future cast' how resilient a given obfuscation type is likely to be in the future as de-obfuscation attacks become more sophisticated.

The counterfeiting of electronics is also a large problem—especially for government systems. Government systems can have development cycles of 5 to even 15 years, with an average platform lifetime of another 15 years. It is not uncommon for commercial electronics to sunset during a mission's lifetime, sometimes even during the development stage. So called 'life-time buys' of the electronics needed are not always sufficient. This situation creates a lucrative demand for obsolete parts, from which counterfeiters seek to profit. To address this, the SURE team is researching and developing the Independent FPGA Functional Test (IFT) platform for Xilinx FPGAs. This tool provides extreme, exhaustive testing of all configurable elements (routing, slices, Block RAM, etc.) of the device to independently determine if it is counterfeit. IFT addresses challenges in scaling to the over 380 million routing settings in an FPGA through intelligent packing, parallelism, and partial runtime reconfiguration to keep the testing time reasonable.



Data charts for comparison of attack resiliency time

USC-LOCKHEED MARTIN QUANTUM COMPUTING CENTER
DIRECTORS DANIEL LIDAR, ROBERT LUCAS, AND FEDERICO SPEDALIERI

Faculty, researchers and students are performing basic and applied research into noisy, intermediate-scale quantum (NISQ) computing devices, and are collaborating with researchers around the world. The USC-Lockheed Martin Quantum Computing Center (QCC) houses a D-Wave 2X quantum annealing system, manufactured by D-Wave Systems, Inc. QCC was the first organization outside of D-Wave to house and operate its own system, and it has conducted pioneering research on three different generations of these early NISQ processors. Operating quantum computing systems is demanding: the temperature of these systems needs to be kept near absolute zero (-273 degrees Celsius), and the devices must be electromagnetically shielded to protect the fragile quantum states from degradation by external noise. The main thrust of the research conducted at QCC has been to understand how this noise can adversely affect the computational power of these devices.

ISI CENTER FOR COMPUTER SYSTEMS SECURITY | DIRECTOR CLIFFORD NEUMAN

ISI's *Center for Computer Systems Security* conducts research and supports educational programs in the crucial disciplines of computer, network, and application security and privacy. Center staff develop new architectures for isolation in networked systems and study the cyber-resilience of critical infrastructure and Internet of Things devices. The Center's current research activities encompass work on blockchain, cryptocurrencies, election systems, the power grid, oil and gas systems, and malware analysis.

In addition to its research activities, Center researchers are frequently called upon by the media to explain events involving privacy, cybersecurity, and cybercrime. The Center director has recently been interviewed on television news, radio, and quoted in print media on topics ranging from ransomware attacks on hospitals and transportation systems, the dark web, privacy implications of high-tech gadgets, security for driverless vehicles, and major data breaches and privacy lapses that are of concern to the public.

Center staff lead several education programs at USC in computer security and privacy. They have redesigned the Master of Science degree in cybersecurity engineering to apply the foundations of security from high-assurance systems to the Internet and cloud-focused computing environments now in demand by consumers. They are currently developing a class on security, privacy, and policy in the age of the Internet for the National Academy of Engineering's grand challenge scholar's programs that will be taught concurrently in linked classrooms through the iPodia platform at universities internationally. This class is focused on both technical and legal considerations for computer security and privacy as operated across political and jurisdictional boundaries on the Internet.

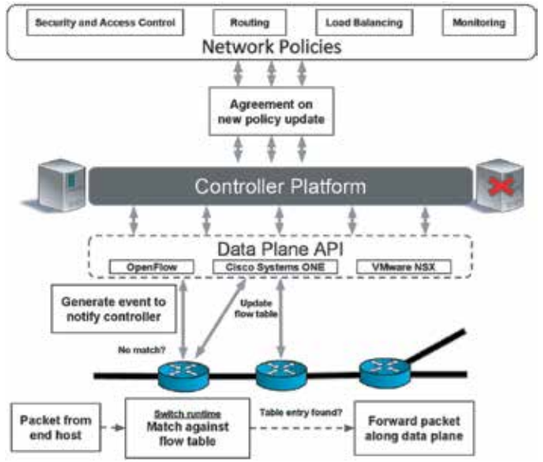
NORTHROP GRUMMAN CYBERSECURITY RESEARCH CONSORTIUM

Northrop Grumman's Cybersecurity Research Consortium (NGCRC), founded in 2009, is a groundbreaking partnership of industry and academia formed to advance research, facilitate collaboration among the nation's top scientists, and accelerate solutions to counter the fast-changing threats from cyberspace. The consortium addresses some of the world's leading cyber problems, including attribution in cyberspace, supply chain risk, and securing critical infrastructure networks. Members of the consortium coordinate research projects, share information and best practices, develop curricula, write joint case studies, and provide numerous learning opportunities and applications for students and the defense community overall.

USC's Viterbi School of Engineering joined the Northrop Grumman Cybersecurity Research Consortium in 2013, with Information Sciences Institute serving as the lead organization for establishing and advancing the partnership and expanding the consortium's breadth of investigation into the most pressing cyber threats to the economy and national security. ISI brought with it a strong reputation for leadership in big data, cybersecurity, computer science, and informatics. The other three partners in this research consortium are also leading cybersecurity research universities—Carnegie Mellon University, Massachusetts Institute of Technology, and Purdue University.

Through NGCRC, ISI researchers are developing a capability for retroactive cross-site sharing of cybersecurity datasets, effectively replaying the relevant portions of data after a determination is made regarding which details are most relevant, and while still complying with policies on data sharing. In other work with NGCRC, ISI researchers are collaborating with Viterbi's Energy Institute and the Center for Energy Informatics to deploy tools to assess operational resilience in critical infrastructure (the smart grid and oil and gas distribution), and develop game-theoretic models to mitigate the impact of attacks on these systems.

Current Project: Distributed Algorithmics for Robust Network Controllers



System architecture for resilient network provenance: The logically centralized controller platform allows dynamic installation of new network policies on the network switching fabric and aggregating data plane state. Observe that some subset of the participating controllers may be potentially malicious or misconfigured and may arbitrarily fail by crashing. The participating controllers "agree" on a new rule to be installed on the switch via a blockchain that ensures integrity and confidentiality of new updates as well as provenance on the current data plane state for forensics and diagnostics.

This research concerns the design and implementation of a scalable and secure distributed controller platform which allows network administrators to concurrently program the data plane without worrying about the synchronization problems that may arise due to malicious/misconfigured controllers, data plane switch/link failures, network asynchrony or modifications to control group membership.

Additionally, our research investigates a novel application of blockchains leveraging the distributed controller platform to perform fine-grained network data plane provenance, thus providing crucial forensics and diagnostics vital to network administrators.

Network administrators operating complex distributed networks often need to answer a diagnostics question querying the distributed state of the network elements—e.g., state of the routing tables—and take appropriate action. However, maintaining the correct distributed state of the network data plane in the presence of Byzantine adversaries is a fundamental challenge that must be addressed towards characterizing the most accurate evolving state of the network data plane. Network administrators typically consider the Byzantine threat model which assumes a completely untrusted environment encompassing a wide range of faults and misbehavior—e.g., cases where a malicious adversary has compromised some of the nodes and switches, but also more benign faults, such as hardware failures or misconfigurations. Moreover, network administrators are currently leveraging software-defined networking (SDN) as an architectural construct by outsourcing the control over the data plane to a logically centralized control plane, thus allowing the control plane to express and compose network policies of varying networking applications and translating these to rules installed onto the switches for handling network flows.

With traditional networking infrastructures, network operators are forced to continuously reconfigure policies in order to respond to a wide range of network events and applications. The idea of programmable networks that allow dynamic programming of network devices via an open interface has thus gathered a considerable amount of attention over the past few years. The advent of SDN has opened up the ability of dynamically deploying several applications—ranging from traffic engineering, inter-domain routing, security and access policy to network virtualization—all of which are vital to existing network infrastructures. A fully centralized system simply cannot adequately provide the required levels of availability, responsiveness and scalability, thus the need for a distributed control plane. This work models the problem of interaction between the data plane and a distributed control plane consisting of a set of failure-prone and potentially malicious control devices, and implements a secure and robust provenance platform that allows network administrators to integrate and verify new network functionality and perform network data plane forensics.

DIRECTOR **STEPHEN CRAGO**

ISI is one of the world's leaders in computing technologies, ranging from basic to applied research. Our Computational Systems and Technology (CS&T) Division focuses on:

- Heterogeneous cloud and embedded computing
- Microarchitecture, integrated circuits, and advanced electronics
- Quantum computing and information
- Reconfigurable computing and wireless networks
- Science automation technologies
- Spaceborne and ground-based processing
- System software
- Trusted and secure electronics and computing

FROM THEORETICAL TO HANDS-ON

Our current initiatives include theoretical adiabatic quantum computing through the USC-Lockheed Martin Quantum Computing Center and hardware security through ISI's Secure and Robust Electronics Center. CS&T projects include system software for heterogeneous clouds and hardware-software design of unique chips and field-programmable gate arrays. We're also exploring applications and algorithms to help process large-scale and real-time streaming data and to solve challenging optimization problems and system engineering for space systems.

CS&T teams are creating wireless networking and edge computing technologies for battlefields and other difficult environments, along with social media platforms for people who lack trustworthy Internet access. We're advancing our scientific automation tools that enable researchers to focus on conducting science, not managing data—and which are already used by astronomers, physicists and earthquake specialists, including the LIGO team that was the first to detect the gravitational waves that Einstein predicted.

INTELLECTUAL LEADERSHIP, DIVERSE EXPERTISE

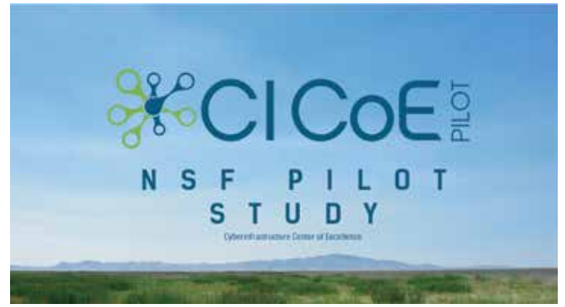
We provide intellectual leadership within the broader research community through participation in workshops like the White House OSTP National Strategic Computing Initiative Workshop, the IEEE Rebooting Computing initiative, and trusted electronics workshops run by the SURE Center.

Our 540 computing technology researchers, research programmers and graduate students represent a wide range of disciplines, including electrical engineering, computer science, physics, and math. CS&T researchers are based in Arlington, Virginia, Marina del Rey, California, and Boston, Massachusetts.

Cyberinfrastructure Center of Excellence Pilot

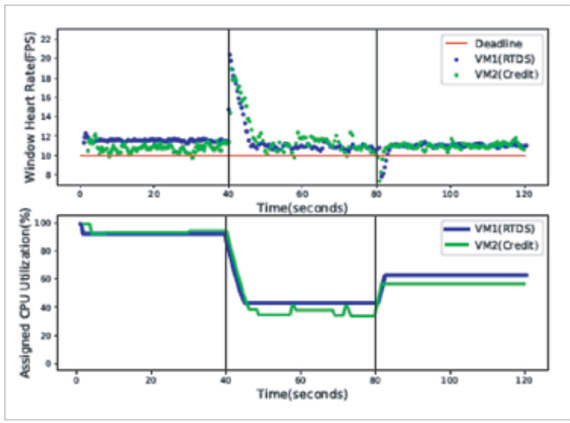
NSF's major multi-user research facilities (large facilities) are sophisticated research instruments and platforms—such as large telescopes, interferometers and distributed sensor arrays—that serve diverse scientific disciplines from astronomy and physics to geoscience and biological science. Large facilities are increasingly dependent on advanced cyberinfrastructure (CI)—computing, data and software systems, networking, and associated human capital—to enable broad delivery and analysis of facility-generated data. As a result of these cyber infrastructure tools, scientists and the public gain new insights into fundamental questions about the structure and history of the universe, the world we live in today, and how our plants and animals may change in the coming decades.

To enable the community building, sharing, and applying of knowledge of best practices and innovative solutions for facility CI, ISI researchers lead a new pilot project for an NSF Cyberinfrastructure Center of Excellence (CI CoE) which aims to support sharing of best practices, software solutions and architectures coming from over 20 of NSF's largest facilities, including NEON, LSST, LIGO and IceCube. Collaborators on the project are the University of North Carolina's Renaissance Computing Institute, the University of Utah, Indiana University and the University of Notre Dame.



Real-Time Cloud and Fog Computing

Although cloud computing has exploded in popularity over the past five years, it suffers from two key challenges: high latency and lack of fine-grained resource management. The high latency to the cloud traditionally prevented real-time applications from depending on a cloud backend. The lack of fine-grained resource management means that workloads cannot burst vertically into additional resources, but instead must burst horizontally, a more coarse and costly approach to cloud scaling. ISI's active research targeting dynamic resource management allows workloads to dynamically scale multi-modal workloads in real-time, adding and subtracting resources to maintain required performance levels. ISI is in discussions with potential transition partners to put its resource management framework into production. At the same time, USC/ISI is extending this work to the fog computing paradigm. Fog computing places cloud-like resources near the computational edge, enabling low-latency communication between edge devices and fog resources. Ultimately, this combination will allow low latency real-time computation for highly dynamic workloads in both public and private clouds The figure shows two scheduling approaches, and in both cases the resource manager adjusts CPU allocation as the workload transitions between modes. This allows more workloads to be packed into fewer computational resources.

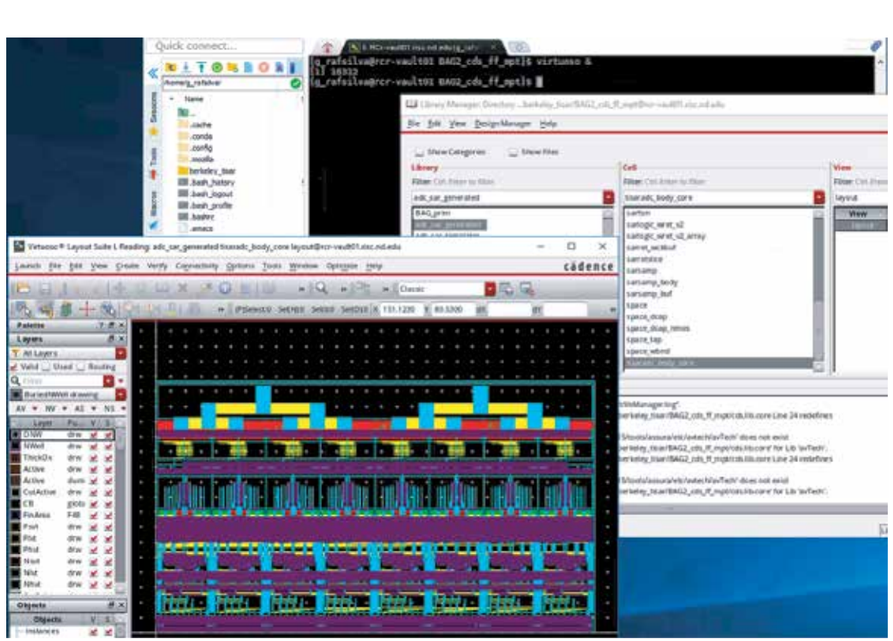


ISI's cloud resource management framework improves utilization by adjusting CPU resources as applications transition between modes of operation. This enables multi-modal real-time applications to successfully execute within cloud and fog platforms.

RACE: Enabling ASIC Design Collaborations in a Secure Environment

Application-specific integrated circuit (ASIC) design flows often face high barriers for adoption (even by experienced design teams) due to a handful of difficulties that include acquiring expert knowledge of new tools, which may involve installing, configuring the environment, and using them; access to new models and libraries (e.g., process design kits, PDKs), which may require solving licensing issues; and managing the execution of design flows in a collaborative manner, which becomes challenging when individual (sub-)teams have their own working environment. To tackle such challenges, leading computer-aided design (CAD) companies targeted the development of virtual private environments for design teams; however, they are tied to a specific vendor, limiting open access collaboration across institutions and adoption of alternative CAD tools.

In partnership with the US Department of Defense (DoD) CRAFT program, and in conjunction with faculty from USC's Department of Electrical and Computer Engineering and researchers from the University of Notre Dame, we have developed the CRAFT Secure Vault, a system that enables sharing of design flows, IP, and best practices across design teams while protecting all information about specific designs being carried out by individual design teams. The Secure Vault streamlines the process of acquiring tools and IP from multiple vendors, installing and configuring these tools on their computing platform, and learning a new design flow using bulky user manuals. The Secure Vault is a secure collaborative chamber which gives users direct access to ASIC design tools and data engineering capabilities, regardless of geography, so that team members can work in a predictable environment that securely defines who has access to what. It also avoids the lengthy process of acquiring tools (and their dependencies), IP, system setup, and administration. Empowered by Blockchain technology, the CRAFT Secure Vault also disposes of automated mechanisms to provide a non-repudiable audit trail of what happened and when. The Secure Vault is hosted on an AWS GovCloud operated by University of Notre Dame, and is controlled unclassified information (CUI) ready. It will be initially used to support the multi-project wafer (MPW) runs for the DARPA CRAFT program.

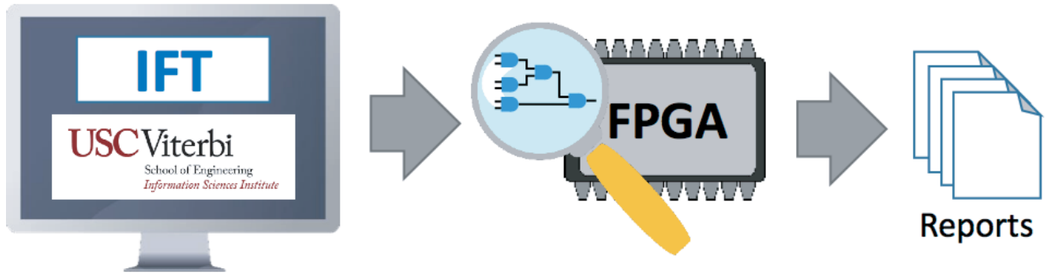


Sample ADC Generator produced within the CRAFT Secure Vault

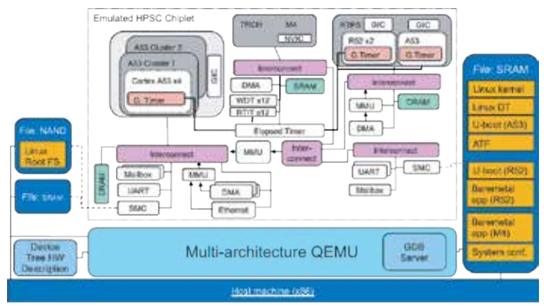
Counterfeit Detection in Field Programmable Gate Arrays

Field programmable gate arrays (FPGAs) are used in many systems that are critical to our infrastructure and national defense. As these systems age, device wear-out may require procurement of new devices from untrusted third-party resellers who may supply used or counterfeit parts in place of authentic chips. ISI is developing a suite of test designs to provide independent functional validation of commercial FPGAs. This validation is an important first step in establishing a trusted supply-chain, determining the usability of devices stored in inventory, and monitoring the health status of fielded systems.

As FPGAs grow in size to over 1 billion transistors, exhaustively testing the millions of logic gates and routing interconnect resources on the chips requires highly scalable solutions. In addition, each family of FPGAs may have dozens of different sized parts requiring a flexible testing solution. The ISI counterfeit detection tools make use of a detailed understanding of the FPGAs, exploit the repeatable structures found on FPGAs, and utilize a collection of internally created CAD tools to create a scalable, finely-tuned suite of test designs to identify faults in a large portion of the underlying FPGA VLSI. ISI and its sponsors are using this test suite to find faults in the devices which may indicate a part has experienced age-related wear-out or be a counterfeit. ISI has developed these tests for the Artix/Kintex/Virtex 7 families of Xilinx FPGA devices and is currently working to port the tests to other families of FPGAs.



ISI's independent functional testing reports any faults in the logic and wires of a field programmable gate array device that may be detected during testing.



ISI's heterogeneous software emulator for the NASA HPSC chiplet

Enabling High Performance and Reliable Processing in Space

Our space agencies increasingly demand high performance and reliable computational resources in space. High-performance on-board processing is an enabling technology for the next generation of Earth observing and deep space missions, both of which require high-performance processing to support autonomy and AI for mission success. ISI has a rich history in high-performance space processing with programs such as the Maestro processor and NASA's High-Performance Spaceflight Computing (HPSC) Chiplet. ISI leads the system software development for both processors. The primary research efforts common to these programs are: 1) architecture modeling to enable software development years ahead of hardware availability, 2) developer tools that enable the efficient use of highly parallel and/or heterogeneous processing elements, 3) software-implemented fault tolerance to allow autonomous detection and correction of failures, and 4) software capable of self-checking the hardware for errors and failures. This figure shows the emulated HPSC Chiplet which is already being used to enable early software development.

Flux-Based Quantum Speedup (FLUQS)

In 2017, ISI began work on the Flux-Based Quantum Speedup (FLUQS) project, sponsored by IARPA under the Quantum Enhanced Optimization (QEO) program. Led by USC, the project is a large collaboration between academic, research and private institutions. It aims at both designing and building the next generation of quantum annealing devices, with the specific goal of enhancing the computational tools required for tackling optimization problems. ISI scientists have taken a leading role on several thrusts of this project that are related to the validation of quantum behavior in the proposed devices, and our understanding of the role quantum mechanics may play in any possible computational speedup. The following represents the highlights of some of the research results we developed as part of this project.

Benchmarking of quantum annealers with optimal annealing times

While the D-Wave quantum annealing processors have grown steadily in size from 4x4 Chimera unit cells in the first generation Rainier processor in 2011 to the current 16x16 unit cells in the DW2000Q processor in 2017, a rigorous assessment of the performance scaling of the processors relative to algorithms run on classical hardware has not been possible. No previous experiments have exhibited an optimal anneal time, which is necessary for determining the true performance scaling with problem size. We have overcome this obstacle by developing a new class of problem instances that utilize small unit cell gadgets that push the optimal annealing time into a range achievable by the processor. Our work demonstrated the first verifiable advantage for these processors over simulated annealing but not over other classical algorithms. We were able to show that the culprit for the processors' relatively poor performance remains its suboptimal temperature, and we are working with collaborators to investigate techniques for overcoming these limitations.

The role of non-stoquastic interactions in quantum annealing

One of the key objectives of the FLUQS project is to determine what might be the necessary ingredients for the next-generation quantum annealer to provide a quantum advantage. One tantalizing prospect that we are investigating is the introduction of non-stoquastic interactions to the annealing protocol. These exotic interactions are appealing for two reasons: non-stoquastic interactions are necessary for universal adiabatic quantum computing, and they prevent efficient simulation of the annealing protocol using quantum Monte Carlo. However, our research shows that generically such interactions make the annealing protocol less efficient; only for special problem instances have advantages been demonstrated to date. Our objective is to better understand how and when such interactions can be used to achieve true quantum advantages.

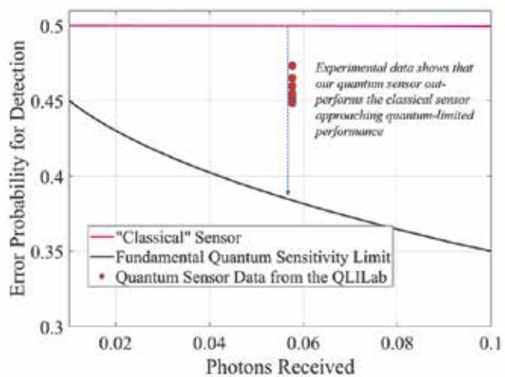
Laser Speckle Contrast Imaging

From mobile phones to airport security checkpoints, biometric systems such as fingerprint scanners and face-recognition systems play a pivotal role in authenticating a person's identity. The universal presence of biometric systems and its responsibility in protecting highly sensitive private information makes it extremely vulnerable to hacking. The Quantum-Limited Information Laboratory at ISI-Boston has developed optical systems based on laser speckle contrast imaging (LSCI) to prevent spoofing attacks on biometric systems. LSCI was originally developed by the biomedical community to non-invasively detect blood flow in intravascular regions of the retina and brain. LSCI involves the imaging of a speckle pattern, which is formed when a laser beam is scattered from an optically rough surface. From the speckle pattern, the presence of blood flow can be detected by computing the speckle contrast, which depends on the velocity of the scattering medium. Biometric systems are typically hacked using 3D printed fingers or face masks, which lack blood flow and can be easily detected by employing LSCI. Moving forward, the project aims to incorporate machine learning principles to enhance the performance of LSCI-based optical systems.

Quantum Limited Optical Sensing

Sensing electromagnetic radiation is a commonplace task in a diverse set of application domains and technologies—including imaging, astronomy and radar systems. Traditionally, the sensitivity limit of a sensor is determined by either noise coming from background sources, or physical shortcomings of the sensor itself, such as intrinsic noise or physical limitations. None of these limitations to sensitivity are *fundamental*, however. They are simply a result of engineering limitations. When the signal contains just a few photons of light, these traditional sensor techniques are incapable of extracting the fragile information contained in that signal. In the Laboratory for Quantum-Limited Information (QLIlab) at ISI-Boston, as part of the Computational Systems & Technology Division, we are investigating the fundamental limits to our ability to measure and extract information from signals, which is determined by the laws of quantum physics. Our team uses quantum theory to calculate this limit and compare the fundamental sensitivity achievable with "quantum sensors" against that which is achieved with state-of-the-art "classical" sensor technology. Evaluating the improvement achievable with a quantum sensor enables us to design and build experimental demonstrations in the laboratory that ***beat*** the sensitivity of classical sensors and approach the *fundamental quantum limit* for *sensitivity*.

A prime example of such an experiment is shown below, where we aim to determine the fundamental quantum limit for discriminating between light that is naturally occurring (such as that from the sun) and light that comes from a laser. While this is a project rooted in fundamental physics, the resulting technology developed can have an application in domains ranging from threat detection for airborne systems to the search for extraterrestrial civilizations.



(left) Experimental work taking data from the quantum sensor experiment in the QLIlab. (right) Experimental data from the quantum sensor measurement showing that our quantum sensor has sensitivity beyond a "classical" sensor. Continuous improvements allow us to approach closer to the fundamental quantum sensitivity limit.

Decision Systems

System Engineering for Space

In 2018 the Decision Systems Group continued expanding its research base into the space sciences—particularly in the area of improving systems engineering and operations as applied to both the space and ground segments. A particularly challenging aspect of the space domain is the ability to produce robust solutions in a timely matter. Traditional approaches to acquisition rely on defining requirements up front, followed by design, development, integration, testing and then deployment—the so-called Waterfall method. Such approaches are rigid and incapable of adapting to changing system requirements and the availability of new technologies. This rigidity often results in cost overruns and delayed product releases that typically result in obsolete and error-prone systems being deployed. In 2018, the team introduced novel Agile and DevOps approaches to current systems acquisition efforts—including the introduction of new metrics for measuring development performance and new approaches to system engineering to handle acquisition efforts that combine both Waterfall and Agile/DevOps. Much of the effort undertaken in 2018 was in support of the successful launch of the first Block III GPS Satellite (SV01) in December 2018.

Block III GPS SV01 Launch — Photo Credit: GeekWire

BICOR

ISI completed the National Institute of Standards and Technology (NIST)-funded *Biomedical Devices and Equipment Consortium Organization to Roadmap Industry (BIOCOR)* project in 2018. BIOCOR was a collaborative initiative focused on enhancing US competitiveness in the biomedical devices and equipment (BDE) sector, and the resulting increase in high-paying high-tech jobs, by encouraging US manufacturing of advanced medical devices. Led by USC/ISI, the team included USC's Keck School of Medicine, Columbia University and the University of Minnesota. Throughout the three-year BIOCOR effort, the team met with key members of the BDE industry, undertook multiple workshops and surveys and participated in multiple conferences and symposiums focused on the BDE industry. The focus of this effort was to determine and prioritize key challenges faced by the BDE industry. From this exploration, ten key challenges were identified along with three focus areas to pursue in the near-term: 1) increase research into digital health; 2) provide small and medium-sized companies access to the latest advanced manufacturing technologies (e.g., automation, molding, 3-D printing, etc.); and 3) provide enhanced training and education—both at the academic and professional level. Included in education and training is access to subject matter experts that provide advice on the latest manufacturing technology, the current and future regulatory environment, and steps to transition from device concept to market-ready product.

BDE INDUSTRY CHALLENGES AND PRIORITIES

- The role of big data in the industry.
- Security and privacy in digital health.
- Support for low-volume manufacturing.
- Need for flexible/adaptable devices.
- Regulatory control—difficult to comply and very difficult to stay current—especially for small manufacturers.
- Workforce sustainability—need highly skilled and motivated workforce. Continuing education is a major challenge.
- Lack of access to the latest technology due to high-cost barrier. This is particularly a problem for small manufacturers.
- Access to funding for R&D and product development.
- Protection of intellectual property (IP).
- Rapid prototyping to help reduce R&D costs. This includes virtual prototyping.

Backpack

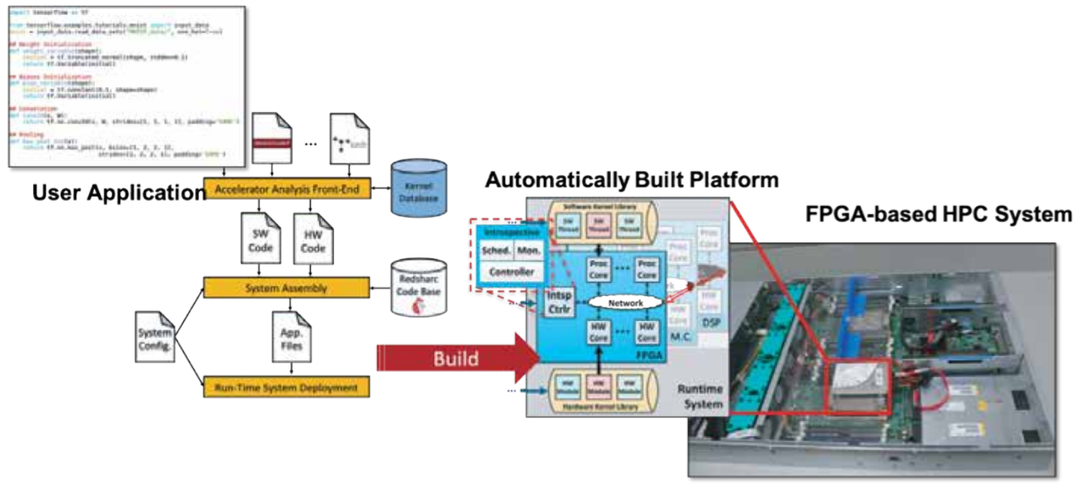
ISI's Backback project is exploring how to enhance communications in societies with restricted or tightly controlled communications. *Backpack: Offline Content Distribution Through Mobile Phones*, expands upon the success of the mobile Android application toolset to accommodate additional modes of data transfer and to develop and deploy an Apple iOS version of the application. During 2018, the team developed a prototype for a mobile/desktop platform which allows seamless content sharing over Wi-Fi or LAN, expanding on its initial use of Bluetooth. A working prototype is already proving simpler to use than current solutions in the "offline content sharing" or "offline educational app" markets.

Performance and Productivity Tools for Applications in Heterogenous Computing

The Reconfigurable Computing Group (RCG) in the Computational Systems & Technology Division has a long history of bridging the gap between end users and heterogenous computing resources. Early work includes the development and release of an open-source toolkit called Tools for Open-source Reconfigurable Computing (TORC) that allows users to apply optimizations within the development flow and customize their own electric design automation (EDA) computer aided design (CAD) tools in ways that vendor tools do not support. More recently, the research focus has shifted to even higher levels of abstraction, where our goal is to support scientists and developers who are unfamiliar and understandably not interested in the low-level expertise usually required to fully realize the true performance potential of the compute resources. The RCG team has developed tools and frameworks to solve these complex productivity and development problems.

Our research has led us to the development of the Hot & sPyC tool suite which enables Earth scientists to target field programmable gate arrays (FPGAs) by writing algorithms in high-level languages (i.e., Python) in minutes to days compared to traditional hardware development timeframes of months to years. The techniques combine source-to-source translation and high-level synthesis (HLS) to convert high-level code written by scientists into highly optimized hardware accelerators. The accelerator cores are then combined with system-on-a chip IP that has been developed for a specific FPGA platform in order to maximize performance while minimizing development time. While initially developed for remote sensing applications, Hot & sPyC is broadly applicable to other domains such as optical signal processing, software-defined radio, and video processing.

Currently deployed techniques for machine learning and artificial intelligence leverage graphics processing units (GPUs) due to performance and relative ease of development. However, there is broad interest in executing machine learning and AI algorithms on FPGAs due to their superior power efficiency and ability to support embedded platforms. Our group is investigating methods to continue raising the abstraction level at which FPGAs can be efficiently programmed by combining machine language toolkits, such as Google's TensorFlow, with techniques such as those used in Hot & sPyC. Such tools would enable AI researchers to directly target cloud-based FPGA resources, such as in AWS-F1 or Microsoft's Brainwave System, or embedded IoT or edge-computing platforms, without needing a detailed understanding of the underlying hardware.



The Reconfigurable Computing Group has developed tools and frameworks easing complex productivity and development

DIRECTOR **ARAM GALSTYAN**

The Artificial Intelligence Division (AID) comprises more than 100 faculty, research staff, USC graduate students, and short- and long-term visiting researchers. Most AID researchers hold graduate degrees in computer science or related disciplines, and many also serve as research faculty in the USC Viterbi School of Engineering—mainly in the Department of Computer Science.

ISI's Artificial Division (AID) is one of the world's largest artificial intelligence (AI) groups. The Division is known especially for its work in natural language processing, machine translation, and information integration. We also explore biomedical data integration and engineering, computational behavior, adaptive robotics, social networks, and video, image and multimedia analysis. We build working prototypes and partner with academia and industry to create commercial applications. These are AID's primary research thrusts.

NATURAL LANGUAGE PROCESSING AND MACHINE TRANSLATION for which AID is internationally renowned; this includes statistical machine translation, question answering, summarization, ontologies, information retrieval, poetry generation, text decipherment, and more.

MACHINE LEARNING AND DATA SCIENCE focusing on developing efficient algorithms to analyze data from a variety of applications areas, including biomedical sciences, computational social science, and cybersecurity.

KNOWLEDGE GRAPHS using artificial intelligence and machine learning techniques to construct large-scale knowledge bases, with applications ranging from combating human trafficking to predicting cyberattacks.

KNOWLEDGE TECHNOLOGIES involving interactive knowledge capture, intelligent user interfaces, semantic workflows, provenance, and collaboration, with a focus on scientific data analysis and discovery.

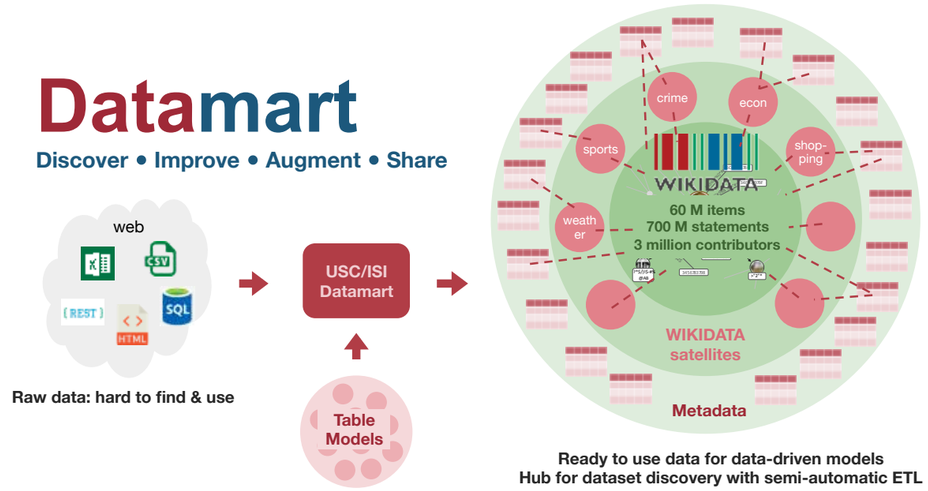
VIDEO, IMAGE AND MULTIMEDIA ANALYSIS including document image processing, face recognition, and biometrics.

BIOMEDICAL DATA INTEGRATION & KNOWLEDGE ENGINEERING providing efficient access to distributed and heterogeneous biomedical data, and developing biomedical informatics systems based on cutting-edge AI techniques.

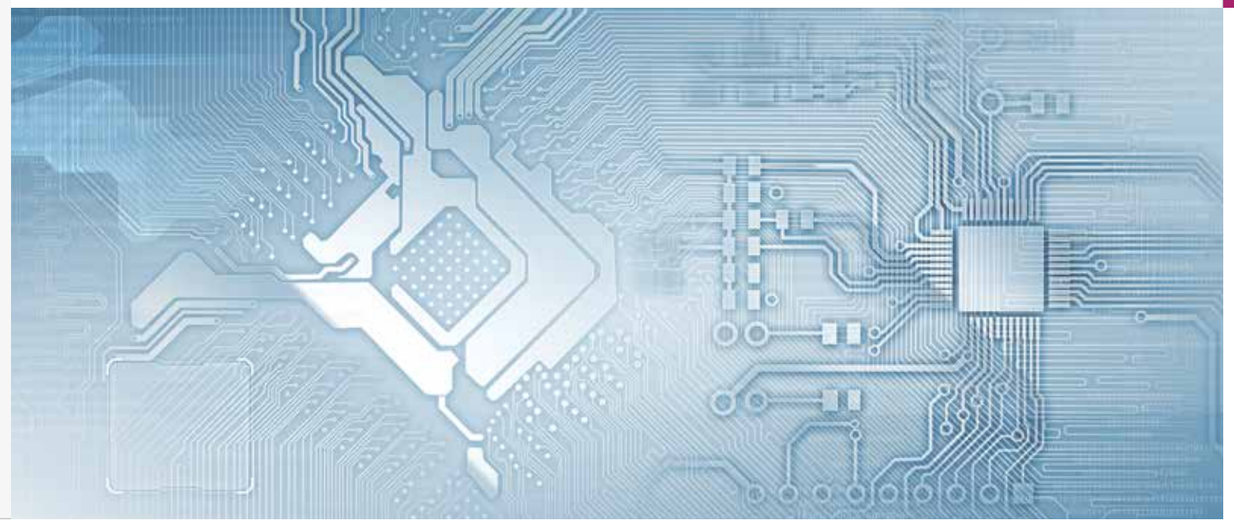
ROBOTICS in particular modular, self-reconfiguring robots and control methods.

Datamart: An Index of Structured Data in The Web

Machine learning models can always be improved by using more data. The goal of the Datamart project is to help data scientists find structured data on the web to complement the datasets used in a model. We are developing techniques to enable search-by-dataset, where a user submits a dataset as a query, and the system returns a collection of datasets that can be used to add features (new columns) or training data (new rows). We are investigating discovery methods that use a combination of search engine queries and focused crawling to identify web pages with links to potentially relevant datasets, information extraction techniques to extract useful metadata from these pages, table understanding methods to identify the variables and values in a data file, and data profiling techniques to build a scalable index containing rich metadata to answer the search-by-dataset queries. The results of these queries are ranked lists of metadata files that Datamart can materialize to retrieve the latest version of the data, reformat to construct a clean tabular representation of the data, and fuzzily join to add columns or rows to an existing dataset.

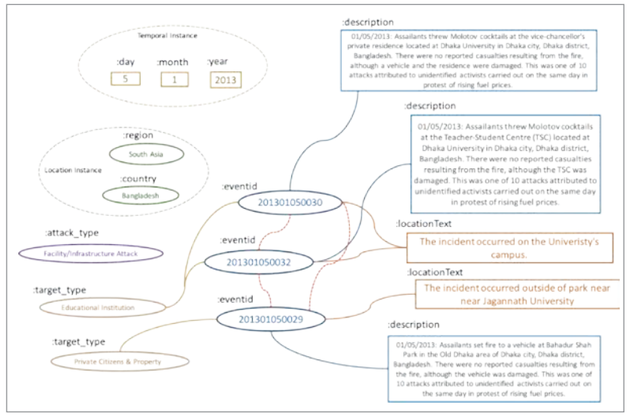


Helping scientists find structured data on the web to complement the datasets used in a model



Complex Entity Resolution

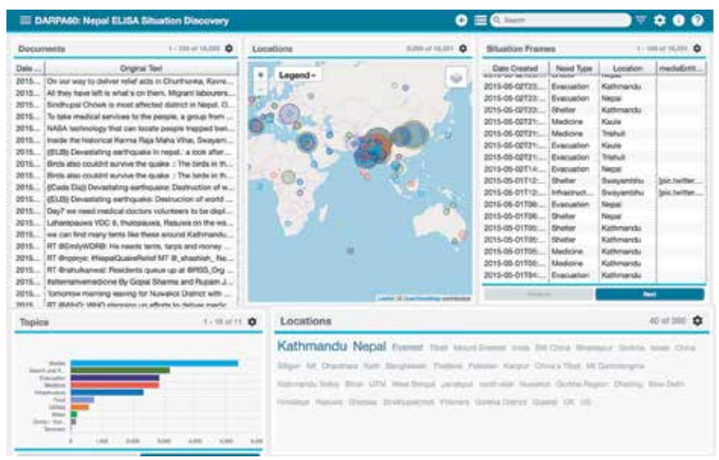
In recent years there has been an increased focus on representing, reasoning over, and conducting inference on events. Unlike ordinary named entities, events are complex data structures, embodied by artifacts like triggers, actors, locations, descriptions and spatiotemporal information sets. In the case of events with geopolitical consequences, such as terrorist attacks, assassinations, or bombings, automatically and accurately predicting links between events is an important research application that can be used to populate and enrich geopolitical, sparse (and proprietarily gathered) knowledge bases with global scope. In ongoing work at ISI, researchers are building a complex entity (aka 'event') resolution system called DeepEvent that relies on advanced representation learning to output typed links between events. DeepEvent was evaluated on the Global Terrorism Database (GTD) and outperformed at least three rival baselines.



Automatically and accurately predicting links between events

Intelligent Crisis Informatics

According to the United Nations Office for the Coordination of Human Affairs (OCHA), in 2016 more than a hundred million people were affected by natural disasters alone, while over sixty million people were forcibly displaced by violence and conflict. Consequently, crisis informatics has emerged as an important interdisciplinary area, with contributions from both social and computational sciences, including machine learning, information retrieval, natural language processing, social networks and visualization. The *Text-Enabled Humanitarian Operations in Real-time* (THOR) system, built in collaboration with Next Century Corporation, is an example of an intelligent crisis informatics platform that was featured in DARPA's 60th anniversary in 2018, and is capable of advanced analytics.



Output from intelligent crisis informatics platform

Natural Language Processing: uroman

In the event of a humanitarian crisis or natural disaster, the natural language processing group at ISI has been focusing on the rapid development of software and systems to allow first responders to more easily help victims communicating in another language. As the language is not known a *priori* and speed is critical, it is important to develop tools for all languages that can be deployed at a moment's notice and without careful thought about the specific language being handled.

We developed *uroman*, a universal romanizer, that can represent any Unicode sequence in a form pronounceable by English speakers. Making such representations is critical for monolingual English speakers attempting to process data written in an unfamiliar script. It can even lead to comprehension without machine translation, as names that are pronounced similarly across languages can be more easily identified when they are romanized. Our team won the award for Best Demo at the 2018 ACL Conference in Melbourne, Australia, for our presentation of *uroman*.

	Original	Romanization
Amharic	በርሊን የጀርመን ዋና ከተማ ነው።	bareline yajaremane waanaa katamaa nawe.
Arabic	المملكة العربية السعودية	almmlka al'rbya als'wdya
Greek	Γερούν Ντάισελμπλουμ	Geroun Daiselbloum
Hebrew	עזרת תורה בירושלים	'zrt tvrh vyrvshlym
Japanese	アメリカ	amerika
Korean	세계에서 6번째로 면적이 넓은 나라이다.	segyeeseo 6beonjjaero myeonjeogi neolbeun naraida.
Mandarin	北卡罗来纳	beikaluolaina
Nepali	तिब्बती भाषामा यसको नाम चोमोलुङ्गमा हो।	tibbatii bhaassaamaa yasako naam comolunggamaa ho .
Tamil	இதன் தலைநகராகச் சென்னை உள்ளது.	itan talainakaraakac cennai ullatu.
Tibetan	ཤ་སྤྲོང་ཁྱེར	lha·sa·grong·khyer

Examples of language handling by the uroman universal romanizer

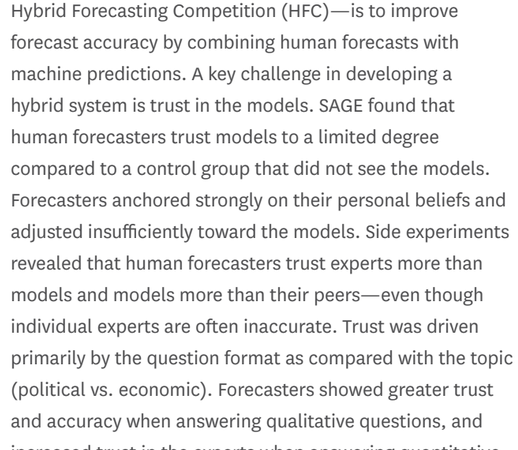
Using Structured Data to Verify Textual Assertions

The rise of "fake news" and misinformation campaigns have had dramatic consequences in the social and political landscape. While a number of approaches have been proposed to use linguistic or social cues to identify attempts to manipulate readers, factually validating assertions still remains a challenging, open research problem. We are developing systems to analyze text to identify quantifiably verifiable statements, which are those that describe a measurable pattern in a dataset. We have formulated this problem as one of mining quantitative indicators and associated trends. Our two-pronged effort involves using commonsense knowledge bases to determine which phrases are most likely to be measurable indicators and then using lexical features, part-of-speech tags, and parse trees to find broad class indicator-trend patterns. For each trend, we are developing methods to match a phrase such as "plummeting" to a set of functional constraints on the feature-representations produced by a time series analysis model. Using this framework, we can translate a news story into a set of mathematical tests on datasets, which ultimately allow us to determine the extent to which the sentences capture truthful and accurate information. We are applying this approach to validate cause-effect assertions in large news corpora for the DARPA Causal Exploration project.

Our system can analyze text to find indicators (such as demand for gas) and quantitative trends (rising) and use publicly available datasets to provide a confidence score for the validity of this text

SAGE: Synergistic Anticipation of Geopolitical Events

Crowdsourcing has shown itself to be an effective way of accurately predicting geopolitical outcomes. The primary goal of the SAGE Team—a performer in IARPA's Hybrid Forecasting Competition (HFC)—is to improve forecast accuracy by combining human forecasts with machine predictions. A key challenge in developing a hybrid system is trust in the models. SAGE found that human forecasters trust models to a limited degree compared to a control group that did not see the models. Forecasters anchored strongly on their personal beliefs and adjusted insufficiently toward the models. Side experiments revealed that human forecasters trust experts more than models and models more than their peers—even though individual experts are often inaccurate. Trust was driven primarily by the question format as compared with the topic (political vs. economic). Forecasters showed greater trust and accuracy when answering qualitative questions, and increased trust in the experts when answering quantitative questions. We found that we can easily manipulate trust by providing simple, clear explanations of the model's past performance. Such performance indicators promoted trust in the models as well as greater improvements in accuracy compared to the models. For the second forecasting season, the SAGE Team developed new interactive charts based on insights that people trust models more when they can alter them. The new forecasting season begins in April 2019.



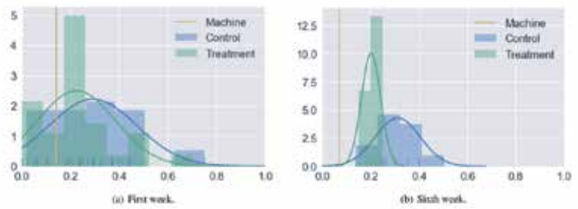


Fig 1 - The probability distribution projected on the correct option (hence, closer to 1 is more accurate) for the following question at two different time windows. This corresponds to the following question: "What will be the South Korean won to one U.S. dollar daily exchange rate on 29 June 2018?" Over time, the forecasters more closely resemble the machine.

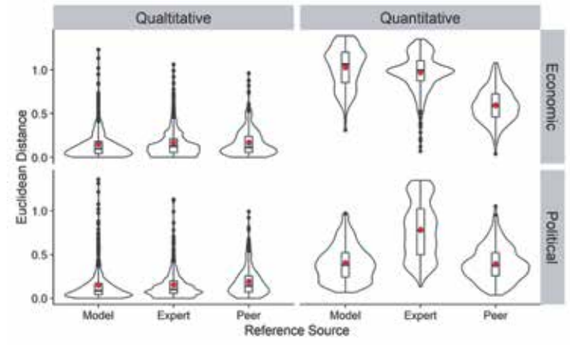


Fig 2 - Side experiment results. Trust (ordinal Euclidean distance) from the reference forecasts by question format and topic.

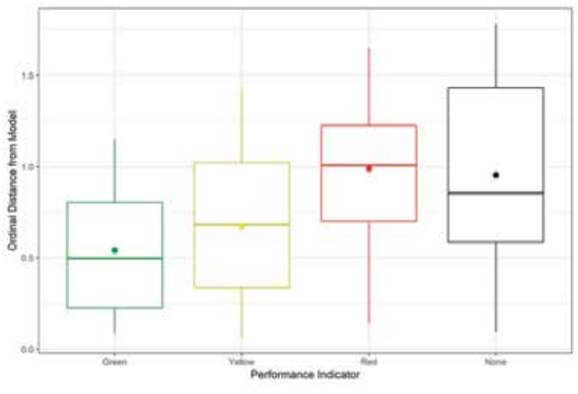


Fig 3 – Public platform experimental results. Trust (ordinal Euclidean distance) from the model prediction by performance indicator level. Black represents control—when no indicator was displayed.

Large-Scale Record Linkage

Record linkage is the problem of identifying records in separate datasets that refer to the same entity in the real world. This problem has been well studied in the database and semantic web communities; many papers have been published, and many systems have been developed to address it. In practice, the existing systems produce low-quality solutions for many datasets or fail to scale to datasets with millions of records. Our work on the Record Linkage Toolkit (RLTK) is developing a Python library of state-of-the-art algorithms using a simple programming interface to enable programmers to easily build high-quality, customized solutions for their record linkage problems. RLTK features a comprehensive library of similarity functions, provides scalable implementations of state-of-the-art blocking algorithms, provides convenient facilities for experimentation and evaluation, and interfaces easily with machine learning libraries such as scikit-learn. We have used RLTK to implement state-of-the-art solutions for multiple record linkage benchmarks, and to solve large-scale record linkage problems in several applications. Our current research focuses on methods to address problems when little or no ground truth datasets are available, a common issue in practical applications. RLTK software, documentation and tutorials are available at https://github.com/usc-isi-i2/rltk.

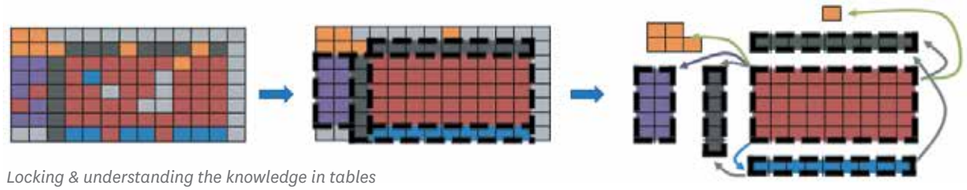
RLTK: Record Linkage ToolKit

license MIT | build passing | pypi package 2.0.0a14 | docs passing

RLTK: Addressing problems when little or no ground truth datasets are available

Unlocking and Understanding the Knowledge in Tables

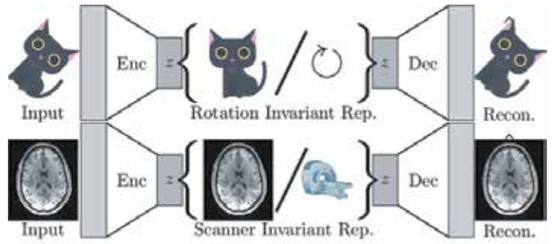
The web has millions of spreadsheets with a wealth of knowledge essential to decision-making and analysis. These spreadsheets present the knowledge in a wide variety of layouts and formats optimized for human consumption, making it difficult for machines to discover and use the knowledge to build models. In the Unlocking and Understanding the Knowledge in Tables project, we are building the tools to: 1) automatically extract the rich data found in tables; 2) make the process easily searchable and accessible to machines and users, ranging from non-experts to experienced analysts; and 3) provide human-comprehensible textual summaries and fact-checking with these datasets. Our approach to automatic extraction involves a pipeline that progresses from understanding the contents of individual cells, to blocks of related cells, to the layout of entire files, using techniques such as deep embeddings of table cells to probabilistic models of layouts. To help users find the right data, we've developed data augmentation techniques that extend the metadata found in datasets using semantic sources (WordNet, ConceptNet, ontologies) and statistical approaches (topic models, Word2Vec). Finally, we've created an extensive corpus of human summaries of quantitative data and are developing methods to summarize complex datasets in the same way a human would describe them. This project has contributed to several ISI efforts including DARPA's World Modelers, Data-driven Discovery of Models, and Causal Exploration projects.



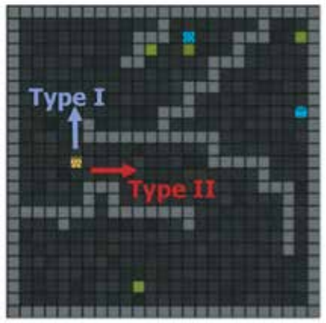
Locking & understanding the knowledge in tables

Fair and Invariant Representation Learning

The removal of unwanted information is a surprisingly common task. Filtering regressors for known covariates or signals, designing scale or translation invariant features, and providing fair predictions with respect to legally protected classes of individuals are all cases of information removal. Even if the unwanted variable is not present in the dataset, other correlated variables might still be exploited as proxies for that unwanted variable. In a sequence of recent and forthcoming work, ISI researchers have developed state-of-the-art methods for the removal of unwanted outside factors; these "invariant representations" are minimally informative of the target variable, while still providing maximal utility to another secondary task (e.g., prediction).



The top row shows that a neural network that detects objects like cats should give results that are invariant to transformations like rotations. In the bottom row we apply this idea to brain scanners so that a disease diagnosis can be made invariant to the type of scanner used.



Next-Generation Social Science

The goal of the DARPA Next-Generation Social Science program is to design new methods and platforms for social science experimentation at scale to enable robust and reproducible results about "what matters most" for emergent social phenomena. Our work focuses on unsupervised machine learning methods to automatically identify and explain factors affecting human decisions in the complex situations that arise within the online social science experiments developed in the program.

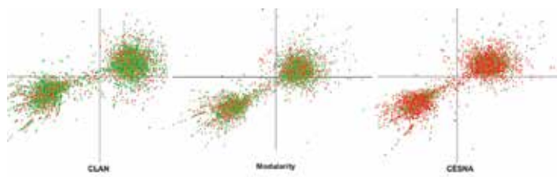
NGS2 explores human behavior through online multiplayer games like the one shown here. We use latent factor discovery to understand how different types of players will choose to act in different situations.

Understanding Group Biases (UGB)

Ethnographies are resources generated by social scientists to understand the beliefs, values, and traditions of a culture. The problem is that they are labor-intensive and difficult to build. The ultimate goal of Understanding Group Biases (UGB) is to do "big ethnography" using machine learning and NLP to automate the laborious task of anthropologists who spend years doing field research in order to study different cultures. UGB aims to detect cultural models using machine learning and natural language processing (NLP) tools applied to huge swaths of web data. With the growth of social media data, this task can be automated and expanded to cover all the data we have access to. Although a wide range of tasks can be covered in addressing the challenges of this project, we divided these tasks into two distinct and major categories:

1. Automatically identify different groups within a given corpus.
2. Given the groups from part 1, identify the beliefs, biases, and cultural norms in each of the groups.

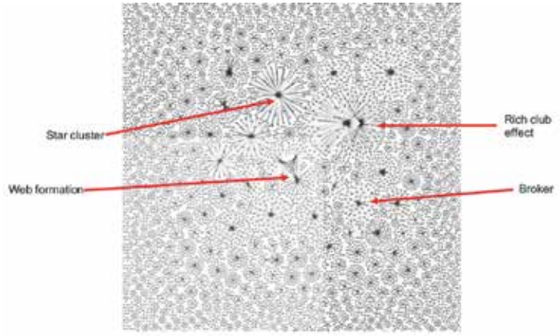
Work has been done to address both of the challenges mentioned above. For instance, in order to automatically identify groups, we introduced a new community detection method. Communities from lowly-connected attributed nodes (CLAN), not only achieved a higher performance accuracy in terms of F1 and Jaccard similarity scores, but also mitigated bias with inclusion of lowly connected nodes. The results of our proposed work are reported in the adjacent figure.



Comparison of our method, CLAN, to the state of the art in the Gamergate dataset: Green represents agreement with the ground truth labels.

Discovery and Dismantlement of Human Trafficking Networks

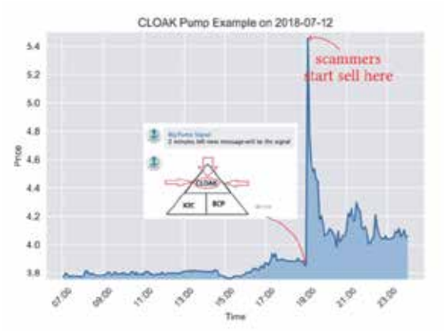
Recent investigative crawls in the online sex advertisement domain have revealed that, in the US alone, the number of (not necessarily unique) sex ads number in the tens, if not hundreds, of millions. An unknown, but non-trivial portion of these advertisements is targeted at human trafficking. Despite the seriousness of this issue, very little is known about the true extent, or underlying social organization of illicit sex activity. Studies at a national level have proven elusive because of data availability. In an interdisciplinary, multi-year collaboration with law enforcement and social scientists, ISI researchers are working to understand the key social forces and indicators that lead to the formation of intricate networks in the sex industry. Our hope is that the indicators and findings will be used to dismantle trafficking networks, and help bring about the rescue of victims and convictions of traffickers.



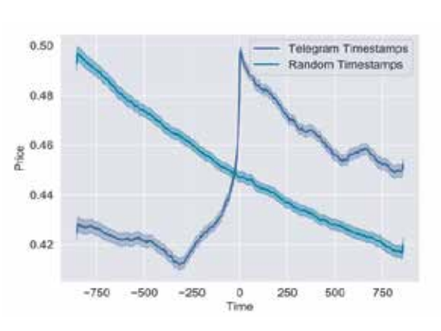
A real-world network of sex workers and their ties, constructed from tens of millions of public sex advertisements crawled by our collaborators over (the now defunct) backpage.com. The network shows phenomena worthy of both investigative and scientific interest, such as the brokerage and a 'rich club' effect.

Detection and Analysis of Cryptocurrency Manipulations on Social Media

Interest surrounding cryptocurrencies—digital or virtual currencies that are used as a medium for financial transactions—has grown tremendously in recent years. The anonymity surrounding these currencies makes investors particularly susceptible to fraud—such as "pump and dump" scams—where the goal is to artificially inflate the perceived worth of a currency, luring victims into investing before the fraudsters can sell their holdings. Because of the speed and relative anonymity offered by social platforms such as Twitter and Telegram, social media has become a preferred platform for scammers who wish to spread false hype about the cryptocurrency they are trying to pump. In this work we propose and evaluate a computational approach that can automatically identify pump and dump scams as they unfold by combining information across social media platforms. We also develop a multi-modal approach for predicting whether a particular pump attempt will succeed or not. Finally, we analyze the prevalence of bots in cryptocurrency-related tweets and observe a significant increase in bot activity during the pump attempts.



Pump announcement of a Telegram Channel "Big Pump Signal" overlaid on the market values of CLOAK. The announcement precedes the price swing.



Average normalized price of each coin centered around the pump timestamps and random timestamps. A pattern of spikes can be observed occurring within one hour of a pump message, followed by a general downward trend.

DIRECTOR **TERRY BENZEL**

With roots in early Internet Infrastructure research and development, ISI's Networking and Cybersecurity Division carries out a broad program of research spanning the areas of networking scientific investigations, Internet operations and governance, cybersecurity research topics, and a variety of infrastructure projects for experimentation and collaboration. These can be grouped into the following general areas:

- Network and security measurement, analysis and defenses
- Network infrastructure supporting science and operations
- Research, methods and infrastructure for cyber experimentation
- Social engineering attacks
- Binary program analysis, vulnerability discovery and reverse engineering
- Modeling human behavior for cybersecurity and social simulation
- Theory and practice of distributed computing

The prevalence and interdependence of cybersecurity, networking and social systems informs the research and development agenda of ISI's Networking and Cybersecurity Division. Our research and development activities are aimed at understanding the underlying Internet, the theory and practice of distributed computing, approaches to analyzing vulnerabilities and scientific approaches to modeling, experimenting and evaluating critical infrastructure systems. ISI's Networking and Cybersecurity Division carries out a broad program of research spanning these areas and applies them to pressing problems. These can be grouped into the following general areas:

NETWORK AND SECURITY MEASUREMENT, ANALYSIS AND DEFENSES

We research methods to observe and collect network and network security data and behaviors. These methods are used to develop novel networking capabilities and network defenses.

NETWORK INFRASTRUCTURE SUPPORTING SCIENCE AND OPERATIONS

Network infrastructure that fosters network and cybersecurity enabled collaborations, driving discovery in science for research, education communities, and Internet users domestically and internationally.

RESEARCH, METHODS AND INFRASTRUCTURE FOR CYBER EXPERIMENTATION

Valid scientific experiments are required to accurately evaluate and assess network systems. Conducting these experiments necessitates modeling multiple, complex network, environmental, traffic, and behavioral effects and systems. Our work creates models, experimentation frameworks, tools and approaches to enhance the science of cyber experimentation and make the experiments reusable, repeatable and robust.

SOCIAL ENGINEERING ATTACKS

Social engineering attacks such as phishing and impersonating are on the rise because often an organization's weakest link in security is the human in the loop. By leveraging the metadata from communication channels, and using techniques to redirect attackers, we can produce new methods for detection and fingerprinting campaigns across multiple attempts.

BINARY PROGRAM ANALYSIS, VULNERABILITY DISCOVERY AND REVERSE ENGINEERING

Binary program analysis is the process of analyzing software programs in their binary form, also called "executable." Our work focuses on reverse engineering to search for vulnerabilities in software that is released without source-code, and to assess the security of software products.

MODELING HUMAN BEHAVIOR FOR CYBERSECURITY AND SOCIAL SIMULATION

Human behavior is a key determining factor in assessing the effectiveness of an organization's cyber defenses, including its policies. Our current research aims to observe and model important aspects of human behavior in order to predict likely responses to security posture and the evolution of information in online social networks.

THEORY AND PRACTICE OF DISTRIBUTED COMPUTING

Understanding the foundations of distributed computing is important for the design of efficient computational techniques across all scientific fields. As a consequence of failures and the asynchrony pervasive in distributed systems, many problems that are trivial to solve sequentially are impossible or infeasible to solve in a distributed fashion, thus presenting us with problems of deep intellectual yet practical interest.

NETWORK AND SECURITY MEASUREMENT, ANALYSIS AND DEFENSES

The ANT lab has been *developing new methods* to *generate* and *share* network data with researchers for more than a decade. We draw on data from **network traffic**—anonymized packet headers, controlled testbed experiments, traffic flow data, and curated data such as distributed denial-of-service (DDoS) attacks; **Internet scanning**—censuses of all IPv4, data about IPv6, network topology, Internet outages; and **application level data**—anonymized Domain-Name-System queries, anycast mapping.

New Data Collection: Near-Real Time Network Outages

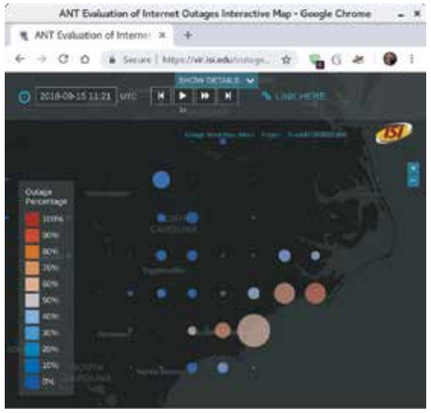
New in 2018 was our development of **near-real time detection of network outages**. Since November 2013, we have been observing Internet outages with Trinocular. Trinocular observes the Internet from six locations around the world, and when combined, provides a picture of global IPv4 reliability.

We have added streaming data processing to Trinocular, providing reports of outages within about one hour of their onset. (In our prior work, results were computed with large, batched computations every few months.)

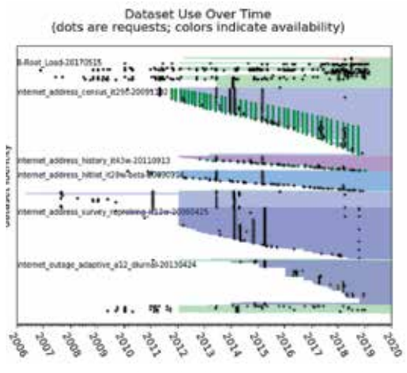
We visualize these results in a website at https://ant.isi.edu/outage/world/. Outages are useful to understanding how the U.S. infrastructure reacts to natural disasters such as hurricanes.

This figure shows Internet outages resulting from Hurricane Florence's flooding and high winds in September, 2018.

We describe potential uses of outage detection to understand government policies in our paper "The Policy Potential of Measuring Internet Outages" in the *Proceedings of the TPRC*, September 2018. Our outage website was initially developed with support from an ISI Michael Keston Endowment and DHS. It is being extended by the NSF and DHS.



Internet outages in the Carolinas, 24 hours after Hurricane Florence made landfall



Dataset generation, distribution, and popularity

Data to Researchers

The ANT lab has been developing new measurement methods for more than a decade, with support from from the DNS and NSF. Between 2006 and December 2018, we provided 1798 datasets (733 TB of data before compression).

This graph shows dataset generation (colored regions) and distributions (dots), and it shows the popularity of our datasets. Each dot is a dataset being sent to a researcher. The data shows several interesting things: the bands at the top show very popular datasets over the entire period, like our curated distributed-denial-of-service (DDoS) attacks. The middle blue areas show our regularly generated Internet scans and our recent Internet outage data. Some of these are subscribed to by external researchers, while other times (the vertical towers of dots) show a research group requesting our entire "back catalog" of multiple years of data to carry out longitudinal analysis.

Understanding Privacy of Network Data

Sharing data requires careful attention to the privacy of network users. ISI has been studying these issues for some time, and has been a leader in understanding and improving privacy of shared DNS data.

In 2018 we distributed new libraries for CryptoPAN, an algorithm initially developed at Georgia Tech. We pioneered its use for DNS, and our library is in use with open-source software from us and from DNS-OARC, an industry research group.

In addition, we have advanced the discussion around DNS privacy through two papers at the NDSS Workshop on DNS Privacy in February 2018: "Analyzing and Mitigating Privacy with the DNS Root Service" and "Enumerating Privacy Leaks in DNS Data Collected Above the Recursive."

Network Traffic Identification for Cybersecurity

Identifying an application from its traffic has been a topic of interest for decades, and many techniques have been proposed. Unfortunately, many current techniques fail on encrypted traffic as the features they rely upon become obfuscated. Modern methods to detect applications within encrypted flows generally utilize machine learning approaches, including k-means clustering, k-nearest neighbors and hidden Markov models. The principal challenge to identifying encrypted applications is defeating the obfuscation caused by encapsulation and potential multiplexing of multiple application flows. Additionally, performing traffic analysis and classification at line-rates is challenging. Our research looks at transformation of flows into waveforms, enabling signal and image processing techniques for rapid application identification without isolating individual flows. In pre-transformation, one is looking at various methods of featurization, including extracting blocks of common application patterns, such as "bulk transfer" and "parallel bulk transfer," and higher-level patterns such as "TLS renegotiation" or "stream buffering." Preliminary results show that these approaches can rapidly identify known applications and have the potential to detect instances of new, distinct applications. These techniques can also be used to identify and isolate anomalous behavior within a flow or across multiple flows. This work has been applied to the GAWSEED project (DARPA's CHASE program) and will be used in the recently awarded APROPOS project (DARPA's Searchlight program).

NETWORK INFRASTRUCTURE SUPPORTING SCIENCE AND OPERATIONS

B-Root DNS Infrastructure

ISI's B-Root DNS critical infrastructure serves as the foundation in a number of efforts to advance the state of the art for the Domain Name System (DNS) in multiple directions. B-Root is one of the 13 different root servers that are at the top of the DNS system ("above" .com and .edu). The Internet's DNS was created at ISI in the late 1980s, and we have managed and maintained the B-Root DNS Root server ever since–with a focus on both a mission of research and as a service to the Internet at large.

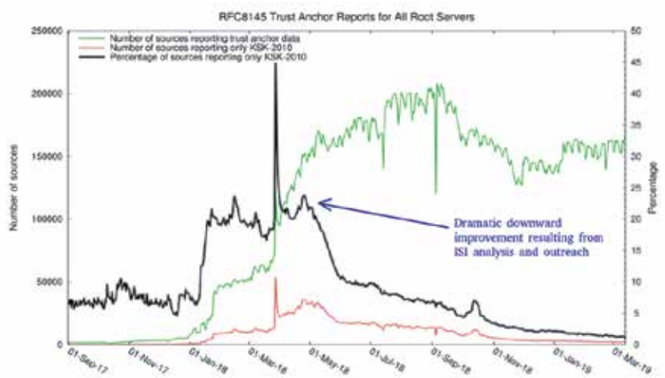
New Service Locations: B-Root currently operates from two locations (Los Angeles and Miami) to provide disaster recovery, capacity, and to reduce latency. In 2018 B-Root installed its first international Anycast instance in Chile, in cooperation with PitChile. We expect this site to go into production service in early 2019.

In the Internet Community: ISI has also played a critical role in establishing an Internet Governance model for the DNS Root Server system. This multi-year cooperative effort is working to put in place oversight, increasing transparency and accountability, in cooperation with the International Corporation for Assigned Names and Numbers (ICANN).

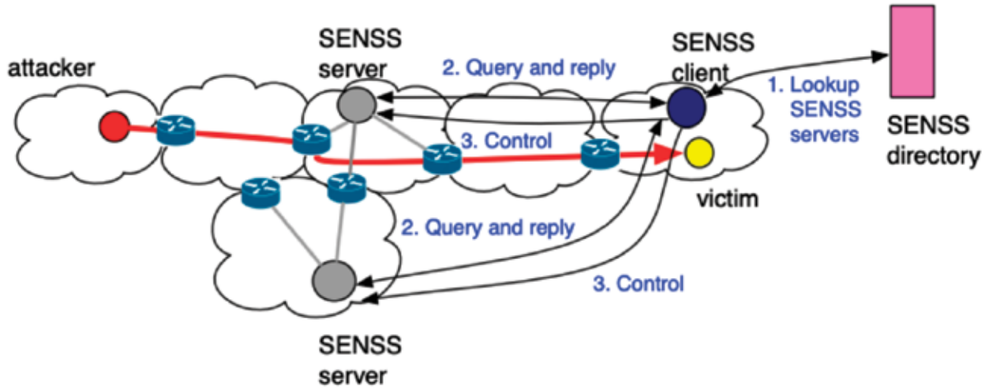
Supporting Research: B-Root actively supports USC and ISI's mission as a research university. We have worked with researchers in the Global Analysis of Weak Signals for Enterprise Event Detection (GAWSEED) project, part of DARPA's Cyber Hunting at Scale (CHASE) program, the DDoS Defense In Depth for DNS (DDIDD) project under the NSF's CICI (Cybersecurity Innovation for Cyberinfrastructue) program, and shared data through the DHS IMPACT program. Insights from B-Root have informed academic work such as "When the Dike Breaks: Dissecting DNS Defenses During DDoS" by researchers from USC/ISI with SIDN Labs, University of Twente (Netherlands), University of Passo Fundo (Brazil), and published in the ACM Internet Measurements conference 2018.

Intersecting Community and Research: B-Root played a critical role in the intersection of the Internet governance and research. DNS Security (DNSSEC) depends on a central cryptographic key managed by the International Corporation for Assigned Names and Numbers (ICANN). ICANN had been planning this change for several years, yet was uncertain about risks surrounding the change and postponed it for a year out of concern.

ISI identified a major source of incorrect use of old cryptographic keys, and addressed the issue through community outreach. The resulting conclusion built confidence and allowed ICANN to successfully carry out the key change in October 2018. A key insight was ISI research identifying one software product that was the source of a majority of errors. The following graph shows errors in the black line. After we contacted the vendor, they rolled out updates and the error reports dropped in half (June and July, 2018), even as reporting increased (green line).



The rise and fall of trust in cryptographic trust anchors



SENSS: A collaborative approach to fight DDoS attack

Defeating Distributed Attacks Through ISP Collaboration

Volumetric distributed denial-of-service (DDoS) attacks can bring any network to a halt. Because of their distributed nature and high volume, the victim often cannot handle these attacks alone and needs help from upstream ISPs. Today's Internet has no automated mechanism for victims to petition ISPs for help in handling attacks, and ISPs themselves do not offer such services. Instead, traffic is usually redirected to cloud-based scrubbing centers, and then routed back to the target of the attack. This approach is costly; it introduces delays, jeopardizes user privacy and does not allow the victim to control the mitigation process.

ISI's STEEL lab has developed a collaborative approach to fight DDoS attacks. The SENSS project—funded by the Department of Homeland Security—enables the victim of an attack to request attack monitoring and filtering on demand, and to pay only for the services rendered. Requests can be sent both to the immediate and remote ISPs in an automated and secure manner, and can be authenticated by these ISPs without having prior trust with the victim. The figure below shows the SENSS architecture and operation.

Simple and generic SENSS APIs enable victims to build custom detection and mitigation approaches against a variety of DDoS attacks. SENSS is deployable with today's infrastructure, and it has strong economic incentives—both for ISPs and for the attack victims.

SENSS is very effective in sparse deployment, offering full protection to direct customers of early adopters, and considerable protection to remote victims when deployed strategically. For example, in 2016 the large DNS provider "Dyn" was hit by a 600 Gbps attack; this disrupted services to more than 1,200 domains. If SENSS had been strategically deployed on only four ISPs close to Dyn, it would have filtered 100% of the attack within seconds.

The SENSS project will be piloted in three academic ISPs across the U.S. in 2019.

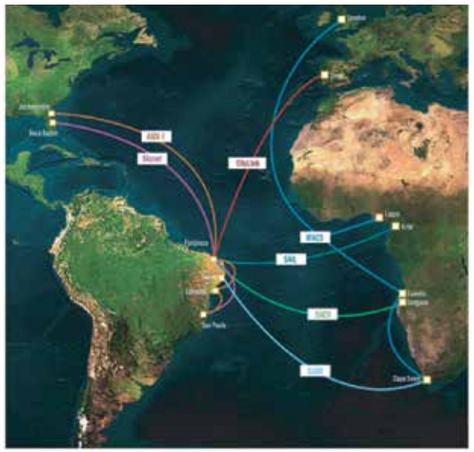
NETWORK INFRASTRUCTURE SUPPORTING SCIENCE AND OPERATIONS

AARCLight—A Series of Undersea Fiber Optic Cable Networks in the South Atlantic

A USC/ISI researcher from the Networking and Cybersecurity Division facilitated a US-Africa side meeting for AARCLight at the 2018 Internet2 Global Summit in San Diego, California. This meeting was aimed at defining consensus on the use of a South Atlantic research and education (R&E) network link to facilitate collaborations between Africa and North and South America. Later that year, ISI represented the project at the Ubuntunet Connect conference in Zanzibar, Tanzania, where we led a quantitative and qualitative survey of network operators in Africa and served as an invited speaker for the NSF-funded AARCLight planning project.

A series of new and existing undersea fiber optic telecommunication networks are connecting continents in the South Atlantic. With the exception of the *SABR*, Fortaleza (Brazil) is the landing point for all the cables listed here.

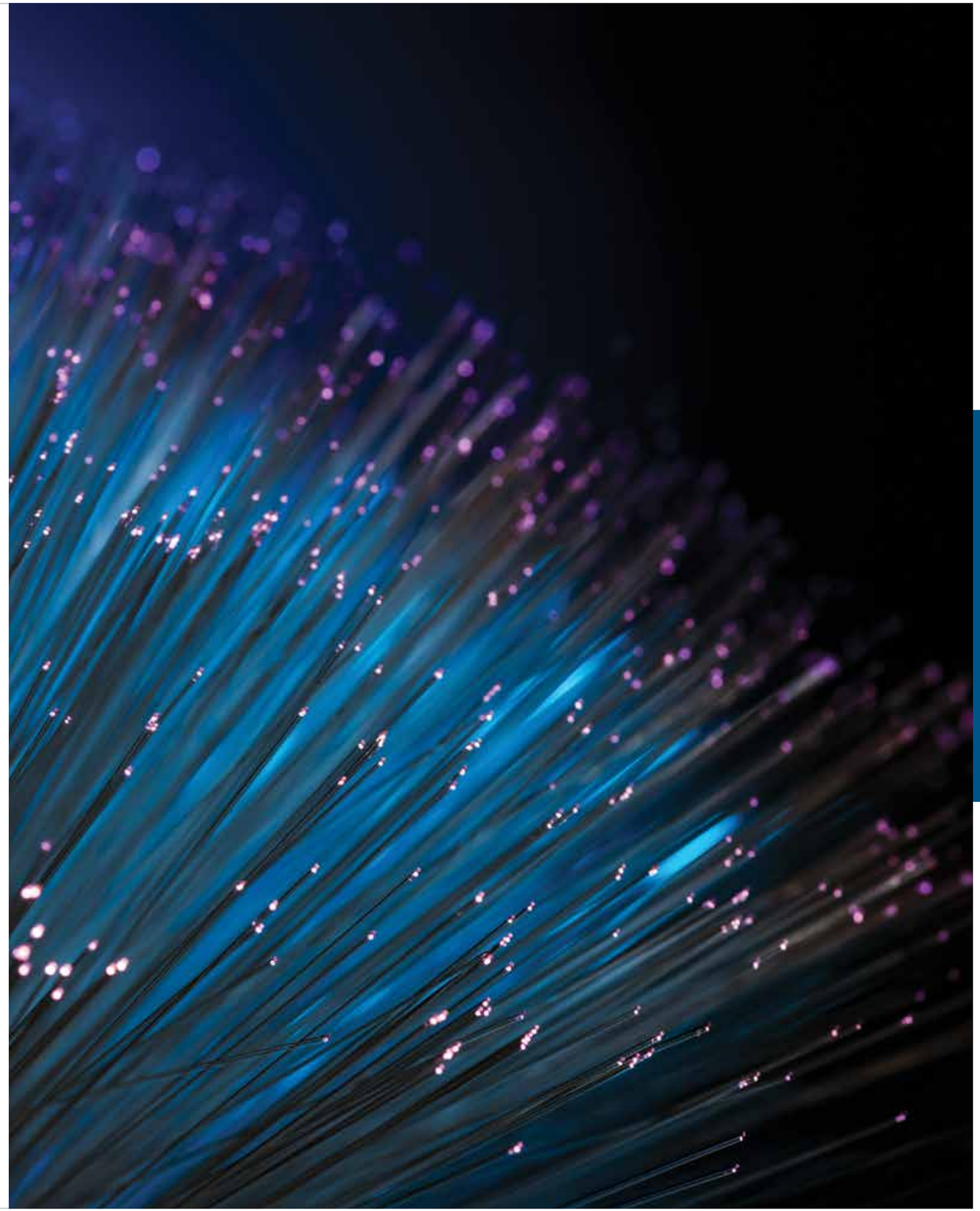
- The *MONET* undersea fiber optic cable system, which connects to Boca Raton and further connects to Miami, Florida (USA) and to Fortaleza and Santos (Brazil), has been operational since May 2018. It spans 10,556 km and has an initial design capacity of more than 64Tbps.
- The 6,165 km undersea fiber optic cable system *South Atlantic Cable System* (*SACS*) between Fortaleza (Brazil) and Sangano (Angola) was completed in the third quarter of 2018 and is operational. *SACS* is owned and managed by Angola Cables who has entered into an agreement with USC/ISI and Florida International University (FIU) for effectively 100 Gbps provisioned over the spectrum using coherent WDM technology. *SACS* offers a total design capacity of 40Tbit/s between Fortaleza (Brazil) and Luanda (Angola).



AARCLight: Undersea fiber-optic cables in the South Atlantic

- Almost at the same time, in September 2018, the completion of the close-to-6000-km undersea fiber optic cable installation, *South Atlantic Inter Link* (*SAIL*), from Fortaleza (Brazil) to Kribi (Cameroon), was announced ready for service.
- The *EllaLink*, an undersea 9300 km fiber optic cable system, from Fortaleza (Brazil) to Sines (Portugal), is currently planned and will be ready for service in 2020.
- The 17,500 km *America Movil* (*AMX-1*) undersea fiber optic cable system, from Fortaleza (Brazil) to Jacksonville and Hollywood (USA), has been operational since 2015. *AMX-1* has multiple landing points in Colombia, Brazil, the Dominican Republic, Puerto Rico, Guatemala, Mexico, and the United States.
- *Seaborn Networks' SABR* undersea fiber optic cable network is currently being developed to connect Cape Town (South Africa) to Recife (Brazil), and eventually the USA. It will be ready for service in 2029.

The AARCLight project has created several components critical to revolutionizing U.S. collaboration with the African continent. It has developed a critical mass of connectivity by signing agreements with funding for spectrum between the U.S. and Luanda, Angola via Fortaleza, Brazil, that could last 35 years. AARCLight has fostered critical research collaborations with the nascent and established African regional network consortiums—the UbuntuNet Alliance and WACREN. It has developed a critical partnership with SA NREN and TENET in South Africa to extend 100GB connectivity from Angola to Cape Town and on to East Africa. AARCLight has further strengthened the partnership with RNP and ANSP, the national R&E networks for Brazil and the State of Sao Paulo where Brazilian collaboration in the operation and the development of U.S.-Brazil-Africa network connectivity will play a critical role.



RESEARCH, METHODS AND INFRASTRUCTURE FOR CYBER EXPERIMENTATION

Scientific Experiment Modelling

Valid scientific experiments are required to accurately evaluate and assess network systems. Conducting these experiments necessitates modeling multiple complex network, environmental, traffic, and behavioral effects and systems. Unfortunately, the majority of experimenters are not experts in all the domains required to conduct these types of experiments. For example, an expert in loss-resilient transport protocols may not be an expert in generating accurate loss and jitter models for evaluation. Technologies must be made available that can help capture the knowledge of one domain expert and enable another to utilize it without becoming an expert themselves.

The Networking and Cybersecurity Division has undertaken a significant research effort to address this issue. Part of this effort is the design of a system which enables modelling experts to contribute tools that are easily usable by other experimenters. Experimenters search through the repository to find modelling capabilities for their specific needs. Additionally, other experts can use these tools to generate models of their specific corner of a field and contribute those models back to the ecosystem. For example, an engineer at CenturyLink could generate a model of their network using the available modeling tools. This enables an experimenter to evaluate their system on an accurate model of the CenturyLink network without requiring detailed knowledge of that network or the skills required to model it.

Capturing and disseminating this knowledge is only part of the overall problem to be addressed. Experimenters must also be educated on what models are actually useful for their experiments. Experimenters typically use overly simplified models or desire perfect models—the former leading to invalid results and the latter being superfluous to the goals of the experiment. To this end, we are working with experimenters to help understand their experimental goals and constraints, and select appropriate models based on them. Our ultimate goal is to encode this information directly within the experiment description and have the experimental infrastructure engine suggest potential models that may satisfy the experimenters goals and constraints. Multiple research efforts within the Networking and Cybersecurity Division are exploring solutions to this problem in parallel.

SOCIAL ENGINEERING ATTACKS

Detecting and Responding to Phishing

Social engineering attacks such as phishing and impersonating are on the rise, as often an organization's weakest link in security is the human in the loop. According to the 2017 Verizon Data Breach Report, nearly half of all documented breaches involved social engineering attacks. Social engineering attacks often have a digital component, such as use of SMS, email, or social media accounts, or use a combination of channels. As such, within the metadata of each channel are clues that could lead to attack detection and in some case, clues as to who the attacker is. By applying NLU and NLP to the language an attacker uses and combining extracted features from the language with metadata from the channel, we can produce new methods for detection and fingerprint campaigns across multiple attempts. This work is currently being developed under the PIRANHA project (DARPA's ASED program), where, in addition to detection, we are developing automatic methods to respond to phishing attacks and trick attackers into giving up details about themselves.

BINARY PROGRAM ANALYSIS, VULNERABILITY DISCOVERY AND REVERSE ENGINEERING

Is Your Software Secure?

Binary program analysis is the process of analyzing software programs in their binary form—also called "executable" (such as .exe files on the Microsoft Windows platform). This process is useful in the context of reverse engineering to search for vulnerabilities in software that is released without source-code, and to assess the security of software products.

The Networking and Cybersecurity Division has been working on developing new approaches in this domain through several projects in collaboration with U.S. and international academic partners.

In a project entitled "A binary analysis approach to retrofit security input parsing routines," we have been working on a new approach for automatically detecting software vulnerabilities at a large-scale, and to directly patch those within the binary (executable) program. During this work, we have discovered multiple zero-day vulnerabilities in real-world software used in production environments, which has, so far, led to a new report in the Common Vulnerability and Exposures (CVE) database (CVE-2018-18311). We are still working on reporting additional vulnerabilities. The first phase of this project also led to a publication in the LangSec workshop, part of the 2018 IEEE Symposium on Security and Privacy. This project is a collaboration with Arizona State University.

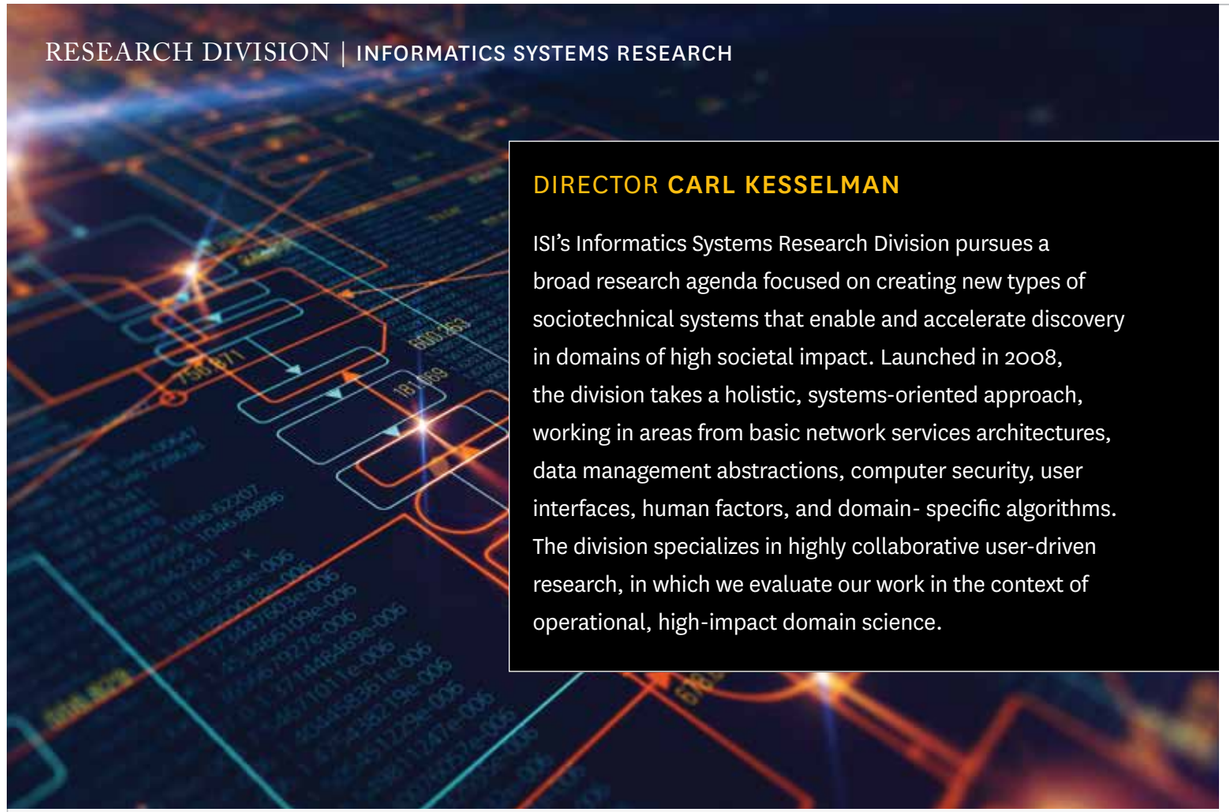
As part of an international collaboration with researchers from the University of California Santa Barbara, Arizona State University, the University of Milan (Italy), CentraleSupélec and INRIA (France), we are focusing on an automated approach to assess the security of boot firmware images, such as those present in most modern Intel platforms through the Unified Extended Firmware Interface (UEFI). Boot firmware has been the target of numerous attacks, giving the attacker control over the entire system while being undetected. Ensuring that boot firmware is tamper-free is therefore a critical step when deploying updates, or before integrating third-party components as part of the development stage. Our approach, entitled "Boot Keeper," leverages state-of-the-art binary program analysis techniques in order to verify a set of key security properties on such firmware images.

MODELING HUMAN BEHAVIOR FOR CYBERSECURITY AND SOCIAL SIMULATION

Human behavior is a key determining factor in assessing the effectiveness of an organization's cyber defenses, including its policies. Our current research aims to observe and model important aspects of human behavior in order to predict likely responses to security posture and the evolution of information in online social networks. Our cognitive agent architecture, DASH, captures important aspects of human online decision-making, including dual-process reasoning and mental models.

Recently, in collaboration with the ISI's AI division, we have scaled DASH to handle distributed simulations with tens of millions of agents. This has been used to predict the behavior of individuals and groups in online settings such as Github, Twitter and Reddit. We also began an integration of our agent simulation with computational game theory in order to increase the range of problems to which it can be applied and the scale of games that can be solved.

Additionally, we began an integration of our agent simulation with computational game theory in order to increase the range of problems to which it can be applied and the scale of games that can be solved.



DIRECTOR **CARL KESSELMAN**

ISI's Informatics Systems Research Division pursues a broad research agenda focused on creating new types of sociotechnical systems that enable and accelerate discovery in domains of high societal impact. Launched in 2008, the division takes a holistic, systems-oriented approach, working in areas from basic network services architectures, data management abstractions, computer security, user interfaces, human factors, and domain- specific algorithms. The division specializes in highly collaborative user-driven research, in which we evaluate our work in the context of operational, high-impact domain science.

In earlier work, the Informatics Systems Research Division developed grid computing infrastructures to support the creation and operation of "virtual organizations" as a foundation for collaboration and discovery. This work focused on understanding methods for sharing computing and storage infrastructure across distributed resource providers and collaborators. The resulting methods played a role in two Nobel prizes—e.g., all the data analysis for discovering the Higgs boson was performed on a global grid infrastructure, and the recent discovery of gravity waves took place on a data grid.

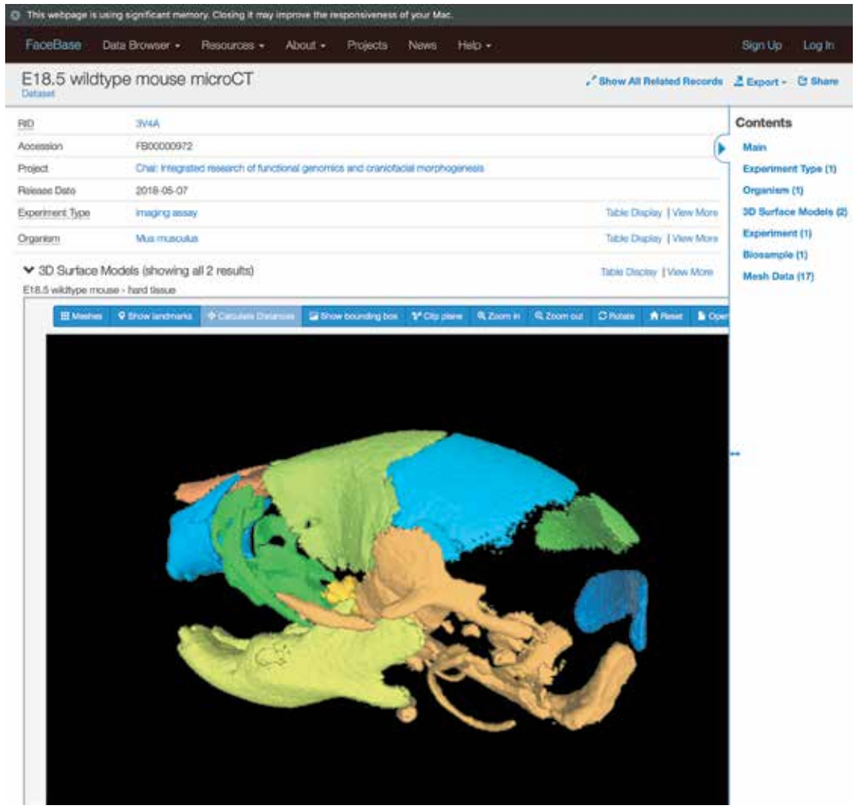
More recently, the division has focused on biomedical applications. Our current collaborations cover a broad range of applications—from basic science to clinical use cases spanning molecular biology, basic neuroscience, neuroimaging, stem cell research, and cranio-facial dysmorphia.

Our researchers work closely with ISI's highly regarded artificial intelligence, networking, and distributed systems experts, as well as with two of USC's nationally ranked Viterbi School of Engineering units: the Daniel J. Epstein Department of Industrial and Systems Engineering, and the Department of Computer Science. The division participates in many collaborative projects with the faculty of Dornsife College, Keck School of Medicine, and Osterow School of Dentistry. Members of the division play a leadership role in the Michelson Center for Convergent Biosciences, including establishing the new Center for Discovery Informatics as part of the university's convergence biosciences initiative. The division also plays a central role in four international consortiums. Most recently, the Informatics Systems Research Division has been a central participant in work by the National Institutes of Health in defining a shared data infrastructure for biomedical research.

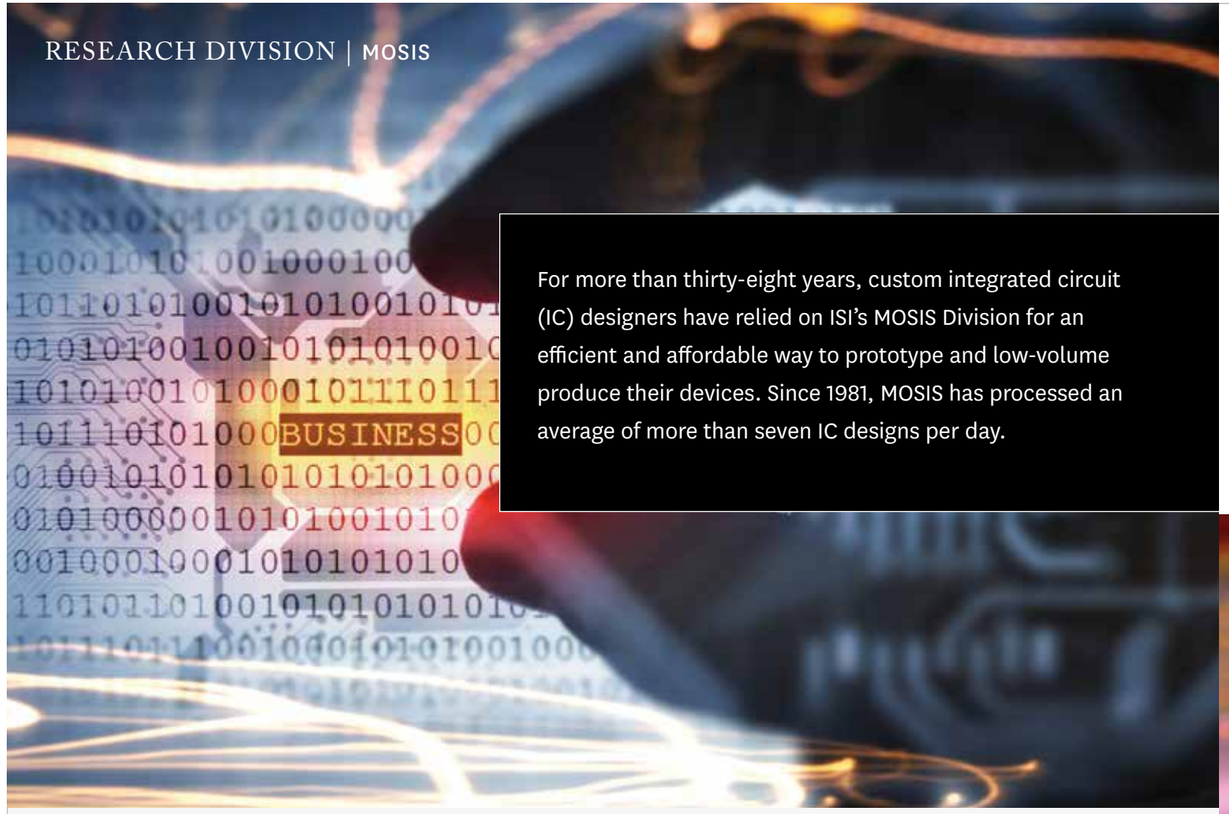
RESEARCH HIGHLIGHTS | INFORMATICS SYSTEMS RESEARCH

ISI's Informatics Systems Research Division develops innovative new approaches to accelerate scientific discovery by focusing on eliminating the complexities associated with assembling, organizing and manipulating complex, large-scale data collections. The goal of the Informatics Division is to understand how to architect, assemble and operate complex social-technical systems that will eliminate the technological barriers of the past years and result in a radically faster time to new discoveries. To better understand the impact of technology on daily work, a core element of the division is the creation of operational scientific infrastructure which is used on a daily basis by diverse scientists and creates a "living laboratory" for understanding the interactions between science as practiced and information technology. The division contributes to many diverse scientific enterprises including developmental biology, protein structures, whole-cell modeling, operations research, organ reconstruction, and advanced genetics.

DERIVA: The Deriva platform developed by ISI's Informatics Systems Research Division is playing a critical role in advancing scientific investigation in many diverse areas. Deriva is being used to empower scientists to create data that is FAIR, i.e., **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable. Deriva is providing the foundation for many scientific domains, some examples of which are shown below.



The Deriva platform is dramatically simplifying the discovery and access to complex scientific data. This figure shows that Deriva automatically organizes and presents data that captures how a specific gene (Pcnt) impacts the development of the genitourinary tract, as indicated by the colors in the microscope images. Without Deriva, scientists would have to organize, by hand, all of this data, and it would be very difficult to assemble and search.



For more than thirty-eight years, custom integrated circuit (IC) designers have relied on ISI's MOSIS Division for an efficient and affordable way to prototype and low-volume produce their devices. Since 1981, MOSIS has processed an average of more than seven IC designs per day.

Many turn to MOSIS for our special expertise in providing multi-project wafers (MPWs) and related services that drive IC innovation. This "shared mask" model combines designs from multiple customers or diverse designs from a single company on one mask set. This is a practical prototyping channel that allows designers to debug and perform essential design adjustments before making a substantial strategic investment. Today, with IC mask and fabrication costs soaring, more designers than ever are using MPWs to manufacture proven devices and prototype new designs on a single wafer.

Beyond MPWs, customers are choosing the services of MOSIS as their resource partner for low-volume production. From design specification interpretation through mask generation, device fabrication and onto assembly, MOSIS is their trusted expert interface to the semiconductor ecosystem. In addition to our commercial service, MOSIS is part of the following active research programs:

DARPA CRAFT (Circuit Realization At Faster Timescales)

Custom integrated circuit design flow and methodology that will accomplish the following:

- Sharply reduce the amount of effort required to design high-performance custom integrated circuits,
- Greatly facilitate porting of integrated circuit designs to secondary foundries and/or more advanced technology nodes.
- Strongly increase reuse of integrated circuit elements.

In support of CRAFT, MOSIS is organizing 16nm CMOS private MPW runs at TSMC.

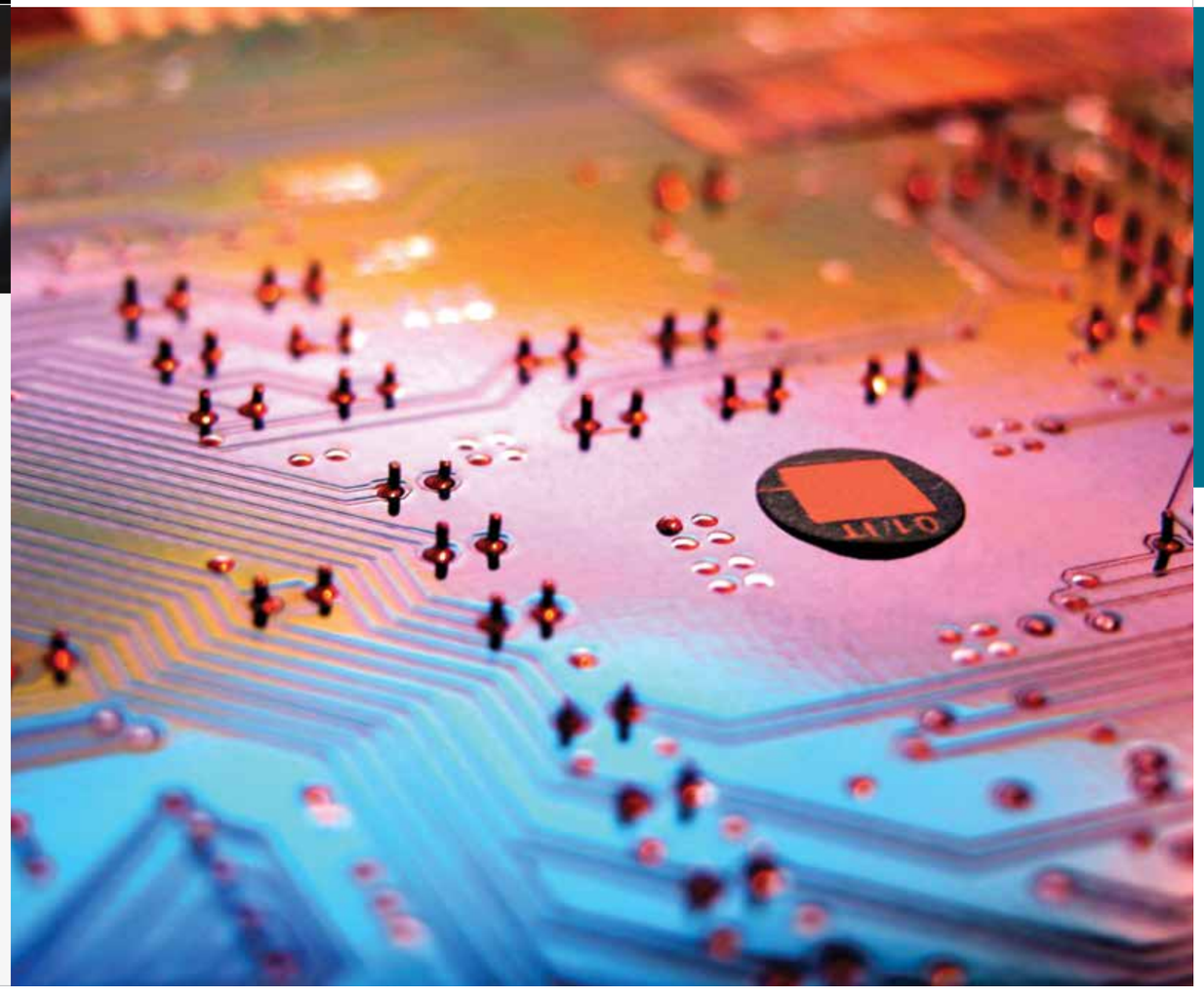
IARPA TIC (Trusted Integrated Chips)

Split-manufacturing, with a new approach to chip fabrication, where security and intellectual property protection are assured. MOSIS is on the government team, and is organizing private MPW runs (130nm, 65nm, and 28nm CMOS).

MINSEC (Microelectronics Needs and Innovation for National Security and Economic Competitiveness)

MOSIS is an integral part of ISI's team to establish a plan and direction for the U.S. Government concerning the microelectronic innovation centers in the U.S. This effort aims to ensure that the U.S. defense and commercial industry do not fall behind the microelectronic IC design and fabrication innovations from other parts of the world—especially the Asian-Pacific region.

Since its inception, MOSIS has been led by only two directors. Wesley Hansford, after 34 years with the MOSIS Division and the last ten years as the division's director, announced his retirement. A search for a new director, one who can ensure that MOSIS continues to offer a value-added role to the microelectronics community, has been initiated. USC's Information Sciences Institute is committed to ensuring that MOSIS adapts to the design environment, provides foundry access, optimizes the design submission to fabrication cycle time, and provides outstanding customer service for the next thirty-plus years.



COSINE

The DARPA SocialSim program that began in 2017 postulates that most of the world population is connected to the global information environment. Therefore, it is of paramount importance to understand how information spreads and be able to capture, at scale and in real-time, the mechanisms that govern the diffusion of online information in an increasingly interconnected world.

Our effort is spearheaded by USC's Information Sciences Institute and includes Indiana University and the University of Notre Dame. This team was selected to tackle the challenges posed by the DARPA SocialSim program with a project entitled *COSINE: Cognitive Online Simulation of Information Network Environments*, led by ISI researchers.

Brainstorming sessions at the DARPA SocialSim Challenge with teammates from ISI, Indiana University, and University of Notre Dame, December 2018.

The goal of COSINE is to create the first-of-its-kind cognitive agent simulation framework for studying multi-scale dynamics of social phenomena in online information environments. Individual agent behaviors within COSINE will be based on the first principles of human behavior, validated through laboratory experiments and empirical analysis. In addition, COSINE's multi-resolution, scalable framework will enable time-resolved massive simulations of dynamic, networked information environments. Online information diffusion is modeled using top-down statistical-physics models, mesoscopic-level compartmental and network-based models, and bottom-up agent-based dynamics. Agent models are based on neurocognitively grounded principles of human behavior that incorporate bounded rationality and cognitive biases within models of attention. The system is calibrated on real-word data collected from a plethora of online platforms.

COSINE is a rich virtual laboratory for studying dynamics of online social phenomena at different temporal resolutions and at multiple scales—from individual to community to global collective behavior. COSINE directly incorporates multiplex networks into interactions between agents, thereby enabling the study of the impact of network structure and multiple communication modalities on emerging social phenomena, as well as how the position of individuals within the network affects their behavior.

In addition, COSINE elucidates how the system responds to endogenous (shifts of attention) and exogenous shocks (e.g., crises, emergencies), and provides for a greater understanding of how the structure of social networks evolves in response to internal and external shocks.

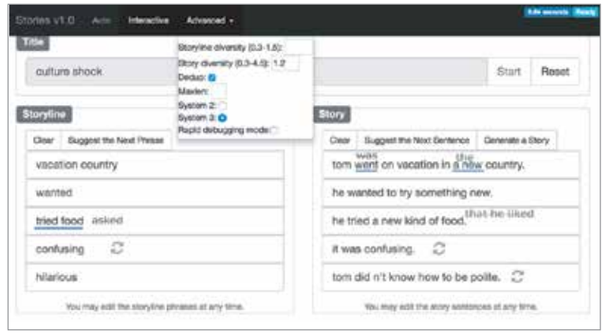
COSINE is bringing a wealth of innovations to agent-based simulations and to the computational social sciences at large. Qualitatively, COSINE lays out first principles to characterize simulation-driven social sciences, provides unprecedented descriptive richness in the modeling of social dynamics, and fosters the development of causal inference methods. Quantitatively, it allows for planetary-scale simulations of online information environments.

The USC Information Sciences Institute was awarded approximately $5 million and has topped the first two DARPA SocialSim Challenges that were aimed at simulating the spread of information on GitHub, Twitter, and Reddit. Furthermore, the output of this project has already appeared in journals such as P*roceedings of the National Academy of Sciences*, *Communications of the ACM*, and many other top conferences in ACM, IEEE, and AAAI.

Communicating with Computer (CwC)

The goal of the Communicating with Computer (CwC) project is to advance the abilities of the human and the computer to collaboratively achieve certain goals, which cannot be achieved with high quality by one party only.

We developed a human-computer collaborative storytelling system, where humans will be teamed up with computers to write short five-line commonsense stories. There has been some prior work on human-computer collaborative writing, but most of these efforts only allow simple interactions, such as each party taking turns writing a sentence or computers making suggestions and humans choosing to accept them or rewrite.



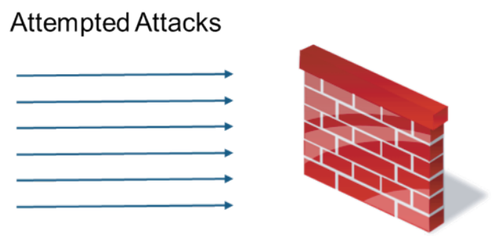
An illustration of our collaborative story composition interface, showing advanced options and annotated real user interactions from an example study

We adopt a plan-and-write strategy, where the two parties will first plan out a storyline (one phrase for each sentence) and then write a story together. The novel planning step helps the parties get on the same page regarding the story; this helps to generate coherent stories. We also maximize the means of collaborations: human can change anything in the storyline or story at any time during the collaboration; they can also change some system configurations to encourage more creative or less creative computer generations.

This screenshot shows the variety of interactions a user can take in the collaboration, and is annotated with an example-in-action. User inserted text is underlined in blue, generated text that has been removed by the user is in a grey strike-through. The refresh symbol marks areas that the user regenerated to obtain a different sentence (presumably after being unhappy with the first result). As can be seen in this example, minor user involvement can result in a significantly better story.

Prediction is Not That Easy

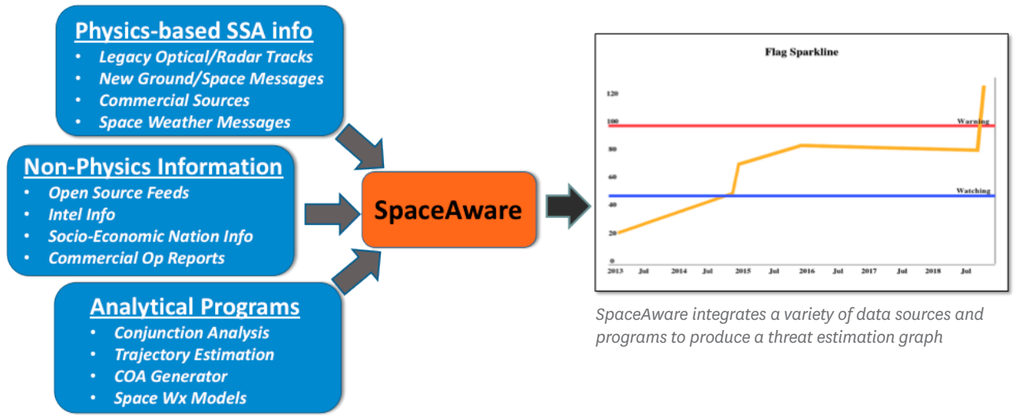
Machine learning methods are famous for leveraging the patterns they find in data to make predictions about the data. It turns out, however, that not all data is created equal, and simple transformations of the data can dramatically reduce the ability of machine learning algorithms to find predictive patterns. Researchers at ISI made this discovery while trying to solve a different problem. As part of an IARPA-sponsored project, they are developing methods for predicting cyber attacks by combining information about past cyber attacks with data harvested from the open and the dark web about potential hacker activities. However, even the most advanced machine learning methods can not beat a simple baseline. Finding the reason for this led to a new insight. Successful cyber attacks, such as those that our researchers were trying to learn from, represent just a tiny fraction of attempted attacks. This is because firewalls, spam filters, and other organizational defenses, stop the vast fraction of attempted cyber attacks. Whereas the attempted attacks have patterns that machine learning algorithms can leverage for predicting new attacks, once these attacks are filtered by the defenses, the patterns are lost. Our research team showed that the loss of information due to filtering is irreversible. Even if there exists some external signal that is highly correlated with what you want to predict, after filtering, the predictive value of the external signal diminishes. This work points to the difficulties of learning from filtered or sampled data. How does this apply to predicting cyber attacks? To sum it up, you are much better off predicting attempted attacks, rather than focusing on what gets through the firewall and ends up in a user's inbox.



Prediction is not that easy. Firewalls filter out most cyber attacks. Predicting future attacks based on the observed (successful) attacks becomes more difficult as the firewalls get better.

SpaceAware

There are a growing number of satellites that are launched into orbit every year; currently there are more than eight thousand active satellites orbiting the Earth. We are developing a system called SpaceAware that continually evaluates the potential threat posture of all satellites and space systems in orbit. To perform this task, we have developed a knowledge graph of space objects that combines physics-based space-situational awareness information (such as optical/radar tracks, space weather, etc.), non-physics information (such as news articles, intel reports, Satellite Catalog, Wikipedia, mailing lists, etc.), and capability-based analytical programs (such as conjunction analysis, trajectory analysis, etc.). All of this information is integrated to provide a comprehensive resource with detailed data about every satellite. Some of the technologies that we are developing and applying to create SpaceAware include extraction of various types of information from text documents (such as satellite name, orbital regimes, mission types, space-related events, etc.), entity resolution to determine the identity discussed in a text document, machine learning of both physics and non-physics characteristics to compare against claimed capabilities, and network analysis to identify possible connections to other entities of interest. All of this compiled information is continuously evaluated to produce a threat-estimation graph for each satellite, which in turn can be used to provide both watch and warning lists of space objects.



SpaceAware integrates a variety of data sources and programs to produce a threat estimation graph

COMPUTATIONAL SYSTEMS AND TECHNOLOGY

Adiabatic Quantum Computation
Tameem Albash and **Daniel A. Lidar**
Journal of Rev. Mod. Phys., 2018

Demonstration of a Scaling Advantage for a Quantum Annealer over Simulated Annealing
Tameem Albash and **Daniel A. Lidar**
Journal of Phys. Rev. X, 2018

Combating Workflow Failures with Integrity-based Checkpoints and Blockchain
Omkar Bhide, Raquel Hill, **Karan Vahi, Mats Rynge** and Von Velch
Proceedings of 6th International Workshop on Distributed Storage and Blockchain Technologies for Big Data, 2018

WRENCH: A Framework for Simulating Workflow Management Systems
Henri Casanova, Suraj Pandey, James Oeth, Ryan Tanaka, Frederic Suter and **Rafael Ferreira Da Silva**
Proceedings of 13th Workshop on Workflows in Support of Large-Scale Science (WORKS 2018)

Pacer: Automated Feedback-Based Vertical Elasticity for Heterogeneous Soft Real-Time Workloads
Yu-An Chen, Geoffrey Tran, Andrew Rittenbach, John Paul Walters and **Stephen Crago**
Proceedings of 2018 IEEE/ACM 11th International Conference on Utility and Cloud Computing (UCC)

The Future of Scientific Workflows
Ewa Deelman, Tom Peterka, Ilkay Altintas, Christopher D Carothers, Kerstin Kleese van Dam, Kenneth Moreland, Manish Parashar, Lavanya Ramakrishnan, et al.
International Journal of High Performance Computing Applications, 2018

Enabling Data Analytics Workflows using Node-Local Storage
Tu Mai Anh Do, Ming Jiang, Brian Gallagher, Albert Chu, Cyrus Harrison, **Karan Vahi** and **Ewa Deelman**
Proceedings of International Conference for High Performance Computing, Networking, Storage, and Analysis (SC18), Poster, 2018

Graphic Encoding of Macromolecules for Efficient High-Throughput Analysis
Trilce Estrada, Jeremy Benson, Hector Carrillo-Cabada, Asghar Razavi, Michel Cuendet, Harel Weinstein, **Ewa Deelman**, Michela Taufer, et al.
Proceedings of 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics

IoT-Hub: New IoT Data-Platform for Virtual Research Environments
Rosa Filgueira, **Rafael Ferreira Da Silva, Ewa Deelman**, Vyron Christodoulou and Amrey Krause
Proceedings of 10th International Workshop on Science Gateways (IWSG 2018)

Quantum-Limited Discrimination Between Laser Light and Noise
Jonathan Habif and Saikat Guha
Proceedings of Laser Science, 2018

Independent Functional Testing of Commercial FPGA Devices
Travis Haroldsen, Matthew French and **Andrew Schmidt**
Proceedings of NRO Verification Sciences and Engineering Workshop, 2018

Off-Diagonal Series Expansion for Quantum Partition Functions
Itay Hen
Journal of Statistical Mechanics: Theory and Experiment, 2018

Solving Quantum Spin Glasses with Off-Diagonal Expansion Quantum Monte Carlo
Itay Hen and **Tameem Albash**
Journal of Physics: Conference Series, 2018

Searching the Sequence Read Archive Using Jetstream and Wrangler
Kyle Levi, **Mats Rynge**, Eroma Abeysinghe and Robert A. Edwards
Proceedings of Practice and Experience on Advanced Research Computing, 2018

Finite Temperature Quantum Annealing Solving Exponentially Small Gap Problem with Non-Monotonic Success Probability
Anurag Mishra, **Tameem Albash** and **Daniel A. Lidar**
Journal of Nature Communications, 2018

A Study of Complex Deep Learning Networks on High-Performance, Neuromorphic, and Quantum Computers
Thomas E. Potok, Catherine D. Schuman, Steven R. Young, Robert M. Patton, **Federico Spedalieri, Jeremy Liu, Ke-Thia Yao**, Garrett S. Rose, et al.
Journal of JETC, 2018

Towards Model Integration via Abductive Workflow Composition and Multi-Method Scalable Model Execution
Rafael Ferreira Da Silva, Daniel Garijo, Scott D. Peckham, **Yolanda Gil, Ewa Deelman** and Varun Ratnakar
Proceedings of Ninth International Congress on Environmental Modeling and Software, 2018

Hot and Spicy: Improving Productivity with Python and HLS for FPGAs
S. Skalicky, **Joshua Monson, Andrew Schmidt** and **Matthew French**
Proceedings of 2018 IEEE 26th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM)

Quantum Annealing of the p-spin Model Under Homogeneous Transverse Field Driving
Yuki Susa, Yu Yamashiro, Masayuki Yamamoto, **Itay Hen**, **Daniel A. Lidar** and Hidetoshi Nishimori
Journal of Phys. Rev. A, 2018

Homogenizing OSG and XSEDE: Providing Access to XSEDE Allocations Through OSG Infrastructure
Suchandra Thapa, Robert W. Gardner, Kenneth Herner, Dirk Hufnagel, David Lesny and **Mats Rynge**
Proceedings of Practice and Experience on Advanced Research Computing, 2018

A Job Sizing Strategy for High-Throughput Scientific Workflows
Benjamin Tovar, **Rafael Ferreira Da Silva**, Gideon Juve, **Ewa Deelman**, William Allcock, Douglas Thain and Miron Livny
Journal of IEEE Transactions on Parallel and Distributed Systems, 2018

Reducing Tail Latencies While Improving Resiliency to Timing Errors for Stream Processing Workloads
Geoffrey Tran, John Paul Walters and **Stephen Crago**
Proceedings of 2018 IEEE/ACM 11th International Conference on Utility and Cloud Computing (UCC)

Workflows using Pegasus: Enabling Dark Energy Survey Pipelines
Karan Vahi, Michael H. Wang, Chihway Chang, Scott Dodelson, **Mats Rynge** and **Ewa Deelman**
Proceedings of 28th annual international Astronomical Data Analysis Software & Systems (ADASS), 2018

Surveying the Hardware Trojan Threat Landscape for the Internet-of-Things
Vivek Venugopalan and Cameron D. Patterson
Journal of Hardware and Systems Security, 2018

Quantum trajectories for time-dependent adiabatic master equations
Ka Wa Yip, **Tameem Albash** and **Daniel A. Lidar**
Journal of Phys. Rev. A, 2018

Improved Metrics for Obfuscated ICs, Government Microcircuit Applications and Critical Technology Conference (GOMACTech)
Mutian Zhu, **Matthew French** and Peter A. Beerel
Proceedings of Government Microcircuit Applications and Critical Technology Conference (GOMACTech), 2018

ARTIFICIAL INTELLIGENCE

Wearable RGB Camera-based Navigation System for the Visually Impaired.
Reham Abobeah, **Mohamed Hussein**, Moataz M. Abdelwahab and Amin Shoukry
Proceedings of VISIGRAPP (5: VISAPP), 2018

Advances in Data Science
Leman Akoglu, **Emilio Ferrara**, Mallayya Deivamani, Ricardo Baeza-Yates and Palanisamy Yogesh
Third Intl Conference on Intelligent Information Technologies, Chennai, India; 11-14 December 2018

Can you Trust the Trend?: Discovering Simpsons Paradoxes in Social Data
Nazanin Alipourfard, Peter G. Fennell and **Kristina Lerman**
Proceedings of Eleventh ACM International Conference on Web Search and Data Mining, 2018

Using Simpson's Paradox to Discover Interesting Patterns in Behavioral Data
Nazanin Alipourfard, Peter G. Fennell and **Kristina Lerman**
Proceedings of Twelfth International AAAI Conference on Web and Social Media, 2018

Adaptive Decision Making via Entropy Minimization
Armen E. Allahverdyan, **Aram Galstyan**, Ali E. Abbas and Zbigniew R. Struzik
International Journal of Approximate Reasoning, 2018

PUBLICATIONS | 2018

Could Social Bots Pose a Threat to Public Health?
Jon-Patrick Allem and **Emilio Ferrara**
American Journal of Public Health, 2018

Automating Ontology Engineering Support Activities with OnToology
Ahmad Alobaid, **Daniel Garijo**, Maria Poveda-Villalon, Idafen Santana-Perez, Alba Fernandez-Izquierdo and Oscar Corcho
Journal of Web Semantics, 2018

Scaling Up Data Integration and Analysis of Sensor Data for Pediatric Asthma
José-Luis Ambite
Proceedings of International Society of Exposure Science and International Society for Environmental Epidemiology (ISES-ISEE) *Joint Annual Meeting, 2018*

BD2K Training Coordinating Center's ERuDIte: the Educational Resource Discovery Index for Data Science
José-Luis Ambite, Lily Fierro, Jonathan Gordon, **Gully A. Burns**, Florian Geigl, **Kristina Lerman** and John D. Van Horn
Journal of IEEE Transactions on Emerging Topics in Computing, Special Issue on Scholarly Big Data, 2018

Title: Proceedings of 12th International Workshop on Semantic Evaluation
Marianna Apidianaki, Saif M. Mohammad, **Jonathan May**, Ekaterina Shutova, Steven Bethard and Marine Carpuat, *2018*

Mining and Forecasting Career Trajectories of Music Artists
Shushan Arakelyan, **Fred Morstatter**, Margaret Martin, **Emilio Ferrara** and **Aram Galstyan**
Proceedings of 29th Conference on Hypertext and Social Media, 2018

Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign
Adam Badawy, Emilio Ferrara and **Kristina Lerman**
Proceedings of 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2018

The Rise of Jihadist Propaganda on Social Networks
Adam Badawy and **Emilio Ferrara**
Journal of Computational Social Science, 2018

The Role of Type and Source of Uncertainty on the Processing of Climate Models Projections
Daniel Benjamin and David V. Budescu
Journal of Frontiers in Psychology, 2018

Partial Shocks on Cooperative Multiplex Networks with Varying Degrees of Noise
Keith Burghardt and Zeev Maoz
Journal of Scientific Reports, 2018

Quantifying the Impact of Cognitive Biases in Question-Answering Systems
Keith Burghardt, Tad Hogg and **Kristina Lerman**
Proceedings of Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018)

Towards Evidence Extraction: Analysis of Scientific Figures from Studies of Molecular Interactions
Gully Burns, Xiangyang Shi, **Yue Wu**, **Huaigu Cao** and **Prem Natarajan**
Proceedings of Second Workshop on Enabling Open Semantic Science (co-located with 17th International Semantic Web *Conference, SemSci@ISWC), Monterey, California, USA, October 8-12th, 2018*

Semantic Software Metadata for Workflow Exploration and Evolution
Lucas Carvalho, **Daniel Garijo**, Claudia Bauzer Medeiros and **Yolanda Gil**
Proceedings of Fourteenth IEEE International Conference on eScience, 2018

Clustering System Data Using Aggregate Measures
Johnnie Chang, Robert Chen, **Jay Pujara** and Lise Getoor
Proceedings of SysML Conference, 2018

Recurrent Neural Networks as Weighted Language Recognizers
Yining Chen, Sorcha Gilroy, Andreas Maletti, **Jonathan May** and **Kevin Knight**
Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language *Technologies, Volume 1 (Long Papers), 2018*

Image-to-GPS Verification Through A Bottom-Up Pattern Matching Network
Jiaxin Cheng, **Yue Wu, Wael Abd-Almageed** and **Prem Natarajan**
Proceedings of Asian Conference on Computer Vision, 2018

PSM-Flow: Probabilistic Subgraph Mining for Discovering Reusable Fragments in Workflows
Chin Wang Cheong, **Daniel Garijo**, Kwok-Wai Cheung and **Yolanda Gil**
Proceedings of 2018 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2018, Santiago, Chile, 3-6 December, 2018

Anytime Focal Search with Applications
Liron Cohen, Matias Greco, Hang Ma, Carlos Hernandez, Ariel Felner, **Satish Kumar Thittamaranahalli** and Sven Koenig
Proceedings of Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13-19 July, 2018

Rapid Randomized Restarts for Multi-Agent Path Finding Solvers
Liron Cohen, Glenn Wagner, David M Chan, Howie Choset, Nathan Sturtevant, Sven Koenig and **Satish Kumar Thittamaranahalli**
Proceedings of the Eleventh International Symposium on Combinatorial Search, Stockholm, Sweden, 14-15 July 2018

Rapid Randomized Restarts for Multi-Agent Path Finding: Preliminary Results
Liron Cohen, Sven Koenig, **Satish Kumar Thittamaranahalli**, Glenn Wagner, Howie Choset, David M. Chan and Nathan Sturtevant
Proceedings of 17th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS, Stockholm, Sweden, 10-15 July, 2018

The FastMap Algorithm for Shortest Path Computations
Liron Cohen, Tansel Uras, Shiva Jahangiri, Aliyah Arunasalam, Sven Koenig and **Satish Kumar Thittamaranahalli**
Proceedings of Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI, Stockholm, Sweden, 13-19 July, 2018

Automatic Classification of Cortical Thickness Patterns in Alzheimer's Disease Patients Using the Louvain Modularity Clustering Method
Fabian W. Corlier, **Daniel Moyer**, Meredith N. Braskie, Paul M. Thompson, Guillaume Dorothee, Marie Claude Potier, Marie Sarazin, Michel Bottlaender, et al.
Proceedings of 14th International Symposium on Medical Information Processing and Analysis, 2018

Predicting Cyber Events by Leveraging Hacker Sentiment
Ashok Deb, **Kristina Lerman** and **Emilio Ferrara**
Journal of Information, 2018

Social Bots for Online Public Health Interventions
Ashok Deb, Anuja Majmundar, Sungyong Seo, **Akira Matsui, Rajat Tandon, Shen Yan**, Jon-Patrick Allem, **Emilio Ferrara**, et al.
Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2018

Automatic Generation of Precisely Delineated Geographic Features from Georeferenced Historical Maps Using Deep Learning
Weiwei Duan, Yao-Yi Chiang, **Craig Knoblock**, Stefan Leyk and Johannes H. Uhl
Proceedings of 22nd International Research Symposium on Computer-based Cartography and GIScience (Autocarto/UCGIS), 2018

Senator, We Sell Ads: Analysis of the 2016 Russian Facebook Ads Campaign
Ritam Dutt, **Ashok Deb** and **Emilio Ferrara**
Proceedings of Third International Conference on Intelligent Information Technologies (ICIIT 2018)

Human Action Recognition Based on Integrating Body Pose, Part Shape, and Motion
Hany El-Ghaish, **Mohamed Hussein**, Amin Shoukry and Rikio Onai
Journal of IEEE Access, 2018

CovP3DJ: Skeleton-Parts-Based-Covariance Descriptor for Human Action Recognition
Hany A El-Ghaish, Amin Shoukry and **Mohamed Hussein**
Proceedings of VISIGRAPP (5: VISAPP), 2018

Multi-Modality-Based Arabic Sign Language Recognition
Marwa Elpeltagy, Moataz Abdelwahab, **Mohamed Hussein**, Amin Shoukry, Asmaa Shoala and Moustafa Galal
IET Computer Vision, 2018

Aligning Product Categories Using Anchor Products
Varun Embar, Golnoosh Farnadi, **Jay Pujara** and Lise Getoor
Proceedings of WSDM Workshop on Knowledge Base Construction, Reasoning and Mining, 2018

LinkedEarth: Supporting Paleoclimate Data Standards and Crowd Curation
Julien Emile-Geay, **Deborah Khider**, Nicholas P. McKay, **Yolanda Gil, Daniel Garijo** and Varun Ratnakar
Journal of Past Global Changes (PAGES) Magazine, 2018

Adding Heuristics to Conflict-Based Search for Multi-Agent Path Finding
Ariel Felner, Jiaoyang Li, Eli Boyarski, Hang Ma, Liron Cohen, **Satish Kumar Thittamaranahalli** and Sven Koenig
Proceedings of Twenty-Eighth International Conference on Automated Planning and Scheduling, ICAPS 2018, The Netherlands, 24-29 June, 2018

Measuring Social Spam and the Effect of Bots on Information Diffusion in Social Media
Emilio Ferrara
Proceedings of Complex Spreading Phenomena in Social Systems, 2018

Constraint Composite Graph-Based Lifted Message Passing for Distributed Constraint Optimization Problems
Ferdinando Fioretto, Hong Xu, Sven Koenig and **Satish Kumar Thittamaranahalli**
Proceedings of International Symposium on Artificial Intelligence and Mathematics, Fort Lauderdale, Florida, USA, 3-5 January, 2018

Solving Multiagent Constraint Optimization Problems on the Constraint Composite Graph
Ferdinando Fioretto, Hong Xu, Sven Koenig and **Satish Kumar Thittamaranahalli**
Proceedings of Principles and Practice of Multi-Agent Systems - 21st International Conference, Tokyo, Japan, 29 December-2 November, 2018

Style Transfer in Text: Exploration and Evaluation
Zhenxin Fu, Xiaoye Tan, **Nanyun Peng**, Dongyan Zhao and Rui Yan
Proceedings of Thirty-Second Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence (AAAI), 2018

A Semantic Model Catalog to Support Comparison and Reuse
Daniel Garijo, Deborah Khider, Yolanda Gil, Lucas Carvalho, Bakinam Essawy, Suzanne Pierce, Daniel Hardesty Lewis, Varun Ratnakar, et al.
Proceedings of 9th International Congress on Environmental Modelling and Software, 2018

Integrating Models Through Knowledge-Powered Data and Process Composition
Daniel Garijo, Yolanda Gil, K.M. Cobourn, **Ewa Deelman**, C. Duffy, **Rafael Ferreira Da Silva**, A. Kemanian, **Craig Knoblock**, et al.
Journal of AGU Fall Meeting Abstracts, 2018

Unsupervised Record Matching with Noisy and Incomplete Data
Yves van Gennip, Blake Hunter, Anna Ma, **Daniel Moyer**, Ryan de Vera and Andrea L. Bertozzi
International Journal of Data Science and Analytics, 2018

MINT: Model Integration Through Knowledge-Powered Data and Process Composition
Yolanda Gil, Kelly Cobourn, **Ewa Deelman**, Chris Duffy, **Rafael Ferreira Da Silva**, Armen Kemanian, **Craig Knoblock**, Vipin Kumar, et al.
Proceedings of 9th International Congress on Environmental Modelling and Software, 2018

P4ML: A Phased Performance-Based Pipeline Planner for Automated Machine Learning
Yolanda Gil, Ke-Thia Yao, Varun Ratnakar, **Daniel Garijo, Greg Ver Steeg, Pedro Szekely**, **Robert Brekelmans, Mayank Kejriwal**, et al.
Proceedings of Machine Learning Research, ICML 2018 AutoML Workshop

Intelligent Systems for Geosciences: An Essential Research Agenda
Yolanda Gil, Suzanne A. Pierce, Hassan Babaie, Arindam Banerjee, Kirk Borne, Gary Bust, Michelle Cheatham, Imme Ebert-Uphoff, et al.
Journal of Communications of the ACM, 2018

Capturing Edge Attributes via Network Embedding
Palash Goyal, Homa Hosseinmardi, Emilio Ferrara and **Aram Galstyan**
Journal of IEEE Transactions on Computational Social Systems, 2018

Embedding Networks with Edge Attributes
Palash Goyal, Homa Hosseinmardi, Emilio Ferrara and **Aram Galstyan**
Proceedings of 29th Conference on Hypertext and Social Media, 2018

GEM: A Python Package for Graph Embedding Methods
Palash Goyal and **Emilio Ferrara**
Journal of Open Source Software, 2018

Graph Embedding Techniques,Aapplications, and Performance: A Survey
Palash Goyal and **Emilio Ferrara**
Journal of Knowledge-Based Systems, 2018

Recommending Teammates with Deep Neural Networks
Palash Goyal, Anna Sapienza and **Emilio Ferrara**
Proceedings of 29th Conference on Hypertext and Social Media, 2018

On Reproducible AI: Towards Reproducible Research, Open Science, and Digital Scholarship in AI Publications
Odd Erik Gundersen, **Yolanda Gil** and David W. Aha
Journal of AI Magazine, 2018

Feature Selection Methods For Understanding Business Competitor Relationships
Rahul Gupta, **Jay Pujara, Craig Knoblock**, Shushyam M. Sharanappa, **Bharat Pulavarti**, Gerard Hoberg and Gordon Phillips
Proceedings of Fourth International Workshop on Data Science for Macro-Modeling with Financial and Economic Datasets, 2018

Identifying Motion Pathways in Highly Crowded Scenes: A Non-Parametric Tracklet Clustering Approach
Allam S. Hassanein, **Mohamed Hussein**, Walid Gomaa, Yasushi Makihara and Yasushi Yagi
Journal of Computer Vision and Image Understanding, 2018

Hidden in Plain Sight: A Machine Learning Approach for Detecting Prostitution Activity in Phoenix, Arizona
Edward Helderop, Jessica Huff, **Fred Morstatter**, Anthony Grubesic and Danielle Wallace
Journal of Applied Spatial Analysis and Policy, 2018

Out-of-the-Box Universal Romanization Tool *uroman*
Ulf Hermjakob, Jonathan May and **Kevin Knight**
Proceedings of 56th Annual Meeting of Association for Computational Linguistics, Demo Track, 2018

Translating a Language You Don't Know in the Chinese Room
Ulf Hermjakob, Jonathan May, Michael Pust and **Kevin Knight**
Proceedings of 56th Annual Meeting of Association for Computational Linguistics, Demo Track, 2018

Incident-Driven Machine Translation and Name Tagging for Low-Resource Languages
Ulf Hermjakob, Qiang Li, Daniel Marcu, **Jonathan May**, Sebastian J. Mielke, Nima Pourdamghani, **Michael Pust**, Xing Shi, et al.
Journal of Machine Translation, 2018

Model of Cognitive Dynamics Predicts Performance on Standardized Tests
Nathan O. Hodas, Jacob Hunter, Stephen J. Young and **Kristina Lerman**
Journal of Computational Social Science, 2018

Trajectory Planning for Quadrotor Swarms
Wolfgang Honig, James A. Preiss, **Satish Kumar Thittamaranahalli**, Gaurav S. Sukhatme and Nora Ayanian
Journal of IEEE Trans. Robotics, 2018

Forecasting Violent Events in the Middle East and North Africa Using the Hidden Markov Model and Regularized Autoregressive Models
KSM Tozammel Hossain, Shuyang Gao, Brendan Kennedy, **Aram Galstyan** and **Prem Natarajan**
The Journal of Defense Modeling and Simulation, 2018

Discovering Hidden Structure in High Dimensional Human Behavioral Data via Tensor Factorization
Homa Hosseinmardi, Hsien-Te Kao, Kristina Lerman and **Emilio Ferrara**
Proceedings of HeteroNAM 2018: First International Workshop on Heterogeneous Networks Analysis and Mining

Always Lurking: Understanding and Mitigating Bias in Online Human Trafficking Detection
Kyle Hundman, **Thamme Gowda, Mayank Kejriwal** and Benedikt Boecking
Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society

Fingerprint Presentation Attack Detection Using A Novel Multi-Spectral Capture Device and Patch-Based Convolutional Neural Networks
Mohamed Hussein, Leonidas Spinoulas, Fei Xiong and **Wael Abd-Almageed**
Proceedings of 2018 IEEE International Workshop on Information Forensics and Security (WIFS)

Bidirectional Conditional Generative Adversarial Networks
Ayush Jaiswal, Yue Wu, Wael Abd-Almageed and **Prem Natarajan**
Proceedings of Asian Conference on Computer Vision, 2018

Unsupervised Adversarial Invariance
Ayush Jaiswal, Yue Wu, Wael Abd-Almageed and **Prem Natarajan**
Proceedings of Advances in Neural Information Processing Systems (NIPS), 2018

CapsuleGAN: Generative Adversarial Capsule Network
Ayush Jaiswal, Wael Abd-Almageed and **Prem Natarajan**
Journal of European Conference on Computer Vision Workshop on Brain-Driven Computer Vision, 2018

Combinatorial Problems in Multirobot Battery Exchange Systems
Nitin Kamra, **Satish Kumar Thittamaranahalli** and Nora Ayanian
Journal of IEEE Trans. Automation Science and Engineering, 2018

PUBLICATIONS | 2018

Discovering Latent Psychological Structures from Self-report Assessments of Hospital Workers
Hsien-Te Kao, Homa Hosseinmardi, Shen Yan, Michelle Hasan, Shrikanth Narayanan, Kristina Lerman and **Emilio Ferrara**
Proceedings of 5th International Conference on Behavioral, Economic, and Socio-Cultural Computing, 2018

Constructing Domain-Specific Search Engines with No Programming
Mayank Kejriwal and **Pedro Szekely**
Proceedings of Thirty-Second AAAI Conference on Artificial Intelligence, 2018

Technology-Assisted investigative Search: A Case Study from an Illicit Domain
Mayank Kejriwal and **Pedro Szekely**
Proceedings of Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, 2018

Thor: Text-Enabled Analytics for Humanitarian Operations
Mayank Kejriwal, Daniel Gilley, **Pedro Szekely** and Jill Crisman
Proceedings of Companion of The Web Conference, 2018

Structured Event Entity Resolution in Humanitarian Domains
Mayank Kejriwal, Jing Peng, Haotian Zhang and **Pedro Szekely**
Proceedings of International Semantic Web Conference, 2018

Investigative Knowledge Discovery for Combating Illicit Activities
Mayank Kejriwal, Pedro Szekely and **Craig Knoblock**
Journal of IEEE Intelligent Systems, 2018

Extracting the Multi-Timescale Activity Patterns of Online Financial Markets
Teruyoshi Kobayashi, **Anna Sapienza** and **Emilio Ferrara**
Journal of Scientific Reports, 2018

Collective Entity Resolution in Multi-Relational Familial Networks
Pigi Kouki, **Jay Pujara**, Christopher Marcum, Laura Koehly and Lise Getoor
Journal of Knowledge and Information Systems, 2018

Deep Neural Networks for Bot Detection
Sneha Kudugunta and **Emilio Ferrara**
Journal of Information Sciences, 2018

Connectivity-Driven Brain Parcellation via Consensus Clustering
Anvar Kurmukov, Ayagoz Musabaeva, Yulia Denisova, **Daniel Moyer** and Boris Gutman
Proceedings of International Workshop on Connectomics in Neuroimaging, 2018

A Forest Mixture Bound for Block-Free Parallel Inference
Neal Lawton, Greg Ver Steeg and **Aram Galstyan**
Proceedings of Uncertainty in Artificial Intelligence Conference, 2018

Computational Social Scientist Beware: Simpsons Paradox in Behavioral Data
Kristina Lerman
Journal of Computational Social Science, 2018

Language, Demographics, Emotions, and the Structure of Online Social Networks
Kristina Lerman, Luciano G. Marin, Megha Arora, Lucas H. Costa de Lima, **Emilio Ferrara** and David Garcia
Journal of Computational Social Science, 2018

Weighted Feature Pooling Network in Template-Based Recognition
Zekun Li, Yue Wu, Wael Abd-Almageed and **Prem Natarajan**
Proceedings of Asian Conference on Computer Vision, 2018

Feature Selection: A Data Perspective
Jundong Li, Kewei Cheng, Suhang Wang, **Fred Morstatter**, Robert P. Trevino, Jiliang Tang and Huan Liu
Journal of ACM Computing Surveys (CSUR), 2018

Exploiting Spatiotemporal Patterns for Accurate Air Quality Forecasting using Deep Learning
Yijun Lin, Nikhit Mago, Yu Gao, Yaguang Li, Yao-Yi Chiang, Cyrus Shahabi and **José-Luis Ambite**
Proceedings of 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2018

Self-Organization of Dragon King Failures
Yuansheng Lin, **Keith Burghardt**, Martin Rohden, Pierre-Andre Noel and Raissa M. D'Souza
Journal of Phys. Rev. E, 2018

Building Linked Data from Historical Maps
Chun Lin, Hang Su, **Craig Knoblock**, Yao-Yi Chiang, Weiwei Duan, Stefan Leyk and Johannes H. Uhl
Proceedings of ISWC 2018 Workshop on Enabling Open Semantic Science (SemSci), 2018

Adiabatic Quantum Computation Applied to Deep Learning Networks
Jeremy Liu, Federico Spedalieri, Ke-Thia Yao, Thomas E. Potok, Catherine D. Schuman, Steven R. Young, Robert M. Patton, Garrett S. Rose, et al.
Journal of Entropy, 2018

Multi-Agent Path Finding with Deadlines
Hang Ma, Glenn Wagner, Ariel Felner, Jiaoyang Li, **Satish Kumar Thittamaranahalli** and Sven Koenig
Proceedings of Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI, Stockholm, Sweden, 13-19 July, 2018

Multi-Agent Path Finding with Deadlines: Preliminary Results
Hang Ma, Glenn Wagner, Ariel Felner, Jiaoyang Li, **Satish Kumar Thittamaranahalli** and Sven Koenig
Proceedings of 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS), Stockholm, Sweden, 10-15 July, 2018

Lifelong Path Planning with Kinematic Constraints for Multi-Agent Pickup and Delivery
Hang Ma, Wolfgang Honig, **Satish Kumar Thittamaranahalli**, Nora Ayanian and Sven Koenig
Journal of CoRR, 2018

Overview: A Hierarchical Framework for Plan Generation and Execution in Multi-Robot Systems
Hang Ma, Wolfgang Honig, Liron Cohen, Tansel Uras, Hong Xu, **Satish Kumar Thittamaranahalli**, Nora Ayanian, Sven Koenig, et al.
Journal of CoRR, 2018

Stack-Pointer Networks for Dependency Parsing
Xuezhe Ma, Zecong Hu, Jingzhou Liu, **Nanyun Peng**, Graham Neubig and Eduard Hovy
Proceedings of 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018

Learning Pose-Aware Models for Pose-Invariant Face Recognition in the Wild
Iacopo Masi, Feng-ju Chang, Jongmoo Choi, Shai Harel, Jungyeon Kim, KangGeon Kim, Jatuporn Leksut, **Stephen Rawls**, et al.
Journal of IEEE Trans. on Pattern Analysis and Machine Intelligence, 2018

Playing Under the Influence (of Twitch): The Effects of Streaming on Gamers Performance
Akira Matsui, Anna Sapienza and **Emilio Ferrara**
Proceedings of UCI Esports Conference (ESC 2018)

ENIGMA-ODS: A Platform for Global Neuroscience Collaborations in the ENIGMA Consortium
Agnes McMahon, **Daniel Garijo**, Ryan Espiritu, Faisal Rashid, MiHyun Jang, Tejal Patted, Varun Ratnakar, **Yolanda Gil**, et al.
Proceedings of INCF Neuroinformatics Congress, 2018

Modeling Evolution of Topics in Large-Scale Temporal Text Corpora
Elaheh Momeni, Shanika Karunasekera, **Palash Goyal** and **Kristina Lerman**
Proceedings of Twelfth International AAAI Conference on Web and Social Media, 2018

From Alt-Right to Alt-Rechts: Twitter Analysis of the 2017 German Federal Election
Fred Morstatter, Yunqiu Shao, **Aram Galstyan** and Shanika Karunasekera
Proceedings of Companion of the The Web Conference, 2018

Identifying Framing Bias in Online News
Fred Morstatter, Liang Wu, Uraz Yavanoglu, Stephen R. Corman and Huan Liu
Journal of ACM Transactions on Social Computing, 2018

Invariant Representation Without Adversarial Training
Daniel Moyer, Shuyang Gao, **Robert Brekelmans**, **Greg Ver Steeg** and **Aram Galstyan**
Proceedings of Advances In Neural Information Processing Systems, 2018

Measures of Tractography Convergence
Daniel Moyer, Paul M. Thompson and **Greg Ver Steeg**
Proceedings of International Conference On Medical Image Computing & Computer Assisted Intervention (MICCAI), CDMRI Workshop, 2018

Towards Quantifying Sampling Bias in Networks
Lisette Espin Noboa, Claudia Wagner, Fariba Karimi and **Kristina Lerman**
Proceedings of International Workshop on Mining Attributed Networks WWW'2018

Towards Understanding the Min-Sum Message Passing Algorithm for the Minimum Weighted Vertex Cover Problem: An Analytical Approach
Masaru Nakajima, Hong Xu, Sven Koenig and **Satish Kumar Thittamaranahalli**
Proceedings of International Symposium on Artificial Intelligence and Mathematics (ISAIM), Fort Lauderdale, Florida, USA,3-5 January, 2018

How Do Organizations Publish Semantic Markup? Three Case Studies Using Public Schema. org Crawls
Daye Nam and **Mayank Kejriwal**
Journal of Computer, 2018

Towards Controllable Story Generation
Nanyun Peng, Marjan Ghazvininejad, Jonathan May and **Kevin Knight**
Proceedings of First Workshop on Storytelling, 2018

Tampering with Twitter's Sample API
Jurgen Pfeffer, Katja Mayer and **Fred Morstatter**
Journal of EPJ Data Science, 2018

Sulcal-Based Morphometry in Parkinson's Disease: A Study of Reliability and Disease Effects
Fabrizio Pizzagalli, Guillaume Auzias, Armand Amini, Joshua Faskowitz, Faisal Rashid, **Daniel Moyer**, Peter Kochunov, Denis Riviere, et al.
Proceedings of 14th International Symposium on Medical Information Processing and Analysis, 2018

Hybrid Link Prediction for Competitor Relationships
Jay Pujara
Proceedings of Fourth International Workshop on Data Science for Macro-Modeling with Financial and Economic Datasets, 2018

How To Efficiently Increase Resolution in Neural OCR Models
Stephen Rawls, Huaigu Cao, Joe Mathai and **Prem Natarajan**
Proceedings of IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition, ASAR 2018, London, UK, 12-14 March, 2018

Uncovering Biologically Coherent Peripheral Signatures of Health and Risk for Alzheimers Disease in the Aging Brain
Brandalyn Cherish Riedel, Madelaine Daianu, **Greg Ver Steeg**, Adam Mezher, Lauren E. Salminen, **Aram Galstyan** and Paul Matthew Thompson
Journal of Frontiers in Aging Neuroscience, 2018

Deep Multimodal Image-Repurposing Detection
Ekraam Sabir, **Wael Abd-Almageed**, **Yue Wu** and **Prem Natarajan**
Proceedings of ACM Multimedia, 2018

DISCOVER: Mining Online Chatter for Emerging Cyber Threats
Anna Sapienza, Sindhu Kiranmai Ernala, Alessandro Bessi, **Kristina Lerman** and **Emilio Ferrara**
Proceedings of Companion of the The Web Conference, 2018

Individual Performance in Team-Based Online Games
Anna Sapienza, **Yilei Zeng**, Alessandro Bessi, **Kristina Lerman** and **Emilio Ferrara**
Journal of Royal Society Open Science, 2018

Non-negative Tensor Factorization for Human Behavioral Pattern Mining in Online Games
Anna Sapienza, Alessandro Bessi and **Emilio Ferrara**
Journal of Information, 2018

Distributed Reinforcement Learning for Multi-robot Decentralized Collective Construction
Guillaume Sartoretti, **Yue Wu**, William Paivine, **Satish Kumar Thittamaranahalli**, Sven Koenig and Howie Choset
Proceedings of Distributed Autonomous Robotic Systems, 14th International Symposium, DARS 2018, Boulder, CO, USA, 15-17 October, 2018

PRIMAL: Pathfinding via Reinforcement and Imitation Multi-Agent Learning
Guillaume Sartoretti, Justin Kerr, Yunfei Shi, Glenn Wagner, **Satish Kumar Thittamaranahalli**, Sven Koenig and Howie Choset
Journal of CoRR, 2018

Learning to Converse with Noisy Data: Generation with Calibration
Mingyue Shang, Zhenxin Fu, **Nanyun Peng**, Yansong Feng, Dongyan Zhao and Rui Yan
Proceedings of IJCAI, 2018

Constraint-Based Learning for Sensor Failure Detection and Adaptation
Yuan Shi, Satish Kumar Thittamaranahalli and **Craig Knoblock**
Proceedings of 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)

Scalable Probabilistic Causal Structure Discovery
Dhanya Sridhar, **Jay Pujara** and Lise Getoor
Proceedings of International Joint Conference on Artificial Intelligence, 2018

Using Noisy Extractions to Discover Causal Knowledge
Dhanya Sridhar, **Jay Pujara** and Lise Getoor
Proceedings of Sixth Workshop on Automated Knowledge Base Construction, 2018

Bots Increase Exposure to Negative and Inflammatory Content in Online Social Systems
Massimo Stella, **Emilio Ferrara** and Manlio De Domenico
Journal of Proceedings of the National Academy of Sciences, 2018

Domain-Specific Insight Graphs (dig)
Pedro Szekely and **Mayank Kejriwal**
Proceedings of Companion of the The Web Conference, 2018

BigNet 2018 Chairs Welcome & Organization
Jie Tang, Michalis Vazirgiannis, Yuxiao Dong, Fragkiskos D Malliaros, Michael Cochez, **Mayank Kejriwal** and Achim Rettinger
Proceedings of Companion of the The Web Conference, 2018

DarkEmbed: Exploit Prediction with Neural Language Models
Nazgol Tavabi, Palash Goyal, Mohammed Almukaynizi, Paulo Shakarian and **Kristina Lerman**
Proceedings of AAAI Conference on Innovative Applications of AI (IAAI), 2018

Alert Generation in Execution Monitoring Using Resource Envelopes
Satish Kumar Thittamaranahalli, Hong Xu, Zheng Tang, **Anoop Kumar**, **Craig Rogers** and **Craig Knoblock**
Proceedings of 31st International Florida Artificial Intelligence Research Society Conference (FLAIRS), 2018

Load Scheduling of Simple Temporal Networks Under Dynamic Resource Pricing
Satish Kumar Thittamaranahalli, Zhi Wang, **Anoop Kumar, Craig Rogers** and **Craig Knoblock**
Proceedings of Thirty-Second AAAI Conference on Artificial Intelligence (AAAI), 2018

On the Linear Programming Duals of Temporal Reasoning Problems
Satish Kumar Thittamaranahalli, Zhi Wang, **Anoop Kumar**, **Craig Rogers** and **Craig Knoblock**
Proceedings of Fifteenth International Symposium on Artificial Intelligence and Mathematics (ISAIM), 2018

Map Archive Mining: Visual-Analytical Approaches to Explore Large Historical Map Collections
Johannes H. Uhl, Stefan Leyk, Yao-Yi Chiang, Weiwei Duan and **Craig Knoblock**
ISPRS International Journal of Geo-Information, 2018

Spatialising Uncertainty in Image Segmentation Using Weakly Supervised Convolutional Neural Networks: A Case Study From Historical Map Processing
Johannes H. Uhl, Stefan Leyk, Yao-Yi Chiang, Weiwei Duan and **Craig Knoblock**
Journal of IET Image Processing, 2018

The Factored Shortest Path Problem and Its Applications in Robotics
Zhi Wang, Liron Cohen, Sven Koenig and **Satish Kumar Thittamaranahalli**
Proceedings of Twenty-Eighth International Conference on Automated Planning and Scheduling (ICAPS), The Netherlands, 24-29 June 2018

Image Copy-Move Forgery Detection via an End-to-End Deep Neural Network
Yue Wu, Wael Abd-Almageed and **Prem Natarajan**
Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)

BusterNet: Detecting Copy-Move Image Forgery with Source/Target Localization
Yue Wu, Wael Abd-Almageed and **Prem Natarajan**
Proceedings of European Conference on Computer Vision (ECCV), 2018

Degree Correlations Amplify the Growth of Cascades in Networks
Xin-Zeng Wu, Peter G. Fennell, **Allon Percus** and **Kristina Lerman**
Journal of Phys. Rev. E, 2018

SlangSD: Building, Expanding and Using a Sentiment Dictionary of Slang Words for Short-Text Sentiment Classification
Liang Wu, **Fred Morstatter** and Huan Liu
Journal of Language Resources and Evaluation, 2018

Toward Relational Learning with Misinformation
Liang Wu, Jundong Li, **Fred Morstatter** and Huan Liu
Proceedings of the 2018 SIAM International Conference on Data Mining

Unknown Presentation Attack Detection with Face RGB Images
F. Xiong and **Wael Abd-Almageed**
Proceedings of IEEE International Conference on Biometrics: Theory, Applications, and Systems, 2018

Message Passing Algorithms for Semiring-Based and Valued Constraint Satisfaction Problems
Hong Xu, Cheng Cheng, Sven Koenig and **Satish Kumar Thittamaranahalli**
Proceedings of Eleventh International Symposium on Combinatorial Search, Stockholm, Sweden, 14-15 July, 2018

The Buss Reduction for the k-Weighted Vertex Cover Problem
Hong Xu, Xin-Zeng Wu, Cheng Cheng, Sven Koenig and **Satish Kumar Thittamaranahalli**
Proceedings of International Symposium on Artificial Intelligence and Mathematics (ISAIM), Fort Lauderdale, Florida, USA, 2-5 January, 2018

Towards Effective Deep Learning for Constraint Satisfaction Problems
Hong Xu, Sven Koenig and **Satish Kumar Thittamaranahalli**
Proceedings of Principles and Practice of Constraint Programming - 24th International Conference, Lille, France, 27-31 August, 2018

A Warning Propagation-Based Linear-Time-and-Space Algorithm for the Minimum Vertex Cover Problem on Giant Graphs
Hong Xu, **Kexuan Sun**, Sven Koenig and **Satish Kumar Thittamaranahalli**
Proceedings of 15th International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research, The Netherlands, 26-29 June, 2018

ELISA-EDL: A Cross-lingual Entity Extraction, Linking and Localization System
Boliang Zhang, Ying Lin, Xiaoman Pan, Di Lu, **Jonathan May, Kevin Knight** and Heng Ji
Proceedings of 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations

NETWORKING AND CYBERSECURITY

Leveraging Controlled Information Sharing for Botnet Activity Detection
Calvin Ardi and **John Heidemann**
Proceedings of the ACM SIGCOMM Workshop on Traffic Measurements for Cybersecurity, 2018

FARM: Architecture for Distributed Agent-based Social Simulations
Jim Blythe and **Alexey Tregubov**
Proceedings of IJCAI/AAMAS Workshop on Massively Multi-Agent Systems, 2018

Usable Security vs. Workflow Realities
Jim Blythe, Vijay Kothari, Sean Smith and Ross Koppel
Proceedings of NDSS Workshop on Usable Security, 2018

Malware Analysis through High-Level Behavior
Xiyue Deng and **Jelena Mirkovic**
Proceedings of CSET, 2018

Who Knocks at the IPv6 Door? Detecting IPv6 Scanning
Kensuke Fukuda and **John Heidemann**
Proceedings of the ACM Internet Measurement Conference, 2018

Detecting ICMP Rate Limiting in the Internet
Hang Guo and **John Heidemann**
Proceedings of the Passive and Active Measurement Workshop, 2018

IP-Based IoT Device Detection
Hang Guo and **John Heidemann**
Proceedings of the ACM SIGCOMM Workshop on IoT Security and Privacy, 2018

Analyzing and Mitigating Privacy with the DNS Root Service
Wesley Hardaker
Proceedings of the SOC NDSS Workshop on DNS Privacy, 2018

The Policy Potential of Measuring Internet Outages
John Heidemann, Yuri Pradkin and Guillermo Baltra
Proceedings of TPRC, the Research Conference on Communication, Information and Internet Policy, 2018

Enumerating Privacy Leaks in DNS Data Collected Above the Recursive
Basileal Imana, Aleksandra Korolova and **John Heidemann**
Proceedings of the ISOC NDSS Workshop on DNS Privacy, 2018

Capturing Domain Knowledge Through Extensible Components
Erik Kline, Genevieve Bartlett, Geoff Lawler, Robert Story and **Michael Elkins**
Proceedings os Testbeds and Research Infrastructures for the Development of Networks and Communications. 13th EAI International Conference, TridentCom 2018, Shanghai, China, 1-3 December, 2018

Security for the Collective Reality of the Smart Home
Ross Koppel, **Jim Blythe**, Vijay Kothari and Sean Smith
Proceedings of SOUPS workshop on Human Aspects of Smarthome Security and Privacy, 2018

Data Privacy and the Elusive Goal of Empowering the User
Vijay Kothari, Sean Smith, **Jim Blythe** and Ross Koppel
Proceedings of CHI workshop on Exploring Individual Differences in Privacy, 2018

Concentrated vs. Distributed Defense as a Deterrent to Cyber Attackers: A Behavioral Analog Simulation Game
Sarah Kusumastuti, Heather Rosoff, **Jim Blythe** and Richard John
Proceedings of INFORMS Annual Meeting, 2018

A Binary Analysis Approach to Retrofit Security in Input Parsing Routines
Jayakrishna Menon, **Christophe Hauser**, Yan Shoshitaishvili and **Stephen Schwab**
Proceedings of 2018 IEEE Security and Privacy Workshops (SPW)

DEW: Distributed Experiment Workflows
Jelena Mirkovic, Genevieve Bartlett and **Jim Blythe**
Proceedings of 11th USENIX Workshop on Cyber Security Experimentation and Test (CSET 18), 2018

AARCLight New opportunities for South Atlantic R&E Network Collaboration Between Africa, Brazil, and the US
Heidi Morgan, Julio Ibarra, Jeronimo Bezerra, Luis Fernandez Lopez, Vasilka Chergarova, Donald A. Cox III, Gabriella E. Alvarez, Michael Stanton, et al.
Journal of UbuntuNet Connect Conference, 2018

NRE-18: Americas Lightpaths Express and Protect Enhances Infrastructure for Research and Education
Harvey Newman, **Heidi Morgan**, Julio Ibarra and Jeronimo Bezerra
Journal of Supercomputing Conference - SCinet Network Research Exhibition, 2018

Generalized Paxos Made Byzantine (and Less Complex)
Miguel Pires, **Srivatsan Ravi** and Rodrigo Rodrigues
Journal of Algorithms, 2018

SENSS Against Volumetric DDoS Attacks
Sivaramakrishnan Ramanathan, Jelena Mirkovic, Minlan Yu and Ying Zhang
Proceedings of the Annual Computer Security Applications Conference (ACSAC), 2018

SDProber: A Software Defined Prober for SDN
Sivaramakrishnan Ramanathan, Yaron Kanza and Balachander Krishnamurthy
Proceedings of the Symposium on SDN Research (SOSR), 2018

Evaluation of Cross-Project Multitasking in Software Projects
Alexey Tregubov, Jo Ann Lane and Barry Boehm
Proceedings of Disciplinary Convergence in Systems Engineering Research, 2018

Does Anycast Hang up on You (UDP and TCP)?
Lan Wei and **John Heidemann**
Journal of IEEE Transactions on Network and Service Management, 2018

GuidedPass: Helping Users to Create Strong and Memorable Passwords
Simon S. Woo and **Jelena Mirkovic**
Proceedings of the 21st International Symposium on Research in Attacks, Intrusions and Defenses (RAID), 2018

LDplayer: DNS Experimentation at Scale
Liang Zhu and **John Heidemann**
Proceedings of the ACM Internet Measurement Conference, 2018

INFORMATICS SYSTEMS RESEARCH

ERMrest: A Web Service for Collaborative Data Management
Karl Czajkowski, Carl Kesselman, Robert Schuler and **Hongsuda Tangmunarunkit**
Proceedings of 30th International Conference on Scientific and Statistical Database Management, 2018

MULTIDIVISION & RESEARCH CENTER

Dropout Approaches for LSTM Based Speech Recognition Systems
Jayadev Billa
Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2018, Calgary, Canada, 15-20 April, 2018

ISI ASR System for the Low Resource Speech Recognition Challenge for Indian Languages
Jayadev Billa
Proceedings of Interspeech 2018, 19th Annual Conference of the International Speech Communication Association, Hyderabad, India, 2-6 September 2018

Combining Rule-Based and Statistical Mechanisms for Low-Resource Named Entity Recognition
Ryan Gabbard, Jay DeYoung, **Constantine Lignos, Marjorie Freedman** and **Ralph Weischedel**
Journal of Machine Translation, 2018

The Locus of Linguistic Variation (Book)
Constantine Lignos, Laurel MacKenzie and Meredith Tamminga
John Benjamins Publishing, 2018

When ACE met KBP: End-to-End Evaluation of Knowledge Base Population with Component-level Annotation
Bonan Min, **Marjorie Freedman**, Roger Bock and **Ralph Weischedel**
Proceedings of Eleventh International Conference on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan, 7-12 May, 2018

One Sense Per Document: Improve Name Finding in the Wild with Document-wide Context
Bonan Min, **Marjorie Freedman** and **Ryan Gabbard**
Proceedings of First Workshop on Knowledge Base Construction, Reasoning, and Mining (KBCOM 2018)

What Can Be Accomplished with the State of the Art in Information Extraction? A Personal View
Ralph Weischedel and **Elizabeth Boschee**
Journal of Computational Linguistics, 2018

EMERGING ACTIVITIES GROUP

Using Historical Practices to Develop Safety Standards for Cooperative On-Orbit Rendezvous and Proximity Operations
David Barnhart, Rahul Rughani, Jeremy Allam, Brian Weeden, Fred Slane and Ian Christensen
Proceedings of 69th International Astronautical Congress (IAC), 2018

Efficient Contextualized Representation: Language Model Pruning for Sequence Labeling
Liyuan Liu, **Xiang Ren**, Jingbo Shang, Xiaotao Gu, Jian Peng and Jiawei Han
Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November, 2018

Empower Sequence Labeling with Task-Aware Neural Language Model
Liyuan Liu, Jingbo Shang, **Xiang Ren**, Frank Fangzheng Xu, Huan Gui, Jian Peng and Jiawei Han
Proceedings of Thirty-Second AAAI Conference on Artificial Intelligence, (30th innovative Applications of Artificial Intelligence), and 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, 2-7 February, 2018

End-to-End Reinforcement Learning for Automatic Taxonomy Induction
Yuning Mao, **Xiang Ren**, Jiaming Shen, Xiaotao Gu and Jiawei Han
Proceedings of 56th Annual Meeting of the Association for Computational Linguistics (ACL), Melbourne, Australia, 15-20 July, 2018, Volume 1: Long Papers

Weakly-Supervised Relation Extraction by Pattern-Enhanced Embedding Learning
Meng Qu, **Xiang Ren**, Yu Zhang and Jiawei Han
Proceedings of 2018 World Wide Web Conference, WWW 2018, Lyon, France, 23-27 April, 2018

First Workshop on Knowledge Base Construction, Mining and Reasoning
Xiang Ren, Craig Knoblock, William Wang and Yu Su
Proceedings of Eleventh ACM International Conference on Web Search and Data Mining (WSDM), Marina Del Rey, CA, USA, 5-9 February, 2018

Mining Structures of Factual Knowledge from Text: An Effort-Light Approach
Xiang Ren and Jiawei Han
Synthesis Lectures on Data Mining and Knowledge Discovery, 2018

Scalable Construction and Reasoning of Massive Knowledge Bases
Xiang Ren, **Nanyun Peng** and William Yang Wang
Proceedings of 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorial Abstracts

Learning Named Entity Tagger Using Domain-Specific Dictionary
Jingbo Shang, Liyuan Liu, Xiaotao Gu, **Xiang Ren**, Teng Ren and Jiawei Han
Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November, 2018

Automated Phrase Mining from Massive Text Corpora
Jingbo Shang, Jialu Liu, Meng Jiang, **Xiang Ren**, Clare R Voss and Jiawei Han
Journal of IEEE Trans. Knowl. Data Eng., 2018

HiExpan: Task-Guided Taxonomy Construction by Hierarchical Tree Expansion
Jiaming Shen, Zeqiu Wu, Dongming Lei, Chao Zhang, **Xiang Ren**, Michelle T Vanni, Brian M. Sadler, Jiawei Han, et al.
Proceedings of 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19-23 August, 2018

Indirect Supervision for Relation Extraction Using Question-Answer Pairs
Zeqiu Wu, **Xiang Ren**, Frank F. Xu, Ji Li and Jiawei Han
Proceedings of Eleventh ACM International Conference on Web Search and Data Mining, (WSDM), Marina Del Rey, CA, USA, 5-9 February, 2018

Hierarchical Graph Representation Learning with Differentiable Pooling
Zhitao Ying, Jiaxuan You, Christopher Morris, **Xiang Ren**, William L. Hamilton and Jure Leskovec
Proceedings of Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems, (NeurIPS), Montreal, Canada, 3-8 December 2018

GraphRNN: Generating Realistic Graphs with Deep Auto-regressive Models
Jiaxuan You, Rex Ying, **Xiang Ren**, William L. Hamilton and Jure Leskovec
Proceedings of 35th International Conference on Machine Learning, (ICML), Stockholm, Sweden, 10-15 July, 2018

Open-Schema Event Profiling for Massive News Corpora
Quan Yuan, **Xiang Ren**, Wenqi He, Chao Zhang, Xinhe Geng, Lifu Huang, Heng Ji, Chin-Yew Lin, et al.
Proceedings of 27th ACM International Conference on Information and Knowledge Management (CIKM), Torino, Italy, 22-26 October, 2018

Dynamic Network Embedding by Modeling Triadic Closure Process
Le-kui Zhou, Yang Yang, **Xiang Ren**, Fei Wu and Yueting Zhuang
Proceedings of Thirty-Second AAAI Conference on Artificial Intelligence, the 30th Innovative Applications of Artificial Intelligence, and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence New Orleans, Louisiana, USA, 2-7 February, 2018

Open Information Extraction with Global Structure Constraints
Qi Zhu, **Xiang Ren**, Jingbo Shang, Yu Zhang, Frank F. Xu and Jiawei Han
Proceedings of Companion of the The Web Conference, Lyon, France, 23-27 April, 2018

PROGRAMS UNDER THE KESTON ENDOWMENT DIRECTORSHIP

In 2015 Michael and Linda Keston created their endowed directorship position with a generous donation. This was followed in early 2016 with the appointment of Dr. Prem Natarajan as the inaugural Michael Keston Executive Director of the USC Information Sciences Institute. Our goal is to use the endowment to support groundbreaking research in the areas of information processing, computer and communications technologies. In consonance with these goals, we established the Michael Keston Lecture Series which features renowned speakers from academia, industry and government, the Michael Keston Researcher-in-Residence Program that allows scientists to visit ISI for a period of time to explore new projects, and the Keston Research Grant to encourage innovative technology that will benefit society. In 2018, five research grants were awarded.

WAEL ABD-ALMAGEED

An Artificial Intelligence-Based Mobile Screening Tool: Fetal Programming in Congenital Adrenal Hyperplasia

Artificial intelligence methods will be used to investigate facial morphology in children with fetal programming due to prenatal hormone exposure. Studies will target children with congenital adrenal hyperplasia (CAH) as a natural human model for excess prenatal testosterone. Classical CAH is caused by a 21-hydroxylase deficiency, affecting 1 in 15,000 with fetal hyperandrogenism due to overproduction of adrenal androgens from week 7 of fetal life. This prenatal hormone exposure represents a significant change to the intrauterine environment during early human development that can adversely program the CAH fetus for postnatal disease. A prototype mobile imagining platform will be designed and built that enables the collection of large-scale facial images in children's clinics, without relying on expensive 3D imaging systems. Further, an artificial intelligence-based 2D-to-3D facial processing pipeline will acquire images of the face in healthy controls and CAH youth and compare facial dysmorphism scoring between CAH patients and those that are unaffected.

DAVID BARNHART

Satbotics Control: How to Merge Biologically Inspired Spacecraft Together

This project will provide support to multiple graduate students to develop a new computational architecture that can enable independent satellites or spacecraft to physically and virtually "aggregate" on orbit. This is a completely new methodology and translates from monolithic to cellular in how space systems are created in the future. The computational architecture is intended to allow seamless merging of sensors/actuators/payloads as "resources" that can then be shared autonomously with all other "cells" to enable greater overall performance and capability on orbit than a single large platform can provide. The basics of this new architecture will be demonstrated on an internal 3-DOF air-bearing testbed, using independent floatbots that simulate independent spacecraft.

ALEXEI COLIN

A Betavoltaic-Powered Transmitter for Continuous Glucose Monitors

The Glutex project aims to develop a long-lived, low-maintenance continuous glucose monitor (CGM) for diabetes patients. A CGM is a wearable device that reports blood glucose levels to the patient. Existing CGMs require the patient to recharge batteries every few days and replace the device semiannually. Glutex eliminates this maintenance by replacing the battery with a betavoltaic energy harvester that lasts up to a decade. Glutex pioneers a circuit that accumulates the small trickle of energy from the harvester and releases it in bursts to power the sensor. A successful prototype opens the path to applying betavoltaic power sources in wearable and implantable medical devices.

IVAN SANCHEZ ESQUEDA

FLEX SYNapses for Smart Wearable Electronics and Skin-Attachable Biosensing Devices

Synaptic transistors on flexible and stretchable substrates can enable the implementation of artificial neural networks and learning algorithms when attached to skin sensors for in situ processing and classification of biological signals collected from wearable devices. They can also enable us to mimic the functions of sensory nerves and construct bioelectronic reflex arcs to actuate electro-mechanical devices. This technology has applications for electrophysiology and medical diagnosis, fitness and activity tracking devices, prosthetics, robotics, etc.

MAYANK KEJRIWAL AND PEDRO SZEKELY

Discovery and Dismantling of Human Trafficking Networks

Human trafficking is a form of modern-day slavery with a significant footprint—even here in the United States. Computational tools and methods, including network analysis and machine learning, can help in data-driven mapping of networks of illicit sex providers, many of whom might be victims of trafficking that is attributable to illicit advertisements posted over the Internet. Researchers are currently working to discover and dismantle such networks, especially for possible underage victims. This effort involves a collaboration with both law enforcement and independent consultations with domain experts in the social sciences.

IN MEMORIAM

MICHAEL KESTON | *Michael and Linda Keston Executive Directorship Endowment*

"Michael Keston remained an engineer at heart, a true builder of things filled with a passion for making this world better by solving the hardest problems. We are filled with gratitude for his generosity, and we will all deeply miss Michael."

Prem Natarajan, Michael Keston Executive Director, USC Information Sciences Institute

In February 2019 the University of California's Viterbi School of Engineering and the Information Sciences Institute were touched by the loss of Michael Keston, esteemed philanthropist, entrepreneur, benefactor, advisor, friend to business and academic leaders, and contributor to causes of great significance in our society. One of those causes was the Michael and Linda Keston Executive Directorship Endowment which the Kestons so generously bestowed on Information Sciences Institute in 2015.

Since that time, the Kestons continued to be instrumental in supporting innovative research here at ISI. Their generous endowment is being used to advance research across the institute—from fighting human trafficking to tracking Internet outages in real-time across the world.

Funding provided by the Keston endowment has already supported the inaugural Michael Keston Lecture Series featuring acclaimed speakers from academia, industry and government, and a researcher-in-residence program, allowing visiting scientists to explore new projects with ISI's world-leading researchers. Additionally, several ISI researchers have been able to explore new projects under a Keston Research Grant that encourages the kind of innovative technology that will ultimately benefit society.

BOB BRADEN | *Legendary Internet Pioneer and Emeritus at USC's Information Sciences Institute*

"We have lost a great mind and a kind heart. Bob was greatly admired by anyone who had the privilege to spend time with him. His legacy is immense."

Terry Benzel, Director of ISI's Networking and Cybersecurity Division

In April 2018, the scientific community lost Robert "Bob" Braden, a legendary Internet pioneer and fellow emeritus at USC's Information Sciences Institute.

Braden joined Information Sciences Institute in 1986 following an 18-year career at the University of California, Los Angeles. Prior to those years at UCLA, he taught at Stanford and Carnegie Mellon universities and spent a year with the University College London. There, he developed the first relay system connecting the Internet with the U.K.'s academic X.25 network.

Braden, who retired from ISI's networking research group in 2016 after a 30-year tenure, played a pivotal role in the creation and development of the Internet. During his long career, he contributed to fundamental Internet communications protocols, operating standards and architecture, and co-edited the Request for Comments (RFC) document series, which laid the foundation for the Internet's technical standards.

A fellow of the Association for Computing Machinery, Braden was also active in the broader research community, leading and participating in multiple key task forces. He played a crucial role in the Internet Engineering Task Force which was essential to forging consensus between Internet researchers, operators and industry.

In 2012, Braden was honored with an ACM SIGCOMM Test of Time Award for a highly influential paper, "*Tussle in Cyberspace: Defining Tomorrow's Internet*," (coauthored with ISI's John Wroclawski and MIT's David Clark and Karen Sollins).

Information Sciences Institute

4676 Admiralty Way #1001; Marina del Rey, CA 90292

310.822.1511 | isi.edu | media@isi.edu

USC Viterbi
School of Engineering