DELIVERING THE FUTURE

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

Information Sciences Institute

2017 ANNUAL REPORT

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); Yigal Arens, Prem Natarajan | **Design** Russo Design, Los Angeles | **Printing**: Typecraft, Inc.

USC Viterbi
School of Engineering

Information Sciences Institute

We live in an era of extraordinary developments in the information sciences—the extent and impact of this are as revolutionary as at any time since the original creation of the field. Iconic organizations with deep roots in computer science dominate the business landscape, exerting powerful, disruptive forces that will, without question, have an enduring impact on the future shape of our societies, cultures, and mores.

In particular, the emergence and convergence of ubiquitous computing, mobile devices, computer networking, and artificial intelligence have together enabled technologies that play an increasingly dominant role in our daily professional and private lives. From social media to the Internet of Things, from conversational assistants and smart speakers to cars that can assist and eventually supplant their drivers—these technologies are collectively transforming our hitherto static physical environments into dynamic smart spaces that interact with us in increasingly powerful ways.

Through foundational and enduring contributions to areas such as networking, scientific computing, natural language processing, and AI, researchers at ISI have contributed to the emergence of this new era throughout its 45+ year history. That spirit of contribution remains strong at ISI, and our research portfolio has grown in diversity and impact. Our researchers are driving novel advances in computer vision, language understanding, cybersecurity, quantum computing, high-performance computing architectures, electronics, networking, and other fields. Importantly, they are also engaged in translational work, developing powerful new capabilities such as the domain insight graph (DIG) system which is currently being used with great impact by law enforcement agencies and non-profits to combat the scourge of human trafficking.

These and other accomplishments have led the institute to yet another year of record success in several dimensions. Now in its second year, our new research center in suburban Boston has grown from four researchers in the spring of 2017 to twelve as I write this letter, capturing a broader trend of growth across ISI. During this past year, we witnessed a record number of new researchers joining the ISI family: a total of 38 researchers and other staff across our many different areas of work and our three locations. The size of our PhD student community is at a historical peak and growing. Overall funding during the fiscal year ending June 2017 was $104.5M, crossing the $100M mark for the first time in the institute's history. For the year ending June 2018, our research expenditures are on track to be significantly higher. Importantly, our engagement with various professional communities continues to be strong.

"GROWTH IS NEVER BY MERE CHANCE; IT IS THE RESULT OF FORCES WORKING TOGETHER"

— James Cash Penney (aka JC Penney)

A significant upcoming milestone is the launch of the USC Computing Forum, an initiative designed to bring together scientists, technology leaders, and decision makers in industry and government, and faculty, students, and researchers in academia for an exchange of ideas, information and opportunities. Charter members of the forum include Amazon, Lockheed Martin, Northrop Grumman, California Resources Corporation, and SAP.

In summary, more than any year in recent history, 2017 was defined by growth in nearly every aspect of our institute. This report celebrates the many accomplishments that powered this growth. And, as the discerning reader will no doubt notice, we have a lot to celebrate!

Premkumar Natarajan
Michael Keston Executive Director, USC Information Sciences Institute
Senior Vice Dean of Engineering, Viterbi School of Engineering
Research Professor of Computer Science with Distinction

Information Sciences Institute's main offices in Marina del Rey are located on the coast of Southern California, with a view of the Pacific Ocean and the Santa Monica Mountains (13-story building to the right).

In Waltham, Massachusetts, offices for Information Sciences Institute are located close to major universities and numerous high-tech companies.

In Arlington, Virginia, offices for Information Sciences Institute are located in close proximity to Washington D.C. and Northern Virginia's technology centers.

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 2017 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.

EXECUTIVE OFFICE | DIRECTORS

YIGAL ARENS | Senior Director for Administrative Affairs

In his new position as Senior Director for Administrative Affairs at ISI, Yigal will be responsible for overseeing administrative operations and fostering new levels of excellence in this era of expanding the scale and diversity of operations. He will work closely with division leaders and administrative departments to ensure that they have the best possible work and support environment for a variety of functions—including recruiting, project management, administrative support, and staff development. Yigal has been with ISI for over 30 years—15 of those years as director of the Artificial Intelligence Division.

JOHN DAMOULAKIS | Senior Director for Special Projects

After more than 15 years with ISI as a division director, John has been appointed ISI's Senior Director for Special Projects, with a focus on developing new, large-scale research opportunities and lines of sponsorship in the areas of electronics and hardware device security. These areas represent an increasingly important direction for the Institute—due in no small part to John's previous efforts. Looking forward, he will work closely with researchers across the Institute and collaborate with the MOSIS Division to grow our presence in the areas of design and fabrication of advanced electronics.

JOHN WROCLAWSKI | Senior Director for Strategic Initiatives

As Senior Director for Strategic Initiatives, John will be working with leaders and researchers across the institute to create a diverse research portfolio and develop new, boundary-spanning initiatives to advance the Institute's strategic vision. Prior to this appointment, he served for 12 years as the director of ISI's Networking & Cybersecurity Division, leading pioneering research in such areas as Internet protocols and architecture, the security of complex networked systems, sensor nets, experimental networking research methodologies, space systems networking, and the policy and economic implications of Internet technical structure and real-world use. John has been with ISI for 15 years.

RESEARCH | DIRECTORS

ELIZABETH BOSCHEE | Director of ISI Boston

Elizabeth joined ISI as a research lead in 2016 and is now the director of ISI's fast-growing Boston office. Prior to joining ISI, Elizabeth was a lead scientist at BBN Technologies where she headed up numerous language understanding efforts with DARPA, IARPA, ONR, and the IC—most recently developing systems that read the news to automatically compile knowledge bases on socio-political events. Current projects include automatically extracting and organizing causal knowledge from semantically diverse information sources, and developing a system for generating English summaries of multilingual content from both speech and text.

TERRY BENZEL | Director of Networking & Cybersecurity Division

Terry has been with ISI for nearly 15 years and served as deputy director of the NCD Division for many of those years. Under her new appointment as division director, the division will continue its role in creating successful, insightful, and challenging research results. Terry brings national visibility to our Institute through her longstanding leadership in the cybersecurity area—especially in experimental methods. She has a unique combination of deep research management expertise, excellent relationships with sponsors, proven success as a project PI, and familiarity with the division, the Institute and the university.

CRAIG KNOBLOCK | Director of Artificial Intelligence Division

In the fall of 2017 Craig was named director of the Artificial Intelligence Division—one of the largest and most respected AI laboratories in the U.S. With over 25 years of experience, he is one of ISI's most respected and successful senior researchers, with a unique record of impactful project definition, funding success, professional community engagement, and strong mentoring support. As director of the IS Division, he will lead a team of over 30 researchers working on leading-edge projects—including biomedical data integration, machine translation, robotics and social networks. Craig is an AAAI Fellow, an ACM Fellow and Distinguished Scientist, and a Senior Member of IEEE.

2017 was an extraordinary year for the Information Sciences Institute in many ways: Over $100 million in new grants and collaborations from key federal agencies and from industry and academia; ISI sponsorship of the new USC Computing Forum; three researchers elected Fellow of the Association for Computing Machinery (ACM); and thirty-seven new highly qualified directors, computer scientists, physicists, engineers, researchers and staff at ISI's facilities in Marina del Rey CA, Arlington VA, and Waltham MA—and even satellite offices located at the Air Force's Space and Missile Systems Center. We are pleased to announce the following new members of the ISI team.

ARLINGTON, VIRGINIA

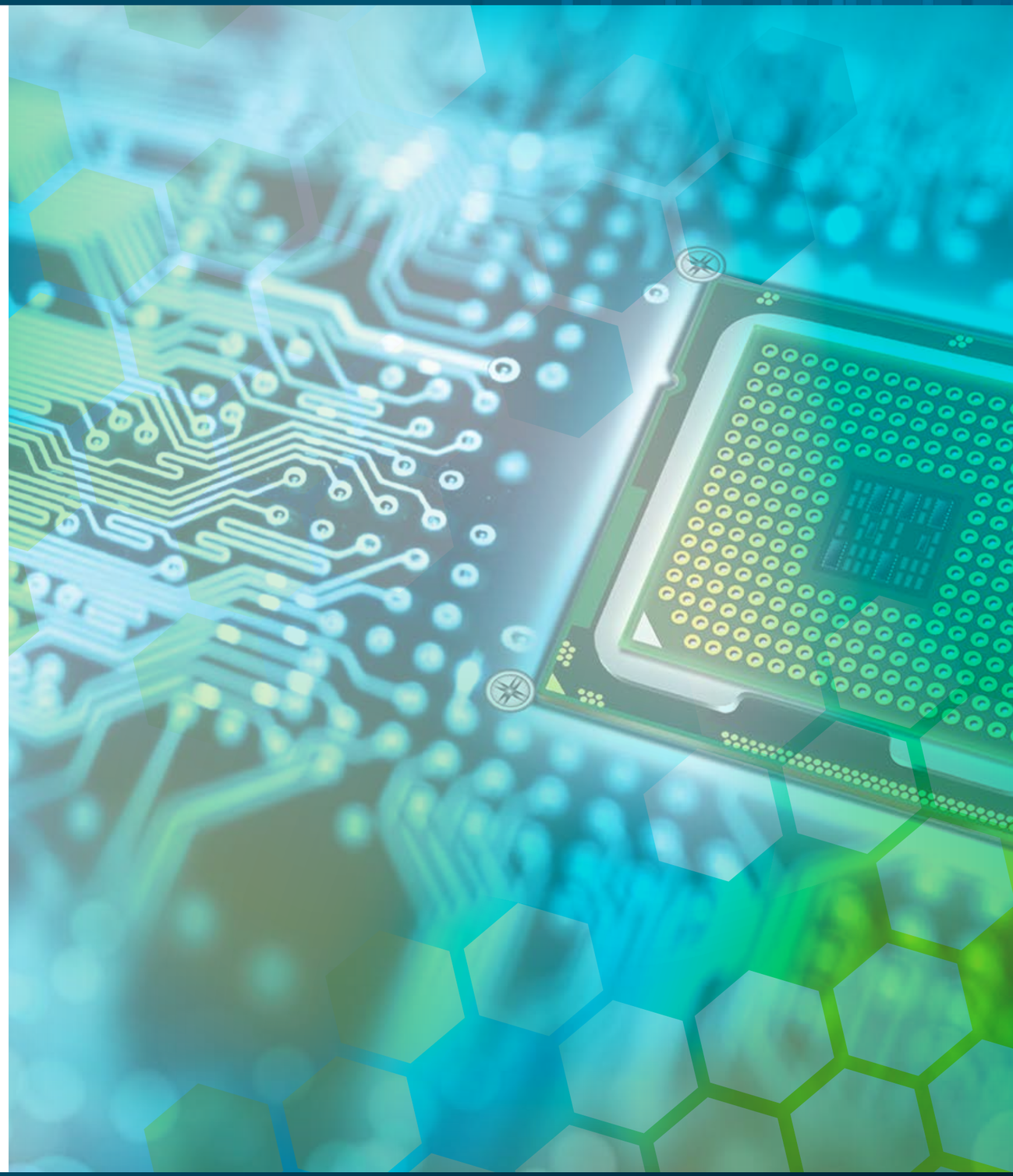
TEAM MEMBER	TITLE	TEAM MEMBER	TITLE
Travis Haroldsen	Computer Scientist	Vivek Venugopalan	Computer Scientist
Andrew Rittenback	Computer Scientist	Ralf Wiegand	Systems Administrator
Samuel Skalicky	Computer Scientist		

BOSTON, MASSACHUSETTS

TEAM MEMBER	TITLE	TEAM MEMBER	TITLE
Marjorie Freedman	Research Lead	Jonathan Habif	Scientist/Experimental Physicist
Ryan Gabbard	Computer Scientist	Constantine Lignos	Computer Scientist
Richard Guilmain	Research Programmer	Ralph Weischedel	Research Team Leader

MARINA DEL REY, CALIFORNIA

TEAM MEMBER	TITLE	TEAM MEMBER	TITLE
Andrés Abeliuk	Computer Scientist	Frederick Morstatter	Computer Scientist
Jeffrey Asher	Computer Systems Engineer	Jay Pujara	Computer Scientist
Rafi Berberian	Systems Admin IV	Srivatsan Ravi	Computer Scientist
Jayadev Billa	Senior Computer Scientist	Marc Romero	Computer Services Consultant
Sarath Burson	Production Coordinator	Edward Russell	Director, Strategic Space Programs
Douglas Cool	Deputy Director, Strategic Space Programs	Anna Sapienza	Postdoctoral Scholar
Caitlin Dawson	Writer/Editor	Leonidas Spinoulas	Computer Scientist
Brian Duffy	Senior Systems Engineer	Robert Story	Research Programmer
Michael Elkins	Senior Computer Scientist	Viviana Teofilo	Project Specialist
Shobeir Fakhrael	Computer Scientist	Satish Thittamaranahalli	Computer Scientist
Malcolm Frazier	Senior DevOps Engineer	Gary Thorne	VLSI Design Engineer
Homa Hosseinmardi	Postdoctoral Scholar	Jane Tungka	Senior Import/Export Trade Analyst
Joséph Mathai	Research Programmer	Lori Weiss	Project Manager
Joel Mathew	Research Programmer	Yixiang Yao	Research Programmer



HONORS

YIGAL ARENS
Recipient of 2017 Distinguished Service Award from the Digital Government Society for his long and "sustained service and tireless dedication" to digital government.

YONATAN BISK
Named to ACM's "Future of Computing Academy" which aims to foster interdisciplinary, next-generation computer scientists to tackle challenges in computing and society at large.

YOUNG CHO
Recipient of 2017 Technology Advancement Grant from the USC Stevens Center for Innovation for his research project "High Accuracy Multicore GPU Power/Temperature Map Generation Using Numerical Software Analysis."

YOLANDA GIL
Elected Fellow of the Association for Computing Machinery (ACM) for "leadership in advancing the use of artificial intelligence in support of science and for service to the community."

CARL KESSELMAN
Elected Fellow of the Association for Computing Machinery (ACM) for "contributions to high-performance computing, distributed systems, and scientific data management."

CRAIG KNOBLOCK
Elected Fellow of the Association for Computing Machinery (ACM) for "contributions to artificial intelligence, semantic Web, and semantic data integration."

FRED MORSTATTER
Among 200 top students and early-career researchers selected to participate in the 2017 Heidelberg Laureate Forum in Heidelberg, Germany.

AWARDS

DANIEL GARIJO
Best Paper Award | "WIDOCO: A Wizard for Documenting Ontologies,"*International Semantic Web Conference (ISWC) Resources Track*.

MARJAN GHAZVININEJAD, XING SHI, JAY PRIYADARSHI, AND **KEVIN KNIGHT**
Best Demo Award | "Hafez: An Interactive Poetry Generation System," *ACL Demo Track*.

MAYANK KERJRIWAL
Distinguished Dissertation Award | 2017 Semantic Web Science Association for "Populating a Linked Data Entity Name System."

JOHANNES H. UHL, STEFAN LEYK, YAO-YI CHIANG, WEIWEI DUAN, AND **CRAIG KNOBLOCK**
IAPR Best Paper Prize | "Extracting Human Settlement Footprint from Historical Topographic Map Series Using Context-Based Machine Learning," *8th International Conference on Pattern Recognition System*.

ANDREW SCHMIDT, GABIEL WEISZ, AND **MATTHEW FRENCH**
Best Short Paper Award | "Evaluating Rapid Application Development with Python for Heterogeneous Processor-based FPGAs," *Proceedings of the 2017 International Symposium on Field Programmable and Custom Computing Machines*.

RICARDO DE OLIVERA SCHMIDT, JOHN HEIDEMANN, AND JAN HARM KUIPERS
Best Paper Award | "Anycast Latency: How Many Sites Are Enough?" *Passive and Active Measurement Conference (PAM)*.

PATENTS

DANIEL LIDAR, **TAMEEM ALBASH**, AND WALTER VINCI
"Nested Quantum Annealing Correction"
No. US-2017-0364362-A1

JOSÉ LUIS AMBITE
Program Committees | International Semantic Web Conference (ISWC); 12th Conference on Data Integration in the Life Sciences (DILS); First Joint Workshop on Health Intelligence/ Fourth Workshop on Expanding the Boundaries of Health Informatics Using AI.
Reviewer | American Medical Informatics Association Symposium; Knowledge and Information Systems; Transactions on Knowledge and Data Engineering.

DAVID BARNHART
Pioneered New Graduate Course | a "hands-on" course for graduate students to learn the operations and practices of real satellite ground-based communications.

GULLY BURNS
Ad Hoc Reviewer | for the NIH BILDS Study Section; active reviewer in the Frontiers System. Programming Committee for Semantic Science Workshop and International Workshop on Capturing Scientific Knowledge.

EWA DEELMAN
Associate Editor | *IEEE Transactions on Parallel and Distributed Systems*.
Editorial Board Member | *International Journal of High Performance Computing Application; ACM International Conference Proceeding Series; Journal of Grid Computing steering committee;* IEEE eScience Conference; IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing.

MATTHEW FRENCH
Technical Program Committee Member | IEEE FCCM 2017 Technical Program Committee.

EMILIO FERRARA
Editorial Boards | *Future Internet and Data Science*.
Founding Editor | *Future Internet: Techno-Social Smart Systems*.
Organizer | *ICDM Workshop: Data Science for Human Performance in Social Networks; CCS workshop on Computational Social Science: Social Contagion, Collective Behavior, and Networks*.

YOLANDA GIL
President-Elect | Association for the Advancement of Artificial Intelligence (AAAI).
Past Chair | Association for Computing Machinery (ACM). Special Interest Group on Artificial Intelligence (SIGAI).
Vice-President | Semantic Web Science Association (SWSA).
Lead Judge | International Science and Engineering Fair (ISEF).
Founding Editorial Board Member | *Human Computation Journal; Journal Data Science: Methods, Infrastructure, & Applications; AI Matters (ACM SIGAI Quarterly Newsletter)*.
Advisory Board Member | *MethodsX Journal Computer Science Section*.
Editorial Boards | *Artificial Intelligence Journal; Journal of Web Semantics; Applied Ontology Journal*.

ITAY HEN
Editoral Member | *Scientific Reports*.
Review Editor | *Frontiers in ICT*.

CRAIG KNOBLOCK
Program Committees | IAAAI, Research Track, International Semantic Web Conference; In-Use Track, International Semantic Web Conference; Extended Semantic Web Conference; WWW Special Track on Semantics and Knowledge; 1st International Workshop on Industrial Knowledge Graphs; 1st Workshop on Enabling Open Semantic Science.
Associate Editor | *Spatial Algorithms and Systems*.
Advisory Board | *ACM Transactions on ARTIFICIAL INTELLIGENCE and Technology*.
Publications Committee | *AAAI Press*.
Finance Committee | IJCAI Inc. Senior Program Committee, AAAI.

KRISTINA LERMAN
Board member | GESIS.
Editorial Board | *EPJ Data Science, Journal of Computational Social Science, Nature Scientific Reports*. World Economic Forum panel on AI and Virtual Reality.

JONATHAN MAY
Co-Organizer | International Workshop on Semantic Evaluation (SemEval).

JOSH MONSON
Technical Committee Member | International Conference on Field Programmable Logic and Applications, and IEEE International Symposium on Highly-Efficient Accelerators and Reconfigurable Technology.

PREM NATARAJAN
Technical Committee Member | International Conference on Document Analysis and Recognition.
Co-chair | ICDAR Workshop on Multilingual OCR.

ANDREW SCHMIDT
Track Co-Chair | ReConFig: High Performance Reconfigurable Computing IEEE FCCM 2017 Technical Program Committee.
Reviewer | Microprocessors and Microsystems Journal.
Co-Chair | Workshop on Enabling Open Semantic Science at International Semantic Web Conference
Co-Chair, Treasurer, Sponsor Chair | Workshop on Capturing Scientific Knowledge at ACM International Conference on Knowledge.

SAMUEL SKALICKY
Technical Program Committee | Journal of Parallel and Distributed Computing, and International Conference on Reconfigurable Computing and FPGAs.

VIVEK VENUGOPALAN
Technical Program Committee | IEEE Field-Programmable Custom Computing Machines (FCCM); IEEE Highly Efficient Accelerators and Reconfigurable Technologies (HEART); IEEE ReConFigurable Computing; IEEE High Performance Computing & Simulation (HPCS); IEEE Conference on Communications and Network Security (CNS); International Workshop on Cyber-Physical Systems Security (CPS-Sec); IEEE International Midwest Symposium on Circuits and Systems (MWSCAS); and Springer *Journal of Hardware and Systems Security*.

POSTDOCTORAL SCHOLARS

Daniel Garijo	Provenance, scientific workflows, linked data, semantic web, artificial intelligence
Deborah Khider	Ontologies, scientific metadata, crowdsourcing, climate dynamics
Anna Sapienza	Computational social science, machine learning, network analysis

PH.D. GRADUATES

PH.D. GRADUATES	ADVISOR	RESEARCH AREA	CURRENTLY AT
Hao Shi	Jelena Mirkovic	Malware analysis	SW engineer at Facebook
Simon Woo	Jelena Mirkovic	Passwords	Assistant Prof, SUNY Korea
Hao Wu	Kristina Lerman	Neural models of social networks	SW engineer at Apple

DOCTORAL STUDENT

DOCTORAL STUDENT	RESEARCH AREA	ADVISOR
Nazanin Alipoufard	Artificial Iintelligence	Kristina Lerman
Nada Aldaraab	Decipherment, machine translation	Kevin Knight
Shusan Arakelyan	Machine learning/data science	Aram Galstyan
Calvin Ardi	Networking & network measurement	John Heidemann
Victor Ardulov	AI/Human-centered machine learning	Shrikanth Narayanan
Adam Badawy	Computational social science	Emilio Ferrara
Guillermo Baltra Elorriaga	Networking & network measurement	John Heidemann
Nathan Bartley	Artificial intelligence	Kristina Lerman
Karel Batic	Signal processing/human-centered machine learning	Shrikanth Narayanan
William Bezouski	Spacecraft localization using computer vision	David Barnhart
Brandon Booth	AI/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
Ashok Deb	Cybersecurity	Emilio Ferrara
Xiyue Deng	Malware analysis	Jelena Mirkovic
Aditya Deshpande	Computer architecture & memory hierarchy	Jeffrey Draper
Tu Mai Anh Do	Workflow systems in HPC environments	Ewa Deelman
Shuyang Gao	Machine learning/information theory	Aram Galstyan
Sahil Garg	Machine learning/information extraction	Aram Galstyan
Majid Ghasemi-Gol	Artificial intelligence	Pedro Szekely
Sarik Ghaszarian	Machine learning/NLP	Aram Galstyan
Marjan Ghazvininejad	Generation	Kevin Knight
Palash Goyal	Machine learning	Emilio Ferrara
Hang Guo	Networking & network measurement	John Heidemann
Lalit Gupta	Quantum computing	Itay Hen
Che-Wei Huang	Affective computing	Shrikanth Narayanan
Basileal Imana	Networking & network measurement	J. Heidemann/A. Korolova
Zoe Gonzalez Izquierdo	Quantum computing	Itay Hen
Ayush Jaiswal	Adversarial learning/multimedia integrity analysis	Prem Natarajan
Wuxuan Jiang	Machine learning/differential privacy	Aram Galstyan
David Kale	Artificial intelligence	Greg Ver Steeg
Hsien-Te Kao	Deep Learning	Emilio Ferrara

Hannes Leipold	Quantum computing	Federico Spedalieri
Wenzhe Li	Machine learning/graph analytics	Aram Galstyan
Zekun Li	Face recognition	Prem Natarajan
Kuan Liu	Scalable learning algorithms for recommender systems	Prem Natarajan
Nikolaos Malandrakis	Natural language processing	Shrikanth Narayanan
Victor Martinez Palacios	Natural language processing	Shrikanth Narayanan
Akira Matsui	Computational behavioral economics and data science	Emilio Ferrara
Mernoosh Mirtaheri	Machine learning/data science	Aram Galstyan
Amrutha Nadarajan	Signal processing/audio/biosignals	Shrikanth Narayanan
Aqib Nisar	Networking	J. Heidemann/E. Katz-Bassett
George Papadimitriou	Execution & anomaly detection in wide area workflows	Ewa Deelman
Pavlos Papadopoulos	Signal processing/audio/speech	Shrikanth Narayanan
Raghuveer Peri	Signal processing/audio/speech	Shrikanth Narayanan
Minh Tran Xuan Pham	Artificial Intelligence	Craig Knoblock
Nina Pourdamghani	Machine translation	Kevin Knight
Abdul Qadeer	Networking & network measurement	John Heidemann
Sivaramakrishnan Ramanathan	Network monitoring and security	Jelena Mirkovic
Stephen Rawls	Deep learning, handwriting recognition, and OCR	Prem Natarajan
Kyle Reing	Machine learning/information theory	Aram Galstyan
A.S.M. Rizvi	Networking & network measurement	John Heidemann
Rahul Rughani	Rendezvous & proximity operations for small satellites	David Barnhart
Ekrram Sabir	Multimedia integrity analysis	Prem Natarajan
Praveen Sharma	Computer architecture and VLSI	Jeffrey Draper
Ruhollah Shemirani	Large-scale genetic analysis, identify by descent	José Luis Ambite
Emily Sheng	NLP for scientific literature analysis	Prem Natarajan
Xing Shi	Machine translation	Jonathan May
Yuan Shi	Artificial intelligence	Craig Knoblock
Jason Slepicka	Artificial intelligence	Pedro Szekely
Serban Stan	Machine learning/graphical models	Aram Galstyan
Dimitri Stripelis	Large-scale data integration	José Luis Ambite
Rajat Tandon	Denial-of-service defenses	Jelena Mirkovic
Nazgol Tavabi	Artificial intelligence	Kristina Lerman
Geoffrey Tran	Fault tolerance in cloud-based analytics	Stephen Crago
Ruchir Travadi	Signal processing/audio/speech	Shrikanth Narayanan
Colin Vaz	Signal processing/audio/speech	Shrikanth Narayanan
Xin-Zeng Wu	Artificial intelligence	Kristina Lerman
Lan Wei	Networking & networking measurement	John Heidemann
Hong Xu	Artificial intelligence	Satish Thittamaranahal
Shen Yan	Machine learning	Emilio Ferrara
Xusen Yin	Decipherment	Kevin Knight
Yilei Zeng	Data Science	Emilio Ferrara
Xiaofan Zhang	Quantum Computing	Itay Hen
Liang Zhu	Networking & networking measurement	Calvin Ardi

Viterbi-India Summer Research Program

STUDENT	ADVISOR
Soumik Ghosh	Itay Hen
Karen Quadros	Prem Natarajan
Bhavya Rachchh	Prem Natarajan

STUDENT	ADVISOR
Jeremy Allam	David Barnhart
Supriya Anand	José Luis Ambite
Chrysovalantis Anastasiou	José Luis Ambite
Aggarwall Ankita	Prem Natarajan
Kamya Batra	Prem Natarajan
Yining Chen	Kevin Knight & Jonathan May
Jiaxin Cheng	Prem Natarajan
Leon Cheung	Kevin Knight
Alyssa Deng	Yolanda Gil
Jiayuan Ding	Pedro Szekely
Ryan Espiritu	Yolanda Gil
Nicole Fronda	Anoop Kumar
Bhavya Gulati	Prem Natarajan
Rahul Gupta	Craig Knoblock
Houman Hajimohamaddi	Alefiya Hussain
Ameya Hanamsagar	Jelena Mirkovic
Renu Hiremath	Prem Natarajan
Pegah Jandaghi	Jay Pujara
Amy Jang	Yolanda Gil
Jie Ji	Daniel Garijo
Rahul Kapoor	Pedro Szekely
Fiona Khatana	Craig Knoblock
Margaret Knoblock	Yolanda Gil
Marcel Lariviere	David Barnhart
Nelson Liu	Kevin Knight & Jonathan May
Fanghao Luo	Pedro Szekely
Sachin Malhotra	Prem Natarajan
Pulkt Manocha	Mayank Kejriwal
Sasha Mayn	Kevin Knight & Jonathan May
Abhishek Megotia	Craig Knoblock
Usama Mehboob	Jelena Mirkovic
Tirth Mehta	Yolanda Gil
Saurabh Mishra	Yolanda Gil

STUDENT	ADVISOR
Daye Nam	Mayank Kejriwal
Abhinav Palia	Jelena Mirkovic
Niharika Patel	David Barnhart
Piyush Patel	David Barnhart
John Pelgrift	David Barnhart
Jing Peng	Mayank Kejriwal
Avanti Prasanna	Prem Natarajan
Jaydeep Ramani	Jelena Mirkovic
Timothy Rhodes	Anoop Kumar
Jocelyn Rodriguez	John Heidemann
Rebecca Rogers	David Barnhart
Sameerdas Savitha	Prem Natarajan
Kaushalkumar Shah	Craig Knoblock
Runqi Shao	Pedro Szekely
Ankit Sharma	David Barnhart
Sukriti Sharma	Prem Natarajan
Henry Sheehy	Itay Hen
Xiangyang Shi	Prem Natarajan
Aakanchha Sinha	José Luis Ambite
Jarrett Smalley	Federico Spedalieri
Onzali Suba	Craig Knoblock
Aditya Sundaram	Mayank Kejriwal
Pallavi Taneja	Craig Knoblock
Sree Priyanka Uppu	Emilio Ferrara
Shreya Venkatesh	Craig Knoblock
Ashwin Venkatesha	Ewa Deelman
Prince Wang	Yolanda Gil
Regina Wang	Yolanda Gil
Hong Xu	Satish Thittamaranahalli
Jingyun Yang	Daniel Garijo
Davut Yavuz	Jelena Mirkovic
Haotian Zhan	Mayank Kejriwal
Zexia Zhang	Daniel Garijo

SCHOLARS AND RESEARCHERS	RESEARCH AREA	ISI HOSTS
Lucas Augusto Carvalho (University of Campinas, Brazil)	*Workflow reuse*	Alefiya Hussain
Aranya Chakrabortty (NCSU)	*Energy CPS*	Alefiya Hussain
Ken Cheong (Hong Kong Baptist University)	*Workflow graph mining*	Alefiya Hussain
Jeff Contri (Penn State)	*Mechanisms*	David Barnhart
Mason Copp (Cal Poly SLO)	*Space engineering*	David Barnhart
Seyed Mehran Dibaji (MIT)	*Dynamics and Controls, Optics*	Alefiya Hussain
Sindhu Ernala (Georgia Tech University)	*Artificial intelligence*	Emilio Ferrara
Sneha Kudugunta (Viterbi-India Program)	*Artificial intelligence*	Emilio Ferrara
Elahe Momeni (Univ. of Vienna)	*Social media modeling*	Kristina Lerman
Emerson Montana (Rice University)	*Space engineering*	David Barnhart
Neeti Narayan (State University of New York)	*Image integrity analysis*	Prem Natarajan
Katayoun Neshatpour (George Mason University)	*Machine learning-based CAD algorithms*	Joshua Monson
Hao Peng (University of Michigan)	*Artificial intelligence*	Emilio Ferrara
Allon Percus (Claremont Graduate Univ)	*Networks*	Kristina Lerman
David Pierce (Penn State)	*Software*	David Barnhart
Iacopo Pozzana (Birbeck Univ. of London, UK)	*Artificial intelligence*	Emilio Ferrara
Abhi Devalapura Rajagopala (Univ. No. Carolina)	*High-performance computing & next-generation memory architecture*	Andrew Schmidt
Miguel Rodriguez (University of Florida)	*Robust retrieval from scientific knowledge bases*	Prem Natarajan
Mohsen Sayyadiharikandeh (Indiana University)	*Artificial intelligence*	José Luis Ambite
Alireza Shabani (USC)	*Quantum computing*	Federico Spedalieri
Pratishtha Sukla (NCSU)	*Networked control systems*	Alefiya Hussain
Eitan Slavick (Santa Monica College)	*Liquid Mirror Development*	David Barnhart
Ana Solis (Santa Monica College)	*Inertial Navigation Systems*	David Barnhart
Ashish Tondwalkar (Thomas Jefferson HS for Science & Technology)	*Object localization through sound using FPGAs & high-level languages*	Andrew Schmidt
Anssi Yli-Jirä (University of Helsinki)	*Non-projective parsing*	Kevin Knight

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:

BATL
Biometric Authentication with a Timeless Learner

Under IARPA's Odin/Thor program, ISI researchers are creating systems and algorithms resilient to presentation attacks (spoofing) of biometric authentication systems—including face, iris and fingerprint systems.

DiSPARITY
Digital, Semantic and Physical Analysis of Media Integrity

Sponsored by DARPA's MediFor program, the ISI research team is focused on identifying and characterizing signs of manipulated images, video and metadata.

ELICIT
A System for Extracting and Organizing Causal Information

Funded by DARPA, ELICIT is developing a framework that integrates concepts of causality, factual knowledge, and meta-reasoning into a model-driven knowledge graph representation that allows decision makers to access relevant knowledge.

ELISA
Exploiting Language Information for Situational Awareness

Under DARPA's LORELEI program, we are developing technology for rapidly constructing information extraction, machine translation, and topic and sentiment processing capabilities for new languages.

GAIA
Generating Alternatives for Interpretation and Analysis

As part of the DARPA AIDA program, the GAIA research team is developing techniques to extract and synthesize knowledge from disparate information sources (multi-lingual texts, including texts embedded in images/videos, speech, image, video, meta-data) to generate hypotheses/interpretations for analyst consumption.

GLAIVE
Graphics-Based Learning Approach Integrated with Vision Elements

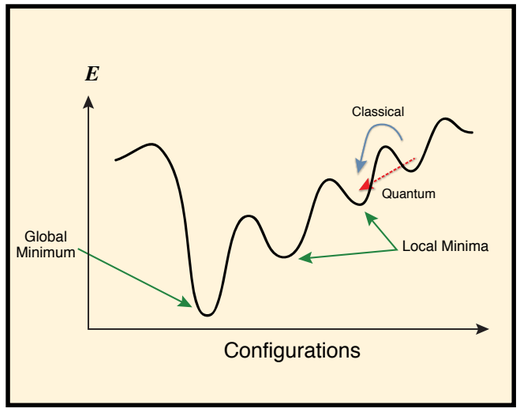
The GLAIVE team is funded by IARPA's Janus program to develop face recognition technology that helps computers recognize individuals in the wild, i.e., in photos and videos captured in uncontrolled settings.

SARAL
Summarization and Domain-Adaptive Retrieval Across Languages

Sponsored by IARPA, the SARAL team is developing a lgorithms and technologies that address a critical need for automated solutions that can identify information of interest across a multiplicity of domains, languages, and scenarios.

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 quantum computing 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 research on three different generations of the processor. Currently, there are only two other operational systems in the US—one installed at NASA Ames Research Center and operated jointly by NASA and Google, and another at Los Alamos National Laboratory. Operating quantum computing systems is demanding: the systems need to be kept at near absolute temperature and electromagnetically shielded to protect the fragile quantum states from degradation by external noise.



Quantum Annealing

THE POSTEL CENTER FOR EXPERIMENTAL NETWORKING

The *Postel Center for Experimental Networking* is an endowed facility focusing on network research and service to the Internet community. In its early years, the Center focused on Internet history and supporting network research tools and resources through visiting scholars. Recently, we explored Internet policy matters and issues underlying our understanding of the Internet architecture and its protocols. We developed reference documents for the Internet community to clarify architectural concepts (tunnels, middleboxes, GRE fragmentation, transport port use) and to extend legacy protocols for new uses (both TCP and UDP option space extensions).

The Postel Center is named in memory of **Dr. Jon Postel**—a brilliant and dedicated scientist who made many key contributions to the formative days of the ARPANET, including protocol design and verification, multimedia computing and communications, electronic commerce, the domain name system, and many other specific Internet protocols. Jon was widely known for the influence he exerted on the management of the Internet, recognizing early on that packet-switching research would need organization and a modicum of discipline if it were to realize its full potential as a universal communication medium. Jon created the activities that eventually grew into the RFC Editor, which issues and controls the many documents that specify how Internet computers interoperate, and also initiated (circa 1981) the Internet Assigned Numbers Authority (IANA), the central coordination function for the global Internet.

Currently we are exploring new missions and future directions, and moving away from support for shared resources that are increasingly available via both public and commercial enterprises. This past year we refocused the Center to expand on ISI's rich history in developing and managing Internet resources. This includes developing materials to educate the Internet community on "best practices" for network conservation, e.g., in the documentation of protocols and management of shared resources, such as number and name spaces. Additionally, we are establishing new collaborations that will help us explore and document ISI's role in Internet history.

ISI CENTER FOR COMPUTER SYSTEMS SECURITY | DIRECTOR CLIFFORD NEUMAN

The ISI *Center for Computer Systems Security* conducts research and provides education in the crucial disciplines of computer, network, and application security. Among the current research activities, Center staff are developing new architectures for isolation in networked systems, and they are studying resilience to cyber-attacks in the critical infrastructure systems of the power grid and oil and gas extraction.

The work on critical infrastructure demonstrates how cyber-attacks affect the operational resilience of the infrastructure, impeding the delivery of power to consumers or oil and gas to refineries. It also identifies remediation strategies among candidate actions. Center researchers are exploring cross-infrastructural dependencies to identify how attacks in one area impact the resilience of other infrastructures.

In addition to its research activities, Center researchers are frequently called upon by the media to explain events involving privacy, cyber security, and cyber crime. In the past year, Center Director Clifford Neuman was quoted in more than 50 publications and appeared more than 15 times in television and national network news segments and radio programs, where topics ranged from security of the Internet of Things, WikiLeaks' disclosure of hacking techniques, breaches of the Dallas emergency warning notification system, numerous ransomweare incidents, security of our election systems, and hacks that occurred during the 2012 campaigns.

ISI's Center for Computer Systems Security was targeted in 2016 as a DHS- and NSA-designated Center of Academic Excellence in information assurance research. Center staff lead several education programs in computer security and have redesigned the master's of science degree in security engineering to apply the existing focus on high assurance systems to the Internet and cloud-focused computing environments now in demand by consumers and security. A new cross-disciplinary class was also introduced that focuses on privacy and the practical legal and international considerations of computer security in today's and future systems.

SPACE ENGINEERING RESEARCH CENTER (SERC) | DIRECTOR DAVID BARNHART

As a joint effort of the USC Viterbi School of Engineering and its Information Sciences Institute (ISI), the Space Engineering Research Center (or SERC) combines the depth of USC academic research in the Department of Astronautical Engineering with the pragmatism of hands-on space development.

Labs and physical spaces are located both on the USC main campus in Los Angeles and in nearby Marina del Rey, where ISI is based. Leadership includes faculty with diverse academic and industry experience from both institutions. SERC collaborates with government labs, industry and other USC schools and departments. Funding is provided by federal agency grants, industry partners and other sources.

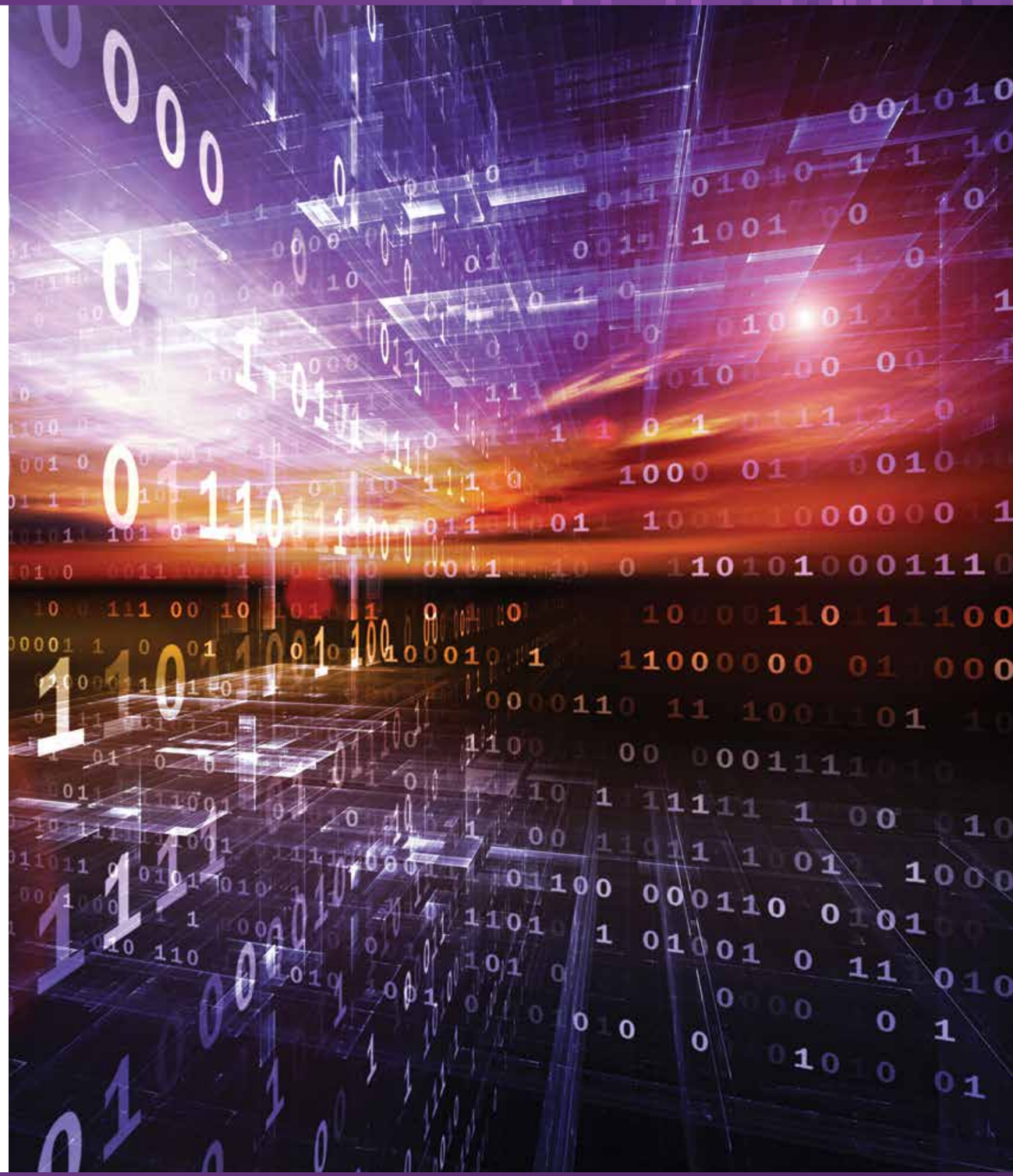
SERC combines disruptive space R&D, hands-on education of the second-generation workforce, and industry collaboration that delivers concrete benefits to our students and corporate partners.

The only "engineering teaching hospital" of its kind within the context of a diverse research university, SERC teaches students to build, test and fly actual spacecraft and satellites. We also prepare them for a radically different, democratized environment in which a burgeoning number of nations and private enterprises will compete to reduce costs, enable novel capabilities and re-envision what space means to Earth.

This year SERC has been working on new innovations in the field of rendezvous and proximity operations related to the emerging missions of space servicing. The Center has been awarded several contracts in this domain—from a docking adaptor system, multi-sensor fusion project, to a major funded research project from ATI and DARPA to establish commercial standards for RPO and spacecraft servicing. The USC/ISI SERC center is part of a unique commercial consortium that will have a global impact on space operations now and into the future. The consortium is called *CONFERS*, or Consortium for Execution of Rendezvous and Servicing Operations.

2017 saw the design and construction of our new optical laboratory focused on the experimental investigation of quantum information. The *Laboratory for Quantum-Limited Information* is focused on demonstrations yielding the fundamental limits of our ability to store and extract information from quantum physical systems—such as single particles of light, atoms and molecules. The laboratory is unique in that it conducts experiments over very broad ranges of optical wavelengths from visible, to near, mid-wave and long-wave infrared. Lab researchers will investigate the generation and measurement of classical (such as laser) and non-classical (such as quantum entangled) light across this spectrum of wavelengths, with the goals of developing a deep understanding of the quantum mechanical properties of this light and enabling revolutionary communications, sensing and computation capabilities.

ISIers Jonathan Habif and Prem Natarajan inspect the new Laboratory for Quantum-Limited Information



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: Scalable and Secure Software-Defined Controllers

With traditional networking infrastructures, network operators are forced to continuously (re)configure policies in order to respond to a wide range of network events and applications. Thus the idea of programmable networks that allow dynamic programming of network devices via an open interface has gathered a lot of attention over the past few years. Software-defined networking (SDN) out-sources the control over the network to logically centralized software, called the *control plane*. The ability of the control plane to "program" the network (the data plane) presents new interesting opportunities to network operators and system designers. The advent of SDN has opened up the ability to dynamically deploy 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.

While the perspective of a centralized control plane simplifies network management, it comes with the usual drawbacks of a centralized approach: a single point of failure and attack, scalability bottleneck, etc. A fully centralized system can simply not adequately provide the required levels of availability, responsiveness and scalability—thus the need for a distributed control plane. Moreover, as SDNs have started to penetrate the wider Internet, an adversary can compromise some of the controllers and make them behave maliciously. This research concerns the design and implementation of a scalable and secure distributed controller platform that 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.

COMPUTATIONAL SYSTEMS AND TECHNOLOGY | DIRECTOR STEPHEN CRAGO

A world leader in computing technologies—from basic to applied research—the CS&T Division conducts research on quantum computing, advanced medical devices, onboard processing for space missions, radiation mitigation, smart grids, cloud computing, and integrated circuits.

ARTIFICIAL INTELLIGENCE | DIRECTOR CRAIG KNOBLOCK

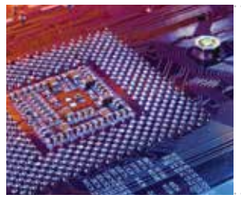
One of the world's largest artificial intelligence groups, the AI Division focuses on natural language processing, knowledge technologies, information integration, robotics, machine intelligence, biomedical knowledge engineering and data integration, and multimedia analysis.

NETWORKING AND CYBERSECURITY | DIRECTOR TERRY BENZEL

The NCD Division focuses on Internet measurements, cyber-physical systems like the smart grid, cybersecurity challenges facing the Internet, methodologies and tools for large-scale cyber and cyber-physical networked systems, clean-break initiatives, and the Domain Name System.

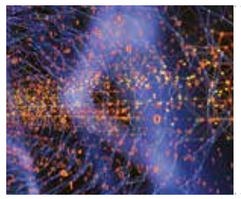
INFORMATICS SYSTEMS RESEARCH | DIRECTOR CARL KESSELMAN

This division has a broad research agenda focusing on creating new types of sociotechnical systems that enable and accelerate discovery in domains of high societal impact. It specializes in highly collaborative user-driven research in which work is evaluated in the context of operational, high-impact science.



MOSIS | DIRECTOR WES HANSFORD

MOSIS is an efficient, affordable way to prototype and volume-produce multi-project wafers (MPW) and related services that drive IC innovation. From design spec interpretation through mask generation, device fabrication and onto assembly, MOSIS is a trusted expert interface to the semiconductor ecosystem.



MULTI-DIVISION AND CENTER RESEARCH EFFORTS

Many of ISI's research efforts are multi-divisional and multi-centered—that is, scientists from different divisions and centers with complementary skills are working together to forge innovative and ground-breaking research, and also working together to solve problems in older programs to bring them up to today's technological standards.

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 is focused on:

BIG DATA · DECISION SYSTEMS · HETEROGENEOUS CLOUD AND EMBEDDED COMPUTING · MICROARCHITECTURE, INTEGRATED CIRCUITS, AND ADVANCED ELECTRONICS · QUANTUM COMPUTING · RECONFIGURABLE COMPUTING AND WIRELESS NETWORKS · SCIENCE AUTOMATION TECHNOLOGIES · SYSTEM SOFTWARE AND COMPILERS · TRUSTED ELECTRONICS AND COMPUTING

From Theoretical to Hands-On

Our current initiatives include theoretical adiabatic quantum computing through the USC-Lockheed Martin Quantum Computation Center and hardware security through the USC/ISI 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 algorithms and data structures to help understand and control large-scale complex systems such as oil field geometry, megacity emergency response and legacy software security.

CS&T teams are creating wireless networking 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—which enable researchers to focus on conducting science, not managing data—already used by astronomers, physicists and earthquake specialists, including the LIGO team that was the first to detect the gravitational waves that Einstein predicted.

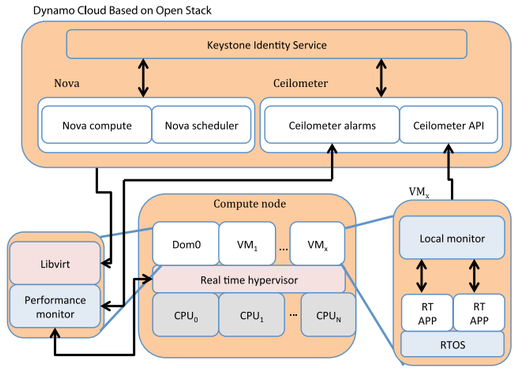
Intellectual Leadership, Diverse Expertise

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 40 computing technology researchers, research programmers and graduate students represent a wide range of disciplines, including electrical engineering, computer science, physics, and math. Led by Stephen P. Crago, CS&T researchers are based in Arlington, Virginia and in Marina del Rey, California.

Dynamo: A Framework for Dynamic Real-Time Workloads in a Reliable Cloud

Today's clouds are capable of high performance at tremendous scale. Unsurprisingly, the cloud represents a high-value target for cyber attackers interested in compromising cloud services and degrading critical infrastructure. The current cloud model pushes the responsibility of reliability onto the end-user, effectively forcing users into existing platform models in order to provide a degree of resilience. While this model is generally effective for existing IT-oriented applications (e.g., databases, directory services, general-purpose computation, etc.), it is poorly suited for mission-critical applications composed of heterogeneous architectures, real-time requirements, and critically, a spectrum of reliability requirements.



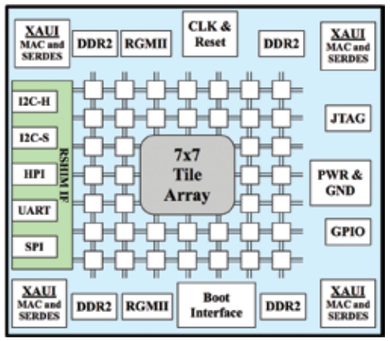
The Dynamo cloud and real-time resource management framework

The *Dynamo* team is developing tools, APIs, and cloud services to enhance the reliability of real-time and heterogeneous workloads in a real-time cloud. Using a dynamic resource management framework, Dynamo enables run-time resource reallocation to enhance the performance and reliability of real-time applications. In 2017, USC/ISI demonstrated the first use of dynamic resource allocation and fault tolerance for streaming analytics frameworks, reducing long tail latency and increasing fault resilience. Dynamo enables developers to leverage redundancy and resiliency for both fault tolerance and performance benefits, allowing adaptive performance and reliability tradeoffs to occur dynamically and automatically.

Dynamo is a component of the cloud computing group at ISI who are finding ways to integrate heterogeneous computing into dynamic public and private secure cloud platforms for data centers, high-performance computing, and embedded computing. ISI is one of the first organizations to develop support for heterogeneous computing in the OpenStack cloud environment, and it is now developing support for real-time applications and the deployment of such technology to new application domains.

Maestro: A Radiation Hardened Space Processor

As Earth-orbiting satellites generate increasingly large datasets, new approaches to data ingest and processing must be considered. One approach is to increase the satellite-to-Earth downlink capacity, and process these datasets on the ground. A second approach, however, would allow the satellite to perform processing and/or pre-processing onboard. This would reduce downlink requirements, and may improve the timeliness of the data through autonomous satellite operation.



The Maestro Processor

The *Maestro processor* and its follow-on, *Promethium*, are 49-core radiation-hardened-by-design (RHBD) processors for space missions. They are derived from the Tilera TILE64 processor, inheriting a full complement of high-speed interfaces, a complete Linux distribution, and a development toolchain. Our team of researchers is developing system software for the Maestro and Promethium processors, including enhancements to their compilers and toolchains.

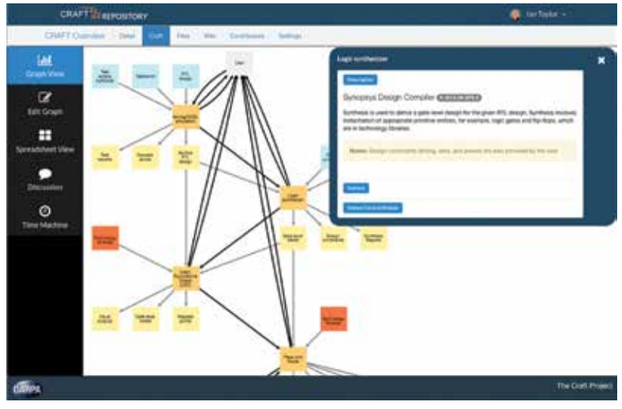
In 2017, USC/ISI developed the initial toolchain for the newly released Promethium processor, which led to improved floating point and I/O performance over Maestro. The toolchain included GCC compiler support, which enables full OpenMP support and additional language front-ends. This has allowed new application development to proceed for both Maestro and Promethium. Further, the switch to GCC resulted in improved performance for many applications. In addition, both OpenMP and the additional language front-ends will dramatically improve software portability to the Maestro and Promethium processors.

RACE: Repository and Workflows for Accelerating Circuit Realization

The Pegasus team (http://pegasus.isi.edu), in conjunction with faculty from USC's Electrical Engineering Department and researchers from the University of Notre Dame, are developing a new environment for collaborative ASIC application-specific integrated circuit design. The goal of this DARPA-funded project is to speed up the ASIC design process by supporting design flow and intellectual property (IP) reuse and sharing, while also enabling knowledge sharing and design automation.

In 2017 the first version of the repository was deployed within the *CRAFT (Circuit Realization at Faster Timescales)* program and used for depositing custom design flows and IP developed by industrial and academic program participants. Because of the sensitive and proprietary nature of information in the system, particular attention is being paid to user authentication and authorization.

The figure to the right is a graphical representation of just a small portion of a sample design flow that can be found in the repository.

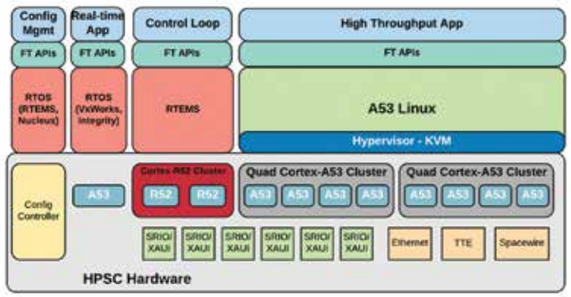


Sample design flow for RACE

System Software and Software Fault Tolerance for NASA's HPSC Processor

The NASA High Performance Spaceflight Computing (HPSC) project will build the next-generation space processor and software ecosystem for NASA. This new processor will revolutionize NASA's onboard processing capabilities, and will provide the computing power to support autonomous operations for deep space missions. It will also provide greater resilience to failure through combinations of radiation hardening and novel software-based fault tolerance.

ISI and The Boeing Company have teamed to provide the system software and processor for the HPSC program. Under this program, ISI will develop an open source emulator, system software, development tools, and novel software fault tolerance for the Boeing-designed heterogeneous system-on-chip.



System software stack for the HPSC processor

The software-based fault tolerance is key to the HPSC program, and enables a software-configurable hardware/software approach to fault tolerance, allowing developers to dynamically tune and reconfigure fault tolerance according to changing mission requirements. Elements of the HPSC software fault tolerance are tuned to the heterogeneous processing elements, providing developers with well-defined APIs on which they can build their applications.

In addition to the software fault tolerance, ISI is developing the core board support packages (BSPs) for major processing systems which will execute multiple coordinating operating systems.

Flux-Based Quantum Speedup (FLUQS)

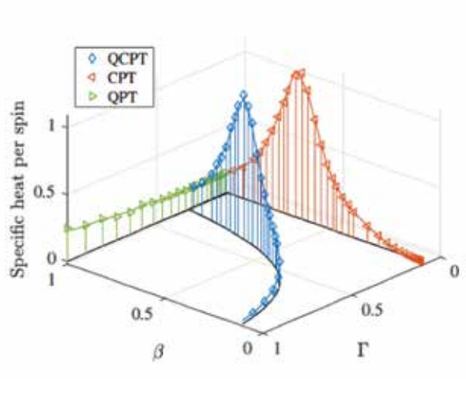
In 2017 our group began work on the Flux-Based Quantum Speedup (FLUQS) project, sponsored by IARPA under the Quantum Enhanced Optimization (QEO) program. This project, lead by USC, is a large collaboration between academic, research and private institutions. It aims at both designing and building the 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.

Development of a New Quantum Monte Carlo (QMC) Method

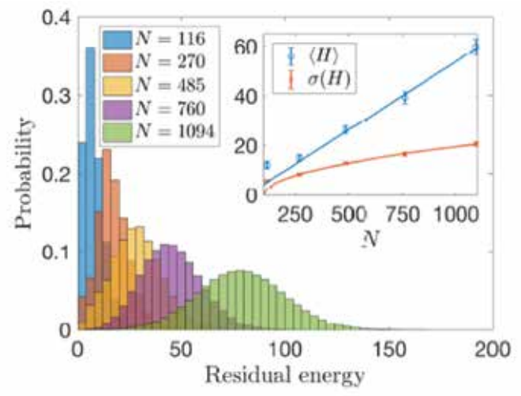
We proposed a Monte Carlo algorithm designed to simulate quantum as well as classical systems at equilibrium, bridging the algorithmic gap between quantum and classical thermal simulation algorithms. The method is based on a novel decomposition of the quantum partition function that can be viewed as a series expansion about its classical part. We demonstrated that the algorithm not only provides a theoretical advancement in the field of quantum Monte Carlo simulations, but it is optimally suited to tackle quantum many-body systems that exhibit a range of behaviors from fully-quantum to fully-classical, in contrast to many existing methods. We demonstrated the advantages, sometimes by orders of magnitude, of the technique by comparing it against existing state-of-the-art schemes, and illustrated how our method allows for the unification of quantum and classical thermal parallel tempering techniques into a single algorithm.

Temperature Scaling Law for Quantum Annealers

Physical implementations of quantum annealing unavoidably operate at finite temperatures. We showed that 1) a fixed finite temperature prevents annealers from functioning as competitive scalable optimizers; and 2) to serve as optimizers annealer temperatures must be appropriately scaled down with problem size. We derived a temperature scaling law dictating that temperature must drop at the very least in a logarithmic manner but also possibly as a power law with problem size. We corroborated our results by experiment and simulations.



Classical parallel tempering (CPT), quantum parallel tempering (QPT), and quantum-classical parallel tempering (QCPT). Results for an instance with $N = 60$ using our generalized parallel tempering algorithm along different curves in the β-Γ plane. Shown are the cases for CPT (red), QPT (green), and a case where $\Gamma = (10\beta)^{-1/2}$ (blue). To verify the accuracy of our algorithm we also show the PIQMC predication (solid line).



Distribution of residual energy, $E - E_0$, from DW2X simulations As problem sizes grow, the distributions become more Gaussian-like. **Inset:** Gaussians' mean (blue) and standard deviation (red) as a function of problem size, averaged over 100 instances per size. The solid lines are fits to a straight line and a square root, respectively, taking into account all sizes but the smallest.

Pegasus: Enabling Production-Grade Science

In the past year, the Pegasus Workflow Management System continued to be employed by a variety of scientific groups to engage in large-scale computational-based, production-grade science. Pegasus is a long-standing collaboration between a group at ISI and the HTCondor team at University of Wisconsin Madison. The system and associated research are primarily funded by the National Science Foundation and the Department of Energy.

In August 2017, for the first time, scientists were able to directly detect gravitational waves—ripples in space-time—and subsequently steer astronomers to observe light produced by colliding neutron stars. Detected by two identical LIGO detectors on August 17th, this was the first time that a cosmic event has been viewed in both gravitational waves and light. Pegasus allowed LIGO scientists to confirm the signal's significance by conducting rigorous offline analyses of massive amounts of data. Pegasus-managed pipelines were run on a variety of computational resources including the LIGO Data Grid (LDG) and the Open Science Grid (OSG). Professor Duncan Brown of Syracuse University, a member of the LIGO Scientific Collaboration, said: "Thanks to our collaboration with the Pegasus, OSG, and Condor teams, we can now turn around our offline analyses in days, not weeks. This is essential for getting confirmations of low-latency alerts and getting our results out to the world."

Visualization of a neutron star merger seen in gravity and matter. Photo/Karan Jani-Georgia Tech.

Proving the existence of gravitational waves as posited by *Einstein's Theory of Relativity* is not the only thing that requires large-scale computing power. California, being an earthquake-prone region, requires good estimates on seismic hazards that provide an insight into how the ground motions will be if an earthquake of a particular intensity strikes on a given fault. Seismologists quantify this seismic hazard for a location using probabilistic seismic hazard analysis, and the results are useful for civic planners, building engineers, and insurance agencies, and they impact billions of dollars a year in construction through building codes. These analyses require a significant amount of computational resources. The Southern California Earthquake Center (SCEC), based at USC, uses a Pegasus-managed workflow infrastructure called Cybershake to generate hazard maps for the Southern California region. In March 2017, SCEC conducted a CyberShake study on DOE systems ORNL Titan and NCSA BlueWaters to generate the latest maps for the Southern California region. Overall, the study required 450,000 node-hours of computation across the two systems.

Bioinformatics is another field where Pegasus is having an impact. An example is Alex Feltus' group at Clemson University, which has two workflows, OSG-GEM and OSG-KINC, for analyzing genes/transcripts in a gene expression matrix (GEM). In the last 12 months, the group used 8.4 million core hours on the Open Science Grid. As part of these computations, Pegasus managed 16 million data transfers totaling 4 petabytes of data.

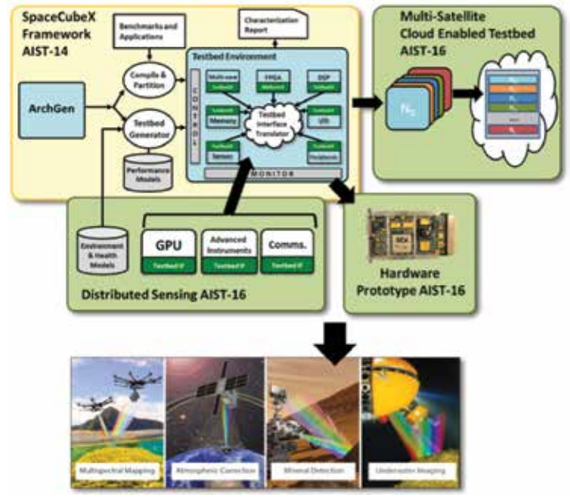
In addition to engaging with users from diverse communities, the team has been hard at work in implementing new capabilities to aid scientists in their work. In September 2017 the team released Pegasus 4.8.0 which included support for application containers and Jupyter notebooks. With support for both Docker and Singularity container solutions, the 4.8 release makes it easier for users to conduct reproducible science on public computational infrastructure. As part of a three-year NSF-funded project called *Scientific Workflow Integrity with Pegasus (SWIP)*, the team is implementing features that increase the trustworthiness in science. The upcoming Pegasus 4.9 release will see Pegasus automatically add data integrity checking steps at various points in the user's workflow to ensure that data has not been altered—either maliciously or accidentally by the infrastructure. http://pegasus.isi.edu

Quantum-Limited Information (QLI)

ISI's new laboratory for quantum limited information (QLI) is dedicated to understanding and demonstrating the fundamental physical limits for extracting information from physical signals. A complete understanding of a physical signal (such as electromagnetic energy like light or RF) is provided only when quantum mechanics is used to mathematically describe the signal. This quantum mechanical description gives us insight into the maximum amount of information that could possibly be extracted from the physical signal—such as detail in an image, data from a communications signal or information from a sensing signal. The QLI laboratory will calculate the fundamental limits for performing information processing tasks such as these, and then build laboratory experiments to demonstrate our ability to achieve these fundamental limits. The laboratory experiments will use traditional light sources such as thermal or laser sources as signals, but will also generate non-classical light, such as entangled light or quantum mechanically squeezed light. Our work will result in revolutionary designs for information processing systems that allow us to communicate and sense at the fundamental limits of nature and deliver capabilities in computing, communications and sensing that are not possible using classical physics.

SpaceCubeX

NASA's next-generation Earth science missions—which will help scientists answer critical 21st-century questions about global climate change, air quality, ocean health, and ecosystem dynamics—are projected to increase on-board processing requirements by 100 to 1,000x over the current generation of satellite systems due to higher resolution instruments and the move to continuous observations. The National Academy of Science (NAS) Earth Science Decadal Survey further defines high-priority Earth remote sensing science targets and the associated measurement requirements, including satellite constellations flying to take advantage of increased temporal sampling, enhanced hyperspectral instrument capability that enables next-generation land imaging, ocean biology and ecology, spectral reflectance measurements, and geo-based hyperspectral IR imagers/sounders.



SpaceCubeX framework

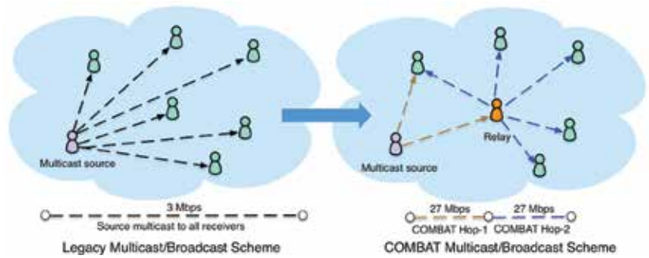
The Reconfigurable Computing Group's SpaceCubeX project has shown that heterogeneous computing (multi-core processors coupled with digital signal processors (DSP) or field programmable gate array (FPGA) co-processors) provides the fastest and most efficient throughput for high data rate applications; the group developed a framework which supports rapid exploration of the trade space of on-board heterogeneous computing solutions. These architectures were benchmarked across a suite of Earth science applications, achieving 20–20,000X performance improvements. The SpaceCubeX project continues to extend its on-board computing analysis framework to include graphics processing units (GPU) to support analysis of and application migration across airborne platforms with high bandwidth instruments and communication links to support next-generation missions, utilize cloud computing to scale to arbitrary constellation sizes, develop prototype hardware for capturing ad hoc processing needs, and demonstrate the technology to support adoption and transition. To date, the team—which includes USC/ISI, NASA GSFC, and NASA Ames—has performed over 500 benchmark experiments over five principal architectures, added GPU and cloud computing support, and begun the development of the SpaceCube 3.0 architecture, whose goal is to become the common processing hardware used on all NASA Earth science missions.

COMBAT: Cooperative Multicast and Broadcast At the Tactical edge

Under a grant from the Office of Naval Research (ONR), ISI's reconfigurable computing team, working with C-3 Comm. Systems and New York University (NYU), is developing advanced MAC and PHY protocols that will improve wireless multicast and broadcast communications at the tactical edge. Wireless communications at the tactical edge are critical for mission success. Rapidly changing channel conditions in this environment deteriorate communications throughput and put mission personnel at risk. Improving the reliability and throughput of these communications links is critical.

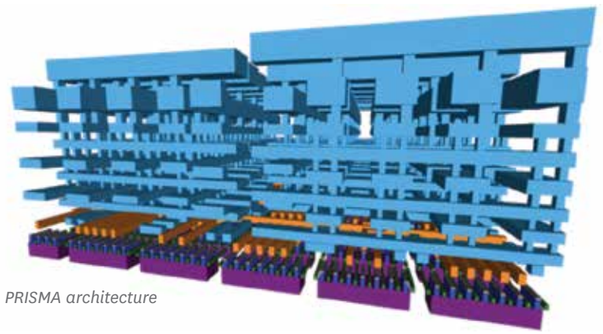
The COMBAT project seeks to improve the performance and reliability of these wireless communications at the link-layer through both PHY and MAC improvements. The COMBAT group is researching advanced techniques for real-time relay selection and improved methods for radio implementation on software-defined radio SDR platforms. Multicast throughput is expected to increase between 1.5-4X (depending on the network configuration) once implementation of the 11-node COMBAT testbed is complete.

Communications at the tactical edge



PRISMA Ptychography-Based Rapid Imaging of Nano-Structures with Multilayer Assemblies

Modern integrated circuits (ICs) are highly complex devices, where variations at the nanoscale can have dramatic impacts on performance and correctness. The ability to image these devices would revolutionize the semiconductor industry. However, imaging these devices, and the dozen or more metal layers that comprise them, is extraordinarily challenging. X-ray ptychography combined with tomography represent today's most promising approach to imaging ICs nondestructively; however, the computational requirements of ptychography and tomography at the sub-10 nanometer scale are substantial. ISI's PRISMA team is developing highly scalable ptychography and tomography algorithms that integrate machine learning with high-performance accelerators to achieve the petaflop-scale performance on petabyte-scale data sizes, enabling sub-10 nanometer x-ray imaging of modern ICs in a matter of days. In 2017 the PRISMA team took the first steps towards an integrated toolchain and developed the first end-to-end tool enabling the full ptychographic and tomographic reconstruction of synthetic imagery. In 2018 the PRISMA team anticipates extending this support to Argonne National Lab's Advanced Photon Source synchrotron.

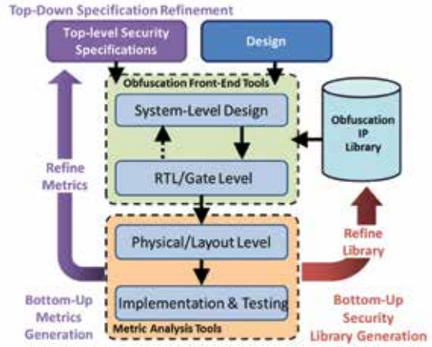


PRISMA architecture

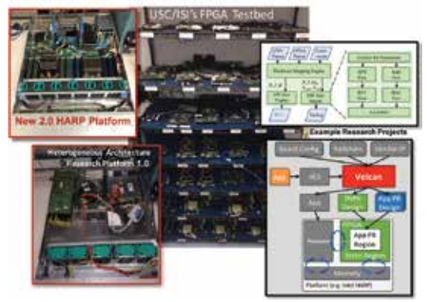
Mirage: A Security Metrics-Driven Obfuscation Design Environment

The globalization of the integrated circuit market has led to the rise of fabless semiconductor companies, leading to significant cost efficiencies. This comes at the expense of security, as companies can no longer control their IP through the entire IC design and production process. In order to leverage state-of-the-art capabilities provided by off-site commercial foundries in a secure manner, solutions are needed to ensure that: 1) the IP is not reverse-engineered and stolen during the fabrication process, and 2) the risk of hardware Trojans is eliminated through either prevention or detection and mitigation. Research and development in these areas has been active, but uncoordinated, resulting in a wide variety of passive and active obfuscation techniques that address specific pieces of circuitry or stages in the IC fabrication process—yet no agreement on metrics, much less end-user tools to select and assess the quality of obfuscation.

USC and ISI are developing *Mirage: A Security Metrics-Driven Obfuscation Design Environment* which treats obfuscation security as a first-class design constraint. This project addresses the increasing concern that IP is not reverse-engineered or stolen during the IC fabrication process. Mirage consists of two principal components: 1) obfuscation front-end design tools that perform design space exploration (DSE) over the range of active and passive obfuscation IP to find the appropriate mix to satisfy product objectives, and 2) metrics analysis tools that measure the achieved obfuscation security level in a unified manner, across combinations of both active and passive obfuscation. Mirage enables designers to select obfuscation IP and assess it in terms of trust, security, and reverse-engineering concerns.



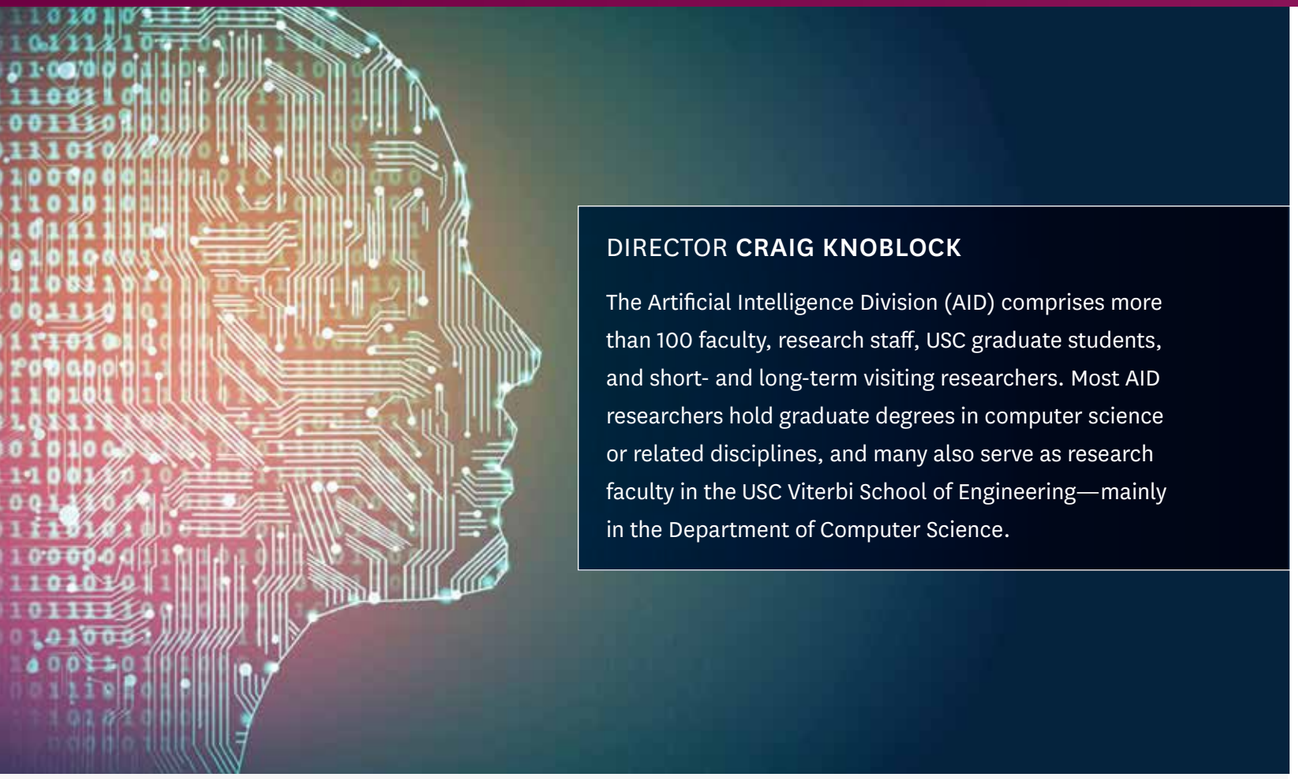
The Mirage Security System



Heterogeneous Accelerator Research Program (HARP)

Heterogeneous Accelerator Research Program (HARP)

During the past year, the Reconfigurable Computing Group at USC's Information Sciences Institute (ISI) and the FPGA/Parallel Computing Group at USC have been working collaboratively as part of Intel's Heterogonous Accelerator Research Program (informally known as HARP). Intel developed HARP to enable research across hardware and software for performance of diverse workloads in the data center as computer architectures move beyond homogeneous parallelism to incorporate application-customized hardware. USC/ISI was selected as one of three universities to receive the new HARP 2.0 platform which includes an Intel Xeon Processor and Arria10 FPGA integrated within the same device package. The USC/ISI team has a combined 16 graduate students, research staff, and faculty members from USC/ISI using the shared resource across research topics that include image processing, distributed big data and graph analytics, natural language processing, software-defined networking, and memory access utilization and optimization. To date, the work has resulted in five publications, four invited talks and demonstrations and two posters, with impressive results in natural language processing (170x speedup), deep learning (5x), high-throughput sorting (2-5x), and 5.8x speedup in convolutional neural networks (CNN) classification on FPGAs. Most recently, HARP is providing a research platform for Project Volcan which aims to bridge the gap between next-generation memory architectures and high-performance computing platforms.



DIRECTOR **CRAIG KNOBLOCK**

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.

AID is one of the world's largest artificial intelligence (AI) groups. It 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 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.

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

Knowledge graphs, using artificial intelligence and machine learning techniques to construct large-scale knowledge bases, with applications ranging from combating human trafficking to predicting cyber attacks.

Biomedical data integration that provides a semantically consistent view of, and efficient query access to, distributed, heterogeneous biomedical data.

Biomedical knowledge engineering, and innovative methods of building biomedical informatics systems based on cutting-edge AI techniques.

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

Machine intelligence and data science, focusing on developing efficient algorithms to analyze data from a variety of applications areas, including computational social science, cybersecurity, and biomedicine.

Video, image and multimedia analysis, including document image processing and face recognition.

BigDataU.org: Using Data Science to Teach Data Science

Data science seeks to provide efficient integration and analysis of increasingly large data sets. In the biomedical sciences the need is particularly acute. The availability of massive datasets in genetics, neuroimaging, mobile health, and other subfields of biology and medicine promises new insights, but also poses significant challenges. To realize the potential of big data in biomedicine, the National Institutes of Health launched the *Big Data to Knowledge (BD2K)* initiative, funding a number of centers of excellence in biomedical data analysis and a training coordinating center (TCC) at USC tasked with facilitating online and in-person training of biomedical researchers in data science.

A major goal of the BD2K TCC is to *automatically* identify, describe, and organize the training resources available on the web, and to provide personalized training paths for users. Given the large amount of available resources and the rapid evolution of data science techniques, we use data science methods to accomplish this goal. In particular, we use machine learning to identify high-quality videos on data science from public repositories like YouTube, where many scientific talks and courses are posted. In addition, we index courses from MOOCs, such as Coursera, Udacity, and EdX; scientific talks, keynotes, and tutorials from videolectures.net; and training materials generated by the BD2K Centers. We use machine learning to describe the contents of the training material—for example, a YouTube video that focuses on support vector machines or on probabilistic graphical models. We have developed an ontology with over 100 concepts to describe these resources. We are developing techniques to organize the resources—for example, one should learn about linear regression before logistic regression, based on an analysis of user clickstream behavior. Finally, we plan to combine the metadata about each training resource with user models to provide personalized learning paths.

Our web portal, **BigDataU.org**, currently indexes over 10,000 training resources, including courses, lectures, and scientific talks—the vast majority in video format. We plan to add several thousand data science books and other training materials shortly.

Doing Bayesian Data Analysis (John Kruschke)

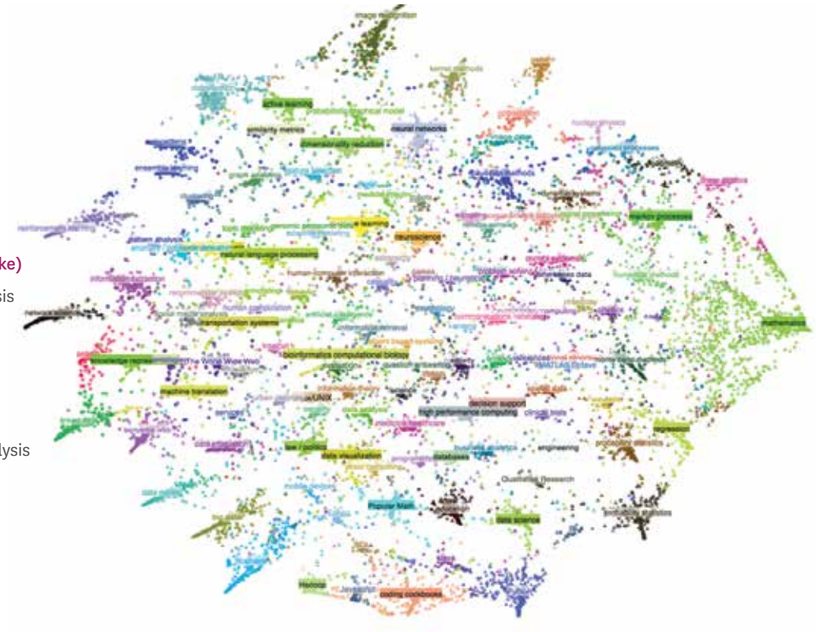
#26 Probability statistics (0.42) / data analysis

#48 Mathematics (0.2) / book theory

#85 Python (0.098 / programming Python

#6 Bayesian methods (0.084) /
Bayesian methods inference model

#11 Regression / models regression data analysis



This image shows a scatterplot derived from a simple topic model generated from BigDataU content and then laid out in two dimensions using the T-SNE embedding method. Each circle or triangle represents a video or book, respectively, and the layout reflects a high-level structure of the whole field.

Understanding Complex Systems through Rapid Model Integration (MINT)

Many complex systems, consisting of interacting phenomena and multi-layered processes, are studied by diverse disciplines with very heterogeneous modeling approaches and forecasting methodologies. The integration of those models is a painstaking and mostly manual process, yet it is required for understanding and predicting the complex behaviors of those systems. Addressing major societal and environmental challenges rests on our ability to rapidly analyze complex systems, predict their future behavior, and design interventions to affect future outcomes. ISI's new MINT project is investigating the use of artificial intelligence technologies to significantly reduce the time needed to develop integrated models of natural processes and human activities that affect resource availability with major economic and social implications. Research includes: 1) Ontologies based on principled methodologies to uniformly describe modeling variables, models, and data; 2) Semantic workflows that use deductive and abductive reasoning to select and integrate models with data transformation steps; 3) Data discovery and integration techniques to find new sources of data and automatically transform it into the format required by models; 4) Knowledge-guided machine learning algorithms for extracting information from remote sensing data to improve model accuracy; and 5) Multi-modal scalable workflow execution to run modeling scenarios.

Image Credits:
http://commons.wikimedia.org/wiki/File:MRI_brain_sagittal_section.jpg
http://commons.wikimedia.org/wiki/File:Earth_Eastern_Hemisphere.jpg

SAFT: Extracting Entities, Relations and Events from Text

As part of DARPA's DEFT program and in collaboration with CMU's Language Technologies Institute, Information Sciences Division has developed a multilingual information extraction and knowledge base construction system called SAFT (for Semantic Analysis and Filtering of Text). SAFT identifies entities and a small number of types of events, links entities into a knowledge base, fills a limited amount of event participant roles, and also detects a number of relationships between entities and events—for example, an organization that a person works for, or a location where a victim was attacked.

The output of SAFT is a knowledge base that lists each entity and each event it locates together with all their properties and relations. This entity and event-centric representation makes it possible to search in ways impossible when using a traditional web-style search, including chains of relationships. For example, one can start with a person of interest, find events he or she participated in, as well as other people related in various ways to him/her, and discover third parties only associated indirectly simply by exploring links in the knowledge base. The SAFT system was successfully evaluated in NIST's 2017 TAC-KBP ColdStart++ task on a corpus of 90,000 English, Spanish and Chinese text documents, and also demonstrated at the DARPA DEFT Demo Day in October 2017.

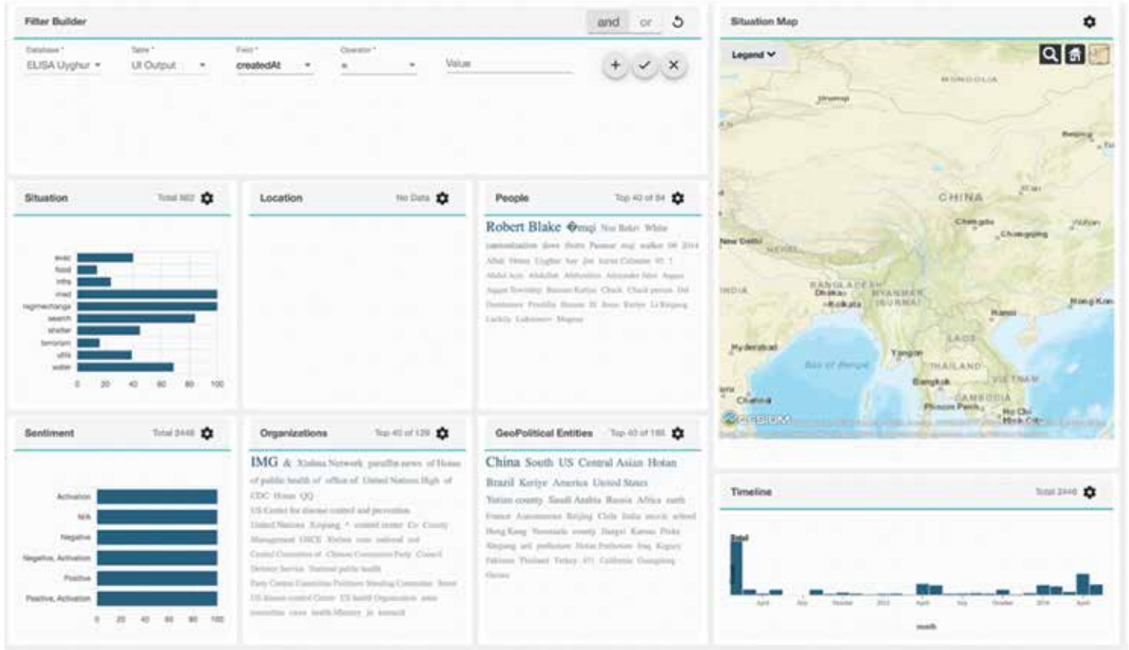
Exploring a SAFT knowledge base around entities related to the 2013 Boston Marathon bombing

Situational Awareness During Crises

The United Nations Office for the Coordination of Human Affairs (OCHA) reported that 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. Using interdisciplinary research from the data sciences to build a visualization and analytics platform for US civilian and military rescue personnel is an important problem with the potential for widespread long-lasting social impact.

Funded under the DARPA LORELEI program, and in collaboration with Next Century Corporation, we built the THOR (Text-enabled Humanitarian Operations in Real-time) framework, to provide humanitarian and disaster relief planners with superior situation awareness through visual and analytical data science. THOR uses open-source technology with an intent to support the Intelligence Community, as well as military and civilian disaster humanitarian relief operations. THOR presents a unique application of data science methods for an important social purpose. At its core, THOR is powered by a domain-specific knowledge graph, which is derived from natural language outputs, and on which further analyses (e.g., entity resolution and event clustering) can be conducted. THOR is designed specifically to account for the challenges in low-resource linguistic environments, heterogeneous data sources and types (with social media only being a subset), arbitrary disasters that may not be known in advance, and the requirement of advanced graphical interactions. We will be demonstrating THOR at the ACM Web Conference in 2018.

The figure shows a screenshot of the interactive tool to visualize humanitarian and disaster relief data extracted from multilingual social media.



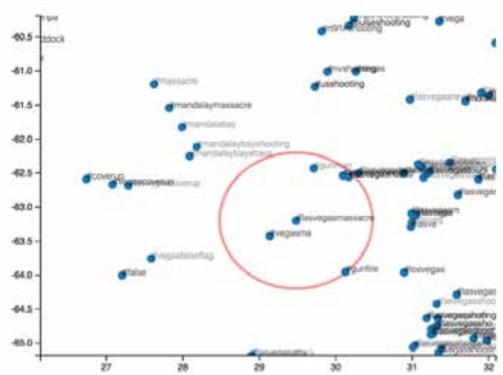
Interactive tool for visualizing humanitarian and disaster relief

Social Sensing and Visual Summarization of Social Media During Disasters

Due to the ubiquity of smart devices and cellular Internet, the latency between the occurrence of an event (such as a disaster) and its public dissemination on social media platforms like Twitter and Facebook by the 'crowd' has virtually vanished. These channels can be used for valuable information-gathering and updating purposes, but the sheer number of real-time messages, many of which can be redundant or irrelevant, makes this a daunting task. This has created an urgent need for automatic summarization and situational awareness tools that field analysts, law enforcement, and public agencies can use in near real time.

We have developed an open-source Twitter-based system, HashViz, that uses unsupervised text mining, representation learning and interactive visualization methods to summarize important information about a disaster from a stream of millions of raw tweets. Using data collected in the aftermath of the October 2017 Las Vegas shooting, as well as data on the many hurricanes during that period, we recently demonstrated how, by intelligently leveraging hashtags, the system can be used to answer key questions about these disasters (including both mundane details like locations, and more interesting phenomena like the circulation of conspiracy theories) with no manual effort. In a set of parallel efforts, we also developed 'social sensors' that analyze and tag social media messages along such dimensions as whether the message is expressing a need (which could be explicit or implicit), providing an update, expressing an opinion, or offering aid or resources to those who help.

The figure shows an interactive visualization of the hashtags related to the recent Las Vegas massacre, clustering hashtags related to the location, and conspiracy theories related to it.



Interactive visualization for hashtags

Web-Scale Network Analysis of Semantic Markups

Schema.org is an initiative that was jointly launched in mid-2011 by major search engines like Google and Bing; it is designed to facilitate a new era of structured and knowledge graph-centric search applications on the Web. The Web Data Commons (WDC) project has crawled increasing amounts of schema.org data in recent years, presenting a golden opportunity for socio-technological data studies that take the semantics of content into account. Until recently, this opportunity had not been realized because, unlike the hyperlinked ('regular') Web or Semantic Web ecosystems like Linked Open Data, schema.org annotations do not have a natural network structure. Recently, we developed a principled methodology for, and conducted empirical studies on, organizations in three different economic sectors (schools, hospitals and museums) that expose semantically linked schema.org annotations. We showed how to use the data that such organizations expose to link their schema.org annotations into a Web-like network and presented robust methods for analyzing this network.

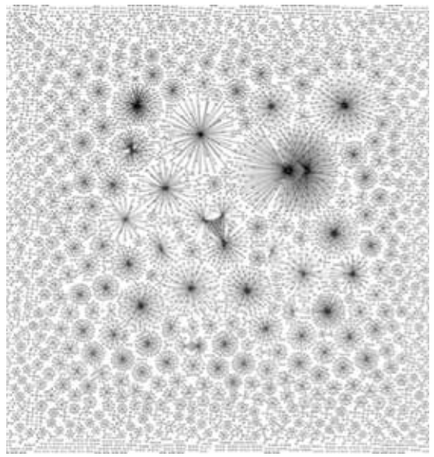
Our results, currently under late-stage review at *IEEE Computer Magazine*, highlight some interesting details about the (longitudinal and cross-sectional) state of semantic markup concerning these three organization types. We find, for example, that schema.org exhibits considerable diversity, with little overlap between the semantic content of different websites. At the same time, schema.org growth does not always exhibit power law distributions, which sets it apart from the early evolution of the hypertext Web. We also quantified the growth and dynamic activity of organizations' schema.org annotations.

Biomedical Domain Insight Graphs

Domain insight graphs are an area of great expansion within the information integration group at ISI. We seek to extend this core technology by applying the infrastructure developed by the DIG group to the more intricate, involved domains of neuroscience and molecular biology. DIG provides a scalable, linked data framework with many powerful, general capabilities. We seek here to leverage these tools in biomedical informatics applications to enable data sharing and a standardized interlinked platform for scientific analysis and reasoning.

Latent Factor Discovery Across Domains with CorEx

Often, the most important factors cannot be directly measured or, even worse, we are not even aware of their existence. Latent factor discovery attempts to identify and reconstruct the most important hidden factors lurking beneath a given dataset. We use the principle of *Total Cor-relation Ex-planation (CorEx)* to find the factors that are the most informative concerning relationships in data without needing explicit domain knowledge. This methodology has yielded valuable insights in several domains with recent notable success in identifying disaster-related documents with small amounts, of data and in analyzing gene expression datasets for personalized cancer treatment. An online tool for exploring latent factors in gene expression is available at http://corex.isi.edu.



Macroscopic trends (grouping of workers by ethnicity) in the illicit sex domain)

Computational Social Science in the Illicit Sex Domain

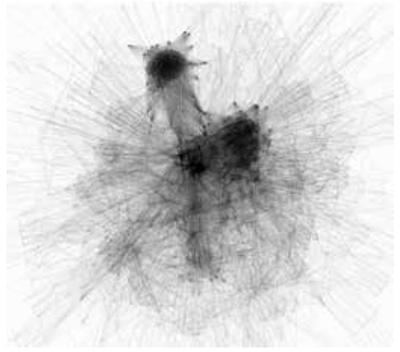
Recent investigative crawls in the online sex advertisement domain have revealed that, in the U.S. alone, the number of (not necessarily unique) sex ads is in the tens, if not hundreds, of millions. An unknown, but non-trivial, portion of these advertisements is targeted at human trafficking. Despite the seriousness of this issue, very little is known about the true extent, or underlying social organization of illicit sex activity, and studies at a national level have proven to be impossible because of a lack of data. In an ongoing collaboration, we are working with social scientists to analyze 25 million sex advertisements that have been crawled from backpage.com, by combining and applying advanced computational tools from several communities, including network science, visualization and record linkage. Using this data, we have constructed a national social network (containing more than 2 million edges) of illicit sex workers, and found evidence for macroscopic trends (e.g., grouping of sex workers by ethnicity) that have long been suspected but never quantified. We are currently studying the specific growth mechanisms in this network, along with mapping out the global properties of the network.

Studying Data Quality Without Gold Standards

Information extraction (IE) is a fundamental AI field that is defined as the automatic extraction of useful information such as named entities and relationships from raw text documents and webpages. Even state-of-the-art IE systems are considerably noisy in the general case. Without manually curated gold standards, it is difficult to compare the quality of different IE systems. Unfortunately, gold standards are expensive to create and may be biased.

Interestingly, noise in IE systems is rarely random—e.g., the city 'Charlotte' is much more likely to get mis-extracted as a name than the city 'Los Angeles.' Furthermore, in practice, if Charlotte is mis-extracted by an algorithm once, it will likely get mis-extracted several times. In state-of-the-art work on data quality, we attempt to use such relational information and 'mis-extraction redundancies' to study, assess and ultimately improve IE quality without a gold standard. We use network theory to characterize the structure of quality (and noise) by developing methods, metrics and visualizations that, despite not requiring a gold standard, are highly correlated with gold standard metrics on real-world data. A journal article that successfully applies our framework to IE in the difficult online sex trafficking domain is currently under review.

This figure shows a visualization of the network of data extracted from sex advertisements on the Web; higher density regions have a higher probability of noisy extractions.



Network of data from web sex advertisements

Entity Resolution (Almost) From Scratch

Entity resolution, also known as 'record linkage,' is a fundamental problem in the information integration community. The problem is defined as the automatic grouping of entities that refer to the same underlying entity. In the relational database community, each record (e.g., describing the personal information of an individual) typically represents an entity, while in many other communities, entities are represented as nodes in a knowledge graph. Entities are connected to other entities using relationships, revealing rich contextual information sets that could potentially be exploited by powerful representation and deep learning methods.

Thus far however, it has been difficult to reconcile recent advances in deep learning with traditional similarity-based entity resolution due to the reliance of solutions on hand-engineered, domain-specific features for achieving good performance. In our recent research, we show how these different approaches can be reconciled using sophisticated graph-theoretic modeling. We also show that our methods can be applied to ordinary records that do not have natural graph representations. The performance of these methods, which are minimally supervised and require no training examples, show that we are one step closer to automating entity resolution without suffering quality degradation.

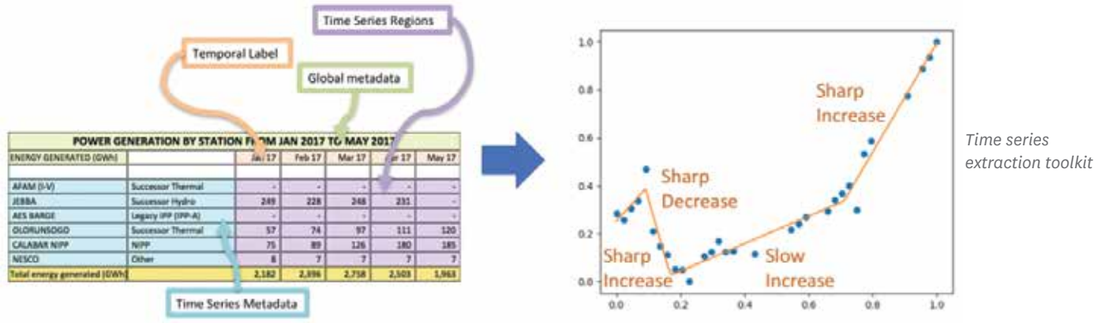
This table shows the ability of our method to handle different entity types such as geopolitical entities and locations. The method groups entities and automatically assigns an appropriate label to each group.

Label	Entities
Everest	Everest, Mount Everest, Mt Everest, Nepal Everest, Everest south side, white" - Everest, white"-Everest
Himalayas	Himalayan, Himalayas
Gorkha	Gorkha, Gorkha Region
Kathmandu	KATHMANDU, KATMANDU, Kath, Kathamandu, Kathmand, Kathmandu, Kathmandu's, Katmandu, Kathmandu
Tibet	Tibet, Tibetan

Entity types

Extracting and Interpreting Time Series for Causal Discovery

A wealth of knowledge is contained in time series data—ranging from commodity prices to censuses of populations displaced by famine. Unfortunately, this knowledge is often stored in spreadsheets that lack a standard format or labeling conventions. As part of the DARPA Causal Exploration project, we are building systems to unlock the data in these spreadsheets and make it easily accessible and interpretable to people. Our systems learn the layout and representation of spreadsheet data automatically, extract useful data into a knowledge graph, and provide a human-interpretable summary of the high-level trends in the data. Our time series extraction toolkit has been applied to datasets ranging from World Bank development indicators to food prices in Africa, generating a corpus of over ten thousand processed time series and accompanying interpretable trends.

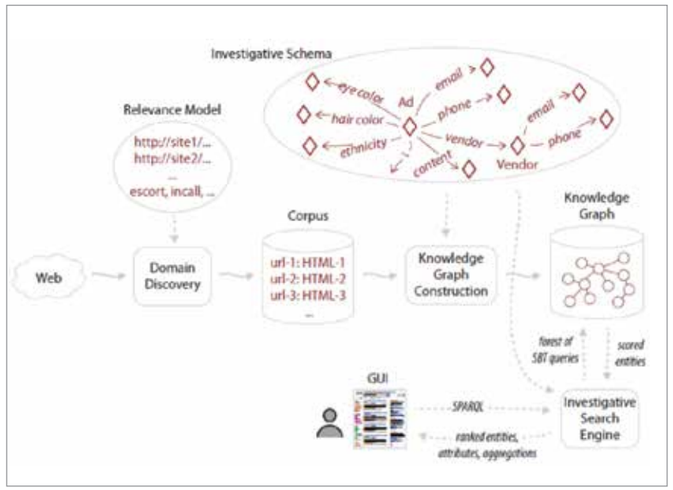


Time series extraction toolkit

Domain-Specific Insight Graphs (DIG)

An important aspect of answering complex questions is assembling and curating datasets that enable modern analytic methods to find hidden patterns and that support sophisticated query and visualization tools to explore connections. DIG is generic technology to make it easy to construct knowledge graphs for specific domains using information on the Web. DIG supports the full pipeline—starting with Web crawling to find and download relevant pages and data, information extraction designed to extract relevant information from structured and semi-structured text, data cataloging to align the extracted data to a common schema, entity resolution to identify information that refers to the same entity in the real world, knowledge graph construction to represent the information in a richly connected structure that enables query and pattern-finding, and a user interface to support sophisticated query and visualization.

The DIG technology (dig.isi.edu) is available as open source using the MIT license, and has been used to build knowledge graphs in many domains—including cyber-attacks, socio-political events, patent trolling, penny-stock fraud, firearms trafficking and human trafficking. Our human-trafficking knowledge graph powers a domain-specific search engine that has been used by law enforcement to investigate human-trafficking cases, including several that have led to rescuing victims and convicting traffickers.

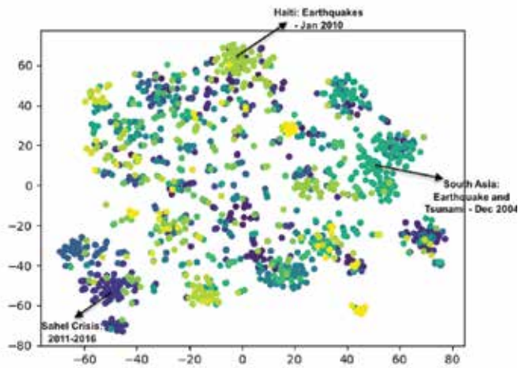


Domain-specific insight graph

Representing and Linking Structured Events Using Deep Learning

In recent years, extracting, analyzing, and even predicting events have emerged as critical problems in several important DARPA and IARPA programs—including use cases such as facilitating accurate geopolitical forecasting, causal exploration, and situational awareness in low-resource regions. While text documents continue to be a primary source of information for event extraction systems, structured datasets like the Global Terrorism Database, ACLED, and PITF contain thousands of ever-updating events that, if integrated and represented uniformly, could potentially be used to build entire geopolitical event dossiers. This is because in addition to a textual description of the event itself, structured sources also contain semantically distinguishable elements such as geolocations, dates, actors—and where applicable, attack (and target) types, weapons and casualties. At the same time, because events have complex information sets, automatically determining when two events are linked in the real world (or are sub-events of a larger underlying event), is an important but unsolved AI problem.

In our work we have built domain-specific models that represent the information sets in events as layers in multi-layer networks. For example, geolocations occupy one layer in this network, text descriptions occupy another layer, and so on. Connections exist between nodes both within and between layers. Once constructed,
we use representation learning techniques (using deep neural networks) to 'embed' events in a vector space, making them amenable to sophisticated but principled mathematical manipulations. Our results, currently under journal review, show that we can successfully predict links between terrorist events, and are also able to predict missing information about events with high accuracy.



Projection of event vector space

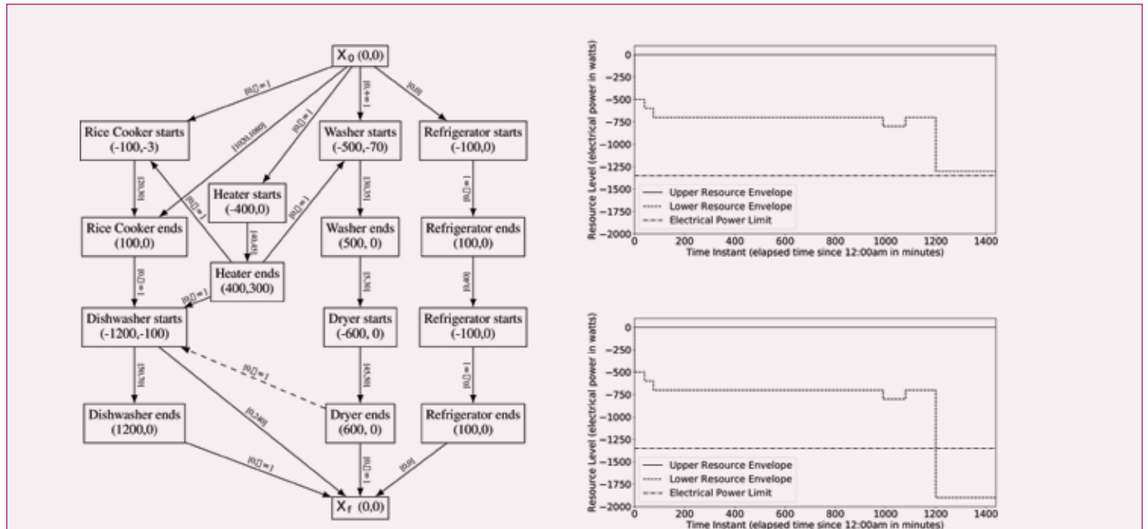
The figure shows a two-dimensional projection of the event vector space, illustrating how sub-events automatically cluster around an event—such as the 2004 earthquake and tsunami in South Asia.

Evidence Extraction

The scientific literature is a written record of the most crucial aspects of human knowledge and is made up of tens of millions of highly technical, structured documents. Although much progress is being made to develop AI methods to process natural language in general, we lack models that effectively process the discourse/argument structure of scientific work. In 2017, AI researchers began an NIH-funded project to extract information from both text and figures of molecular interaction experiments in an effort to build 'machine reading' systems for scientific evidence. This focus differs from standard methods that make no distinction between concrete scientific observations and the interpretive restatement of interpretations echoed from cited documents (which are notoriously unreliable). This project also involves team members from computer vision, natural processing, and the MINDS group.

Resource Envelopes

Look-ahead reasoning is a successful paradigm in artificial intelligence, with applications in constraint satisfaction, game-tree search, robust planning, etc. Unfortunately, look-ahead reasoning in the context of execution monitoring implies the forward projection of world states, i.e., understanding all possible ways in which the current world state can evolve. This is a hard problem in general because the number of possible future world states that are consistent with the current world state is just too large. To mitigate this, we proposed the idea of resource envelopes. Here, anything important to the execution of a plan is first deemed a resource. Then, for each individual resource, we can efficiently reason about all possible ways in which its production or utilization can change. These possibilities lie between an upper and lower envelope at all times, and they can be constructed using an efficient maxow procedure. Therefore, resource envelopes revive look-ahead reasoning in execution monitoring and constitute a cornerstone idea in this context. The idea of resource envelopes has been successfully applied in the SCHARP project at the Information Sciences Institute.



Resource envelopes and look-ahead reasoning

Web Text-Based Network Industry Classifications (WTNIC)

Businesses compete in many different ways, and understanding this competition is important for everyone—from government regulators to venture capitalists. In the NSF-funded WTNIC project, ISI is collaborating with professors in the Marshall School of Business at USC and the Tuck School of Business at Dartmouth to build better competition graphs. Our approach involves analyzing two decades of data that includes hundreds of millions of company webpages to predict competition between firms and understand how these competitive relationships evolve over time. Extracting data from webpages allows us to identify finer-grained relationships and extend coverage to privately traded startups, while feature selection algorithms and efficient hash-based comparisons allow us to scale to the massive dataset size while still maintaining performance.

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 three general areas:

- RESEARCH, METHODS AND INFRASTRUCTURE FOR CYBER EXPERIMENTATION
- NETWORK INFRASTRUCTURE SUPPORTING SCIENCE, DATA AND OPERATIONS
- ROLES OF PEOPLE IN CYBER SCENARIOS

Collaboration and community underlie all of our research areas—from experimental infrastructure co-development and the large user community to data sharing and Internet operations. In addition to these activities, we initiated three new educationally focused collaborations. In one, we are working with a consortium of universities to understand how people learn by using testbeds, while the other collaborations are focused on opportunities for high school and undergraduate students.

Research Methods and Infrastructure for Cyber Experimentation

This research area brings together strength in developing experimental methodologies and tools for research across a wide range of challenging problems of increasing scale and complexity.

Network Infrastructure Supporting Science, Data, and Operations

A deep understanding of the Internet's structure, behavior and operations focuses on Internet measurement, data sharing, and developing research topics drawn from operational support for the core Internet. In addition, high-performance advanced networking supports emerging national research platforms for global science initiatives.

Roles of People in Cyber Scenarios

Humans are often seen as the weakest link in cybersecurity. Modeling, analyzing and developing methods for understanding and adapting the human in cyber systems is a significant challenge. New initiatives in this area address behavior toward passwords and techniques for modeling and experimentation.

Sharing Network Security Data

The ANT Lab (Analysis of Network Traffic) has been generating and sharing network data with researchers for more than a decade. We draw on data from multiple sources, producing multiple kinds of output. Can our data help you? Contact us: https/ant.isi.edu/datasets/

Data Sources and New Kinds of Data

Network traffic: We collect network traffic with passive observers at several locations, anonymize it, and provide anonymized packet headers or network flow data.

Curated network traffic: We extract "interesting" network events, such as DDoS attacks, and document them to provide ready-to-use datasets to test intrusion detection and response.

Internet census and survey data: We report IPv4 address space use from regular scans of the entire IPv4 address space (censuses), and repeated surveys of 2% of the address space.

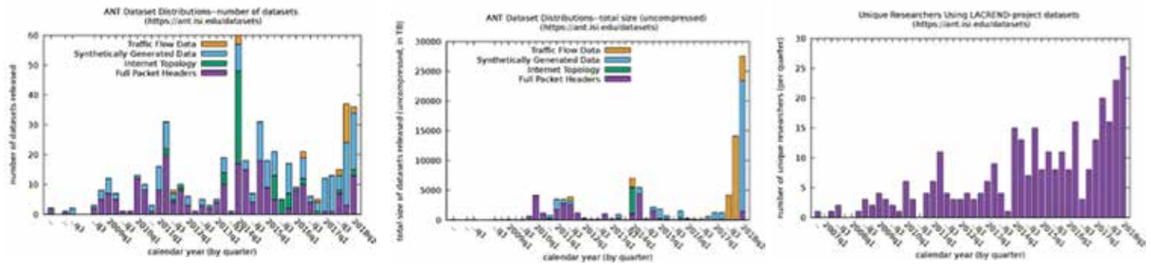
Internet hitlists: To assist IPv4 topology studies, we provide lists of target addresses in millions of networks that are most likely to reply to traceroutes.

Paper-specific datasets: We also curate datasets from published papers to make them available to other researchers.

Data to Researchers

A focus for our research group has always been to share data with other researchers in support of science. Since 2004 we have curated and provided cybersecurity datasets as part of the https://impactcybertrust.org/ program.

Since 2016, we have provided 1578 datasets (378 TB) to 238 unique researchers, with use growing. The graphs below represent this dataset use.



Dataset use since 2016

Outcomes: New Research Results

Our data feeds into new research publications from USC and other organizations. There have been three recent publications based on our work:

Anycast datasets: Verfploeter: Broad and Load-Aware Anycast Mapping, by de Vries, Schmidt. Hardaker, Heidemann, de Boer and Pras. ACM Internet Measurement Conference, 2017. https://ant.isi.edu/datasets/anycast \verfploeter

B-Root and Backscatter data: Detecting Malicious Activity with DNS Backscatter Over Time, by Fukuda, Heidemann, and Qadeer. ACM/IEEE Transactions on Networking, 2017. https://ant.isi.edu/datasets/dns_backscatter

Census data: Do You See Me Now? Sparsity in Passive Observations of Address Liveness, by Mirkovic, Bartlett, Heidemann, Shi, and Deng. IEEE Workshop on Traffic Monitoring and Analysis, https://ant.isi.edu/datasets/address

Large Synoptic Survey Telescope (LSST) Scaling Issues and Network Needs

ISI and Florida International University jointly presented their paper "Large Synoptic Survey Telescope (LSST) Scaling Issues and Network Needs" at the first National Research Platform (NRP) held at Montana State University in August 2017. The purpose of this workshop was to bring together representatives from interested institutions to discuss implementation strategies for deployment of interoperable Science DMZs at a national scale. The viewpoints of administrators, campus IT managers, ESnet, CENIC, Internet2, XSEDE, and the National Science Foundation were presented. Sessions of the NRP workshop were also devoted to science-driver application researchers, describing their needs for high-speed data transfer—including their successes and frustrations. Discussions focused on requirements from the domain scientists and the networking architecture, policies, tools and security necessary to deploy a 200-institution National Research Platform. All participants were encouraged to ask questions and share their thoughts on the topics under discussion.

ISI presented the challenges and the ongoing efforts involved with the LSST data management facilities and network requirements; the telescope's new astronomical scientific instruments demand increased real-time data transfers among scientists throughout the world. In addition, planning discussions concerning the end-to-end path, deployment calendar, network topology, SLA requirements for networking, and network monitoring and performance were presented as part of the initial draft for the LSST 2018 network.

For the last three years, the National Science Foundation (NSF) has made a series of competitive grants to over 100 U.S. universities to aggressively upgrade their campus network capacity for greatly enhanced science data access. NSF is now building on that distributed investment by funding a $5-million, five-year award to UC San Diego and UC Berkeley to establish a Pacific Research Platform (PRP)—a science-driven high-capacity data-centric "freeway system" on a large regional scale. Within a few years, the PRP will give participating universities and other research institutions the ability to move data 1,000 times faster compared to speeds on today's inter-campus shared Internet.

Modeling Human Behavior to Improve Tools and Policies for Cybersecurity

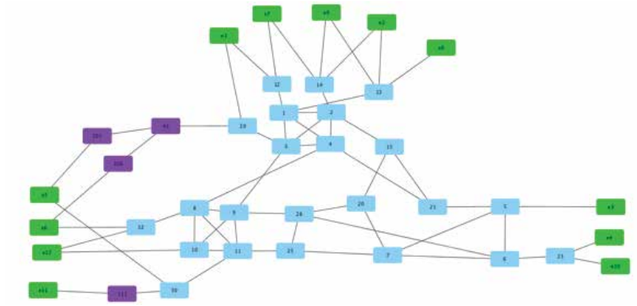
Human behavior is a key determining factor in assessing the effectiveness of an organization's cyber defenses, including policies. Our current research aims to observe and model important aspects of human behavior in order to predict the likely consequences of changes to the security posture of an organization.

The Networking and Cybersecurity Division has designed a cognitive agent architecture, DASH, that captures important aspects of human decision-making in cybersecurity, including dual-process reasoning and mental models. In collaboration with computer scientists and social scientists from USC, Dartmouth College, University of Pennsylvania, and Arizona State, DASH has been used to build agents that model defenders, attackers and end users. DASH simulations have demonstrated, for example, how increasing password strength requirements at a site can weaken the site's security by coupling it to otherwise independent sites due to prevalent coping strategies. We recently began working with DASH agents to enable high-fidelity modeling of online social behavior at scale, in order to gain a deeper understanding of the global information environment. With support from DARPA's Social Sim program and in collaboration with researchers from Indiana and Notre Dame, we are developing distributed simulations with millions of DASH agents—as well as techniques to optimize the simulation predictions using available data.

Toward Advanced and Realistic Network and Security Experimentation

As network systems continue to increase in complexity, it is imperative to be able to conduct scientific experiments to accurately determine the behavior of these systems under a variety of conditions. This requires the underlying models and assumptions to also be highly accurate within the bounds and constraints of the given experiment. Ensuring this fidelity while also providing a highly scalable experimental infrastructure is a difficult challenge.

In the past, the Networking and Cybersecurity Division has used the network emulation provided by DETER and modeling capabilities such as extensible components to be able to accomplish this goal. These technologies have enabled countless researchers to conduct accurate scientific experiments within the boundaries that are provided. However, in certain cases we may want to push the modelling and emulation accuracy beyond what DETER can provide, especially when trying to demonstrate system features to potential transition partners. This is exceptionally true when trying to meet the goals of DARPA's EdgeCT program, whose goal is to bolster the resilience of communication over IP networks by instantiating new capabilities at the WAN edge that mitigate impairments in the network backbone. To this end, our team has constructed the iLab in partnership with CenturyLink, the second-largest telecommunications provider in the United States. This laboratory environment replaces the emulation of DETER with the real hardware equipment that is deployed by CenturyLink in their portion of the Internet core and generates traffic models that are based on their operational traffic. This is critical for our April 2018 demonstration event, where multiple potential transition partners will be in attendance to observe the EdgeCT technologies and possibly build future business relationships. Demonstrations using real-world routers and telecommunications devices fielded in CenturyLink's operational network provide the necessary confidence that transition partners within private industry and the DoD require. This environment also provides ISI with the unique opportunity to compare our emulation and modeling capabilities against the real hardware being modelled in a realistic environment—thus allowing us to improve our modeling accuracy and provide higher fidelity in future experiments.



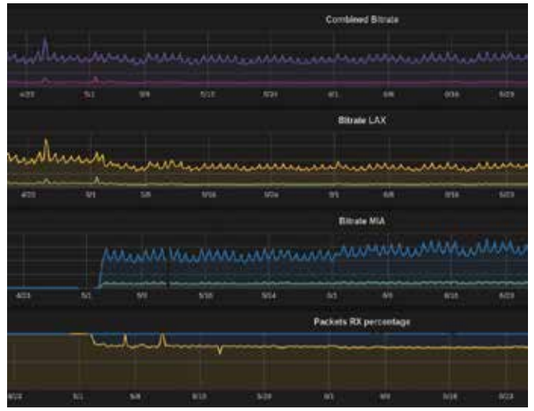
Communication over IP networks

Additionally, we are continuing to enable advanced experimentation by developing new and enhanced emulation capabilities. The DARPA Dispersed Computing (DCOMP) program seeks to create scalable, robust decision systems that enable secure, collective tasking of computing assets in a mission-aware fashion by end users with competing demands, across large numbers of heterogeneous computing platforms. As part of DCOMP, we are creating a novel wireless emulation framework that allows thousands of wireless devices to communicate with high fidelity in a PHY-layer agnostic fashion. By leveraging our extensible components technology and the flexible models it provides, we can rapidly deploy a prototype wireless emulation engine with sufficient accuracy for the early stages of the program, while refining and enhancing it as the program develops. Further, we can leverage the same extensible components capability to emulate additional network components required for DCOMP that we do not currently offer, such as network switches or GPU servers. By using this technology, we can easily allow an experimenter to instantiate and deploy one of these components within their experiment with confidence that the component is accurately modeled—and at the same time, not require the experimenter to have domain knowledge of the component being modeled. The combined effect of the progress made in these two directions will result in more accurate, realistic and better scientific experiments in the future.

B-Root DNS Infrastructure

ISI's B-Root DNS critical infrastructure serves as the foundation of 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. 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. To further both the service to the community and to promote new research opportunities, in May 2017 we added the first anycast instance for B-Root in Miami, Florida (in partnership with Florida International University), and we plan to add an Internationally located third site in the coming year.

ISI is playing a critical role in the development of an initial Internet governance model for the DNS Root Server system. We are working within ICANN to define an open, transparent, and accountable policy framework and governance model for the management of the Internet's top-most DNS root servers (such as B-Root). The planned initial release is expected in mid-2018. It will establish a framework for root server system oversight that has been missing since the untimely 1998 death of ISI's Jon Postel, Internet Hall of Fame pioneer.



This graph shows the traffic arriving at B-Root before and after the MIA anycast site was activated.

B-Root actively supports USC and ISI's mission as a research university. The experience gained from operating B-Root helps ISI researchers understand real-world constraints and challenges—such as routing challenges in supporting anycast networks and the implications of large-scale denial-of-service attacks. These experiences have translated into new peer-reviewed technical papers at networking conferences (for example, *Verfploeter: Broad and Load-Aware Anycast Mapping*), joint work between ISI and University of Twente in the Netherlands, and made possible through the addition of B-Root's second site; and *Recursives in the Wild: Engineering Authoritative DNS Servers*, joint work between ISI, University of Twente, and SIDN Labs. Work on B-Root resulted in multiple new research opportunities with NSF, DARPA, and DHS in 2017, with more opportunities already underway for 2018.

Cybersecurity Experimentation of the Future

As a leader in cybersecurity experimentation, ISI has seen the emergence of a new experimental paradigm—*Cybersecurity Experimentation of the Future (CEF)*—which is targeted at catalyzing a new generation of experimental cybersecurity research. Building on the experience and technology of the DHS S&T-funded DETER Cyber Security Testbed, a transition path has been identified that will create an overarching framework for interconnecting the explosion of testbeds into a coherent set of capabilities. This broader architecture, framework and sets of tools transitions from supporting primarily academic research and education to supporting a wide range of researchers, educators and cyber-solution developers and government users.

CEF Vision. This is now being realized through a first-generation framework, *ceftb,* that serves to enable the following:

Leverage the proliferation of Testbeds.

Raise the level of abstraction for cyber-experimentation and test beyond academic research across to the broad cyber solution developers and government users.

Create open community of testbed providers and experimenters.

Define architectures and infrastructure to interconnect: 1) established testbeds with substantial existing infrastructure; 2) small-to-medium specialized university or commercial labs; 3) individual unique components in industry and national labs.

Provide tools and methodologies for experiment lifecycle sharing, reuse, validation.

CEF Capabilities. Taken together, these areas paint a vision for a new generation of experimental cybersecurity research—one that offers powerful assistance towards helping researchers shift the asymmetric cyberspace context to one of greater planning, preparedness, and higher assurance-fielded solutions. These new capabilities will enable partners, agencies and private industry to conduct the following:

Identify innovative solutions to fill mission-need gaps and define future technology roadmaps through the use of advanced experiment, test and validation services of the CEF.

Evaluate and improve the effectiveness of security solutions, and promote rapid technology innovation adoption through rigorous scientifically based assessment.

Develop and standardize architectures that enable rapid insertion and integration of existing, as well as future, cyberspace defense technologies and infrastructures through shared integration and development environments and tools.

Contribute to proofs-of-concept of draft standards and specifications through agile experimentation and piloting activities via the CEF Ecosystem.

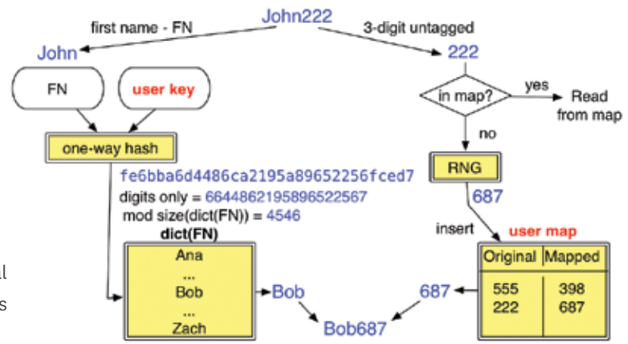
Understanding How Users Choose and Reuse Passwords

It is no secret that users have difficulty choosing and remembering strong passwords—especially when asked to choose different passwords across different accounts. While research has shed light on password weaknesses and reuse, less is known about user motivations for following bad password practices. Understanding these motivations can help us design better interventions that work *with* the habits of users and not *against* them.

ISI researchers, in collaboration with the University of Illinois at Chicago, investigated these questions via an innovative human user study that will be published in the prestigious *ACM Conference on Human Factors in Computing Systems (CHI)* in 2018.

In the study, real user passwords are collected and analyzed. To protect user privacy, passwords are transformed into different strings in a non-reversible manner. The resulting strings, however, keep their semantic structure (e.g., personal name+number), thus allowing researchers to study passwords that are similar but not the same, and to reason about password structure. The study also includes users' beliefs and reasoning about their password choices.

Collecting both actual and user-perceived data about passwords enables a comparison between the users' actual behaviors with their intentions. Researchers found that user intent often mismatches practice.



Semantic transformation example: John222 is first semantically tagged as a first name and 3-digit number combination. John is then transformed into another name (Bob) using per-participant random key and one-way hash. The 3-digit number 222 is transformed into another random 3-digit number 687, which is stored in the per-participant map. After the participant completes the study, their key and map are destroyed, making the transformation non-reversible.

All users underestimate the number of accounts they have; they narrate more rational reuse strategies than they exhibit; and they narrate different password-composition strategies than those they employ. This cognitive inability of users to accurately track their password habits, along with misconceptions about risk and convenience, fosters bad password habits. All users in the study reused their passwords: 98% reused them without any modifications, and 90% of users had weak passwords, vulnerable to offline guessing attacks.

Cyber-Physical Systems

Our cyber-physical systems research continues to advance the state-of-the-art in computational and experimental methods for the study of large-scale, complex cyber-physical systems, and then to apply these methods to address critical infrastructure challenges across domains such as energy systems, smart cities, transportation, natural gas, and water supplies. We continue to work toward our core goal, which is to create new methodologies and tools that will provide clear, correct, and actionable insights into the design and operation of these increasingly central components of modern society with projects funded by NSF and DHS. Combining a strong, multidisciplinary team at ISI with collaborations across multiple USC centers and laboratories, private and public corporations, and other academic institutions, our research synthesizes ideas and methods from disciplines that include statistics, information theory, systems, optimization, economics, policy, network science, and control theory. These disciplines provide techniques for understanding complex systems and for presenting design principles and architectures that allow for system performance quantification, resilience, and optimal management.

Novel Architecture for Experimentation (in support of DCOMP)

Cybersecurity and network testbed architects must address changing demands from researchers that reflect the challenges of studying novel network protocols and system protoypes in forward-looking scenarios. ISI's newest testbed design integrates several key technical advances to deliver a cost-effective and rapidly deployable DCOMP cluster of 1,440 embedded nodes. The design combines Intel Atom small form-factor nodes custom-manufactured for high-density testbed deployment and manageability with commodity high-speed networking switches. Rapid integration of our CEF next-generation testbed software with this cluster yields a solution to rapid reconfiguration—a key experimental challenge.

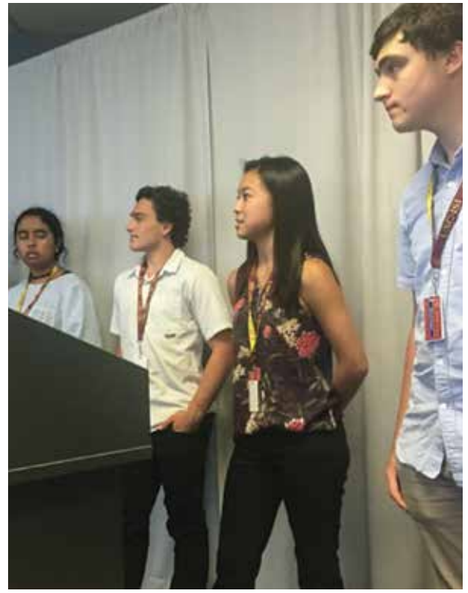
To support interactive experimentation, testbed nodes must reboot and reload quickly while testbed networks must rapidly change complex interconnectivity. This is especially critical for sharing limited resources among different researchers working on multiple simultaneous experiments with similar requirements. The DCOMP cluster enables multiple distinct experiments of varying sizes to be simultaneously instantiated and conducted with minimal cross-experiment interference, as well as single large-scale experiments using all nodes.

ISI's design for OS image loading leverages high bandwidth distributed storage networks and collaboration with the open source developer community behind the latest technologies in the secure pre-boot space—such as LinuxBoot. The benefits of rapid reconfiguration include dynamic, high-tempo reallocation of resources from one experiment to another for short durations, as well as refreshing experiments to a known and precise state in support of scientific repeatability.

The DARPA Dispersed Computing (DCOMP) program seeks to create scalable, robust decision systems that enable secure, collective tasking of computing assets in a mission-aware fashion by end users with competing demands, across large numbers of heterogeneous computing platforms. Researchers are designing systems able to operate in environments where network connectivity is highly variable and degraded. The envisioned computing paradigm aims to enable the strategic, opportunistic movement of code to data, and data to code, in a fashion that best suits user, application, and mission needs. Provisioning the DCOMP testbed cluster enables ISI to provide experimentation capabilities necessary to support multiple research project teams throughout the life of the program.

DETER Lab Hosts High School Students

For a second summer, the Networking and Cybersecurity Division's DETERLab hosted ten high school students for a five-week internship at ISI. The internship focused on leveraging the experimentation tools in DETERLab to model and evaluate cyber security threat and develop solutions. In the first summer, the group of students worked on developing a representative model of the school campus as a pro-consumer of electricity. They computed the power that it could generate annually based on rooftop space, and gathered information on power consumed by the school from its local facilities. Based on that information, they developed models to buy and sell electricity from the local power utility—all with a focus on how to develop long-term sustainability on campus. This year the students self-organized into three groups—each focused on a different area of research. The first group evaluated the impact of attacks on remote stock trading. They created an online trading market based on past history data and studied the impact of attacks that delay or block trading. The second group studied the Internet protocol suite, TCP and IP, and emulated a range of recent attacks reported in the research literature. The third group created a framework to analyze network traffic and generate narratives to describe what is observed in the data.



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.

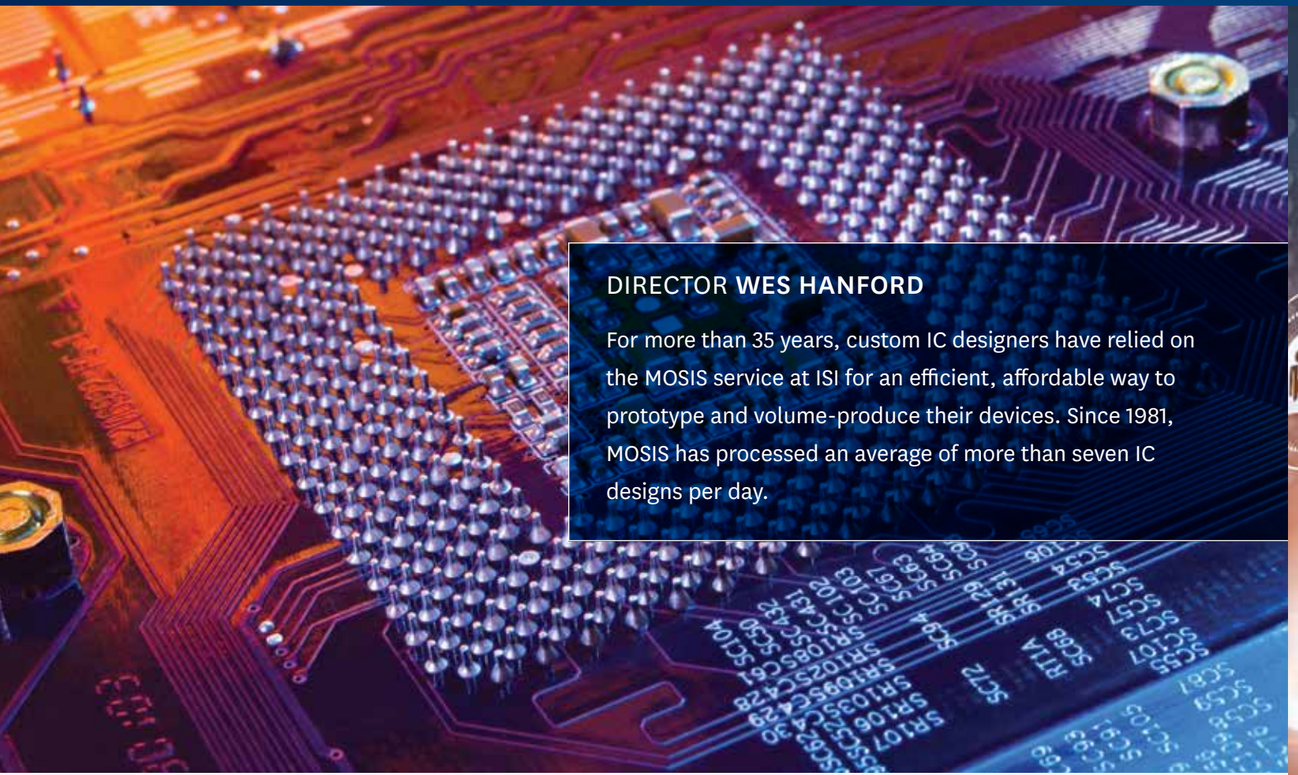
More recently, the division has focused on biomedical applications. Our current collaborations cover a broad range of applications—from basic science to clinical use cases spanning molecular biology, basic neuroscience, neuroimaging, stem cell research, and cranio-facial dysmorphia.

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 was choosing to house the data collection for the GenitoUrinary Development Molecular Anatomy Project—an international consortium devoted to understanding the development of the genitourinary track in humans.

ISI's Informatics Systems Research Division is developing innovative new approaches to accelerating scientific discovery by focusing on eliminating the complexities associated with assembling, organizing and manipulating complex, large-scale data collections. With the advent of "big data," the situation has in many ways become worse. 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.

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., Findable, Accessible, Interoperable, and Reusable. Application of Deriva was featured in a cover article in the *Journal of the American Society for Nephrology* (shown here). This work, led by USC's Broad Center for Stem Cells and Regentitive Medicine, presents a significant advance in our understanding of how kidneys form in humans and mice. The data, obtained and contributed to by our Informatics Division, is made available to a global community of researchers as part of the GenitUnary Tract Molecular Atlas (GUDMAP,) which was created and is hosted by ISI.

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.



DIRECTOR **WES HANFORD**

For more than 35 years, custom IC designers have relied on the MOSIS service at ISI for an efficient, affordable way to prototype and 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. It’s a practical prototyping channel that allows designers to debug and perform essential design adjustments before making a substantial strategic investment. Today, with mask 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 increasingly choosing MOSIS as their resource partner for volume-production. From design spec 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: 1) sharply reduce the amount of effort required to design high-performance custom integrated circuits, 2) greatly facilitate porting of integrated circuit designs to secondary foundries and/or more advanced technology nodes, and 3) strongly increase reuse of integrated circuit elements. In support of CRAFT, MOSIS is organizing 14/16nm private MPW runs.

DARPA DAHI (Diverse Accessible Heterogeneous Integration) Transistor-scale heterogeneous integration processes to combine advanced compound semiconductor (CS) devices, and other emerging materials and devices with high-density silicon complementary metal-oxide-semiconductor (CMOS) technology. The goal of DAHI is to establish a manufacturable, accessible foundry technology for the monolithic heterogeneous co-integration of diverse devices and complex silicon-enabled architectures on a common substrate platform. MOSIS is teamed with Northrop Grumman Corp., and organizing private MPW runs (65/45nm).

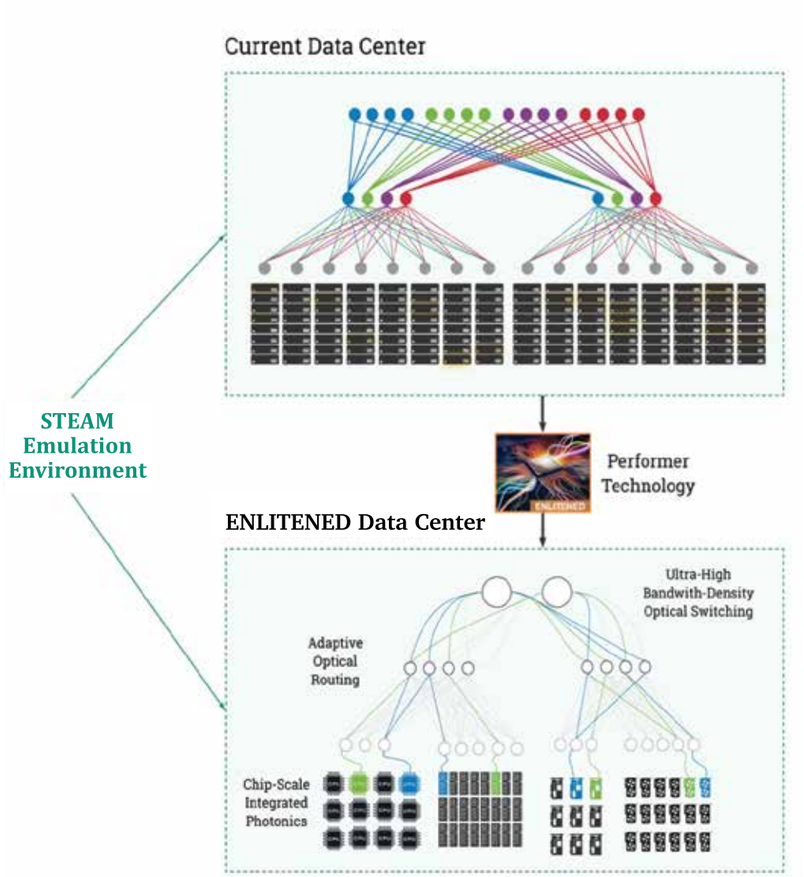
IARPA TIC (Trusted Integrated Chips) Split-manufacturing, with a new approach to chip fabrication where security and intellectual property protection are assured. MOSIS is on the government team, and organizing private MPW runs (130/65/28nm).

STEAM

STEAM is a testbed and evaluation environment funded at $1M to support the ARPA-E ENLITENED program, whose goal is to conduct R&D of optical networking technologies that can reduce the energy consumption of datacenters by at least 2x. The technologies being developed in the program may be broadly portioned into two categories—chip-scale photonoics (optical networking directly to chips) and warehouse-scale photonics (ultra-high bandwidth density switching and routing).

STEAM researchers from multiple divisions are creating an environment where they can design systems of the future before the underlying technologies exist. Emulations of future technologies are created according to their design specifications and forecasted performance characteristics. The emulated versions of these technologies are then placed into system-level models side-by-side with technologies of today—both real and emulated.

The result is the ability to do comparative analysis into the future as the diagram shows. If we suppose that technology like adaptive optical routing and switching exist at terabit bandwidths (with associated control plane), and there are low insertion losses across the network, port densities of thousands per device, and disaggregated integrated photonic devices such as CPS, RAM DIMMs and GPUs that can connect directly to these networks with the protocols to coordinate them, then what does the ENLITENED datacenter look like? STEAM allows us to explore that question.



GPU graphic by Batthias Smit from Noun Project. Licensed under Creative Commons.

ENLITENED graphic taken directly from ARPA-E Program page.

COSINE: Cognitive Online Simulations of Information Network Environments

COSINE is a five-year, $5M collaborative project involving multiple divisions at ISI and collaborating researchers at Indiana University and the University of Notre Dame.

The goal of COSINE is to create a first-of-its-kind cognitive agent simulation framework for studying multiscale 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 will be modeled using top-down statistical-physics models, mesoscopic-level compartmental and network-based models, and bottom-up agent-based dynamics. Agent models will be based on neurocognitively grounded principles of human behavior that incorporate bounded rationality and cognitive biases within models of attention. The system will be calibrated on real-word data collected from a plethora of online platforms. COSINE will create a virtual laboratory for studying the dynamics of online social phenomena at different temporal resolutions and at multiple scales, from individual to community to global collective behavior.

EFFECT: Forecasting Cyber Attacks

In late fall of 2016, dozens of popular sites, including Twitter, Netflix, and Paypal, became unreachable due to a massive cyber attack directed against the infrastructure of the Internet. Attackers exploited a vulnerability in the software used by Internet of Things (IoT) devices that enabled them to commandeer vast numbers of such devices and use them to unleash a massive denial-of-service attack against the domain name system services provider. The overwhelmed domain name services provider stopped routing Internet traffic to its proper destinations, which resulted in widespread disruptions of normal activities.

The attack did not come out of the blue. In fact, there were signs suggesting it would happen weeks prior to the attack. If caught early enough, these signs could have allowed organizations, such as the DNS services provider, to anticipate and prepare for the attack. The first signs emerged in a blog post that described a vulnerability in the operating system of IoT devices which would enable an attacker to remotely gain control of the device; a software kit exploiting this vulnerability was then publicly released. The news spread rapidly online, and it took only a little more than three weeks for hackers to weaponize the software to take over thousands of IoT devices and launch an attack.

The goal of the IARPA-sponsored EFFECT project (funded at $15.2M) is to monitor publicly available data sources to anticipate cyber attacks. ISI is leading a team that is working to address this challenge. The team includes researchers from multiple ISI divisions whose researchers have complementary skills in information extraction, data fusion and integration, cyber security, socio-behavioral modeling, and machine learning. We are building a system that monitors a multitude of data sources on the open and dark web, extracts indicators of the preparation and planning of cyber attacks, and analyzes these indicators to produce warnings of attacks. Members of the EFFECT team also include academic and industry partners: Arizona State University, Raytheon BBN, Hyperion Gray, Lockheed Martin ATL, and Ruhr U. Bochum.

BATL (Biometric Authentication with a Timeless Learner)

Today's access to enterprise, personal, and societal information increasingly relies upon technological adjudication of identity. BATL researchers are creating systems and algorithms resilient to presentation attacks on biometric authentication systems—including face, iris and fingerprint systems. Also under research are methods for detecting previously unseen presentation attacks—which then enable them to learn about the characteristics and nature of those attacks. Collaborators in this $13.8M research effort include the Switzerland-based Idiap Research Institute, Darmstadt University of Applied Science, and Northrup Grumman.

DiSPARITY (Digital, Semantic and Physical Analysis of Media Integrity)

Nefarious actors on social media and other platforms often spread rumors and falsehoods through images whose content and/or metadata (e.g., captions) have been modified so as to provide visual substantiation of such rumors. Sponsored by DARPA's MediFor program,ISI's DiSPARITY research team is focused on automatically identifying and characterizing signs of manipulated images, video, and metadata. This four-year, $8.6M research effort will develop algorithms to automatically detect and characterize the visual and semantic attributes of manipulated digital media—as well as the genealogy of a manipulated image given a world data set. Key challenges to achieving these goals include the wide variety of imaging devices, the increasing sophistication of manipulation tools and techniques, including rapid advances in computer graphics technology, and the sheer volume of data to be analyzed. In addition to the research team at ISI/USC, the DiSPARITY team includes researchers from the University of Erlangen-Nuremburg and the University of Naples.

ELICIT (A System for Extracting and Organizing Causal Information)

The goal of the ELICIT project is to develop an innovative, optimized knowledge organization system that integrates powerful concepts of causality, factual knowledge, and meta-reasoning into a model-driven knowledge graph representation to provide operational environment knowledge from these heterogeneous and semantically diverse sources to military planners. Today those planners face complex operational environments, influenced by a diverse spectrum of political, territorial, economic, ethnic, and religious factors. Existing methods to holistically understand these environments are both labor-intensive and time consuming. A $6.4M research effort sponsored by DARPA, ELICIT includes researchers from ISI, Rensselaer Polytechnic Institute, CMU, and Lockheed Martin ATL.

ELISA (Exploiting Language Information for Situational Awareness)

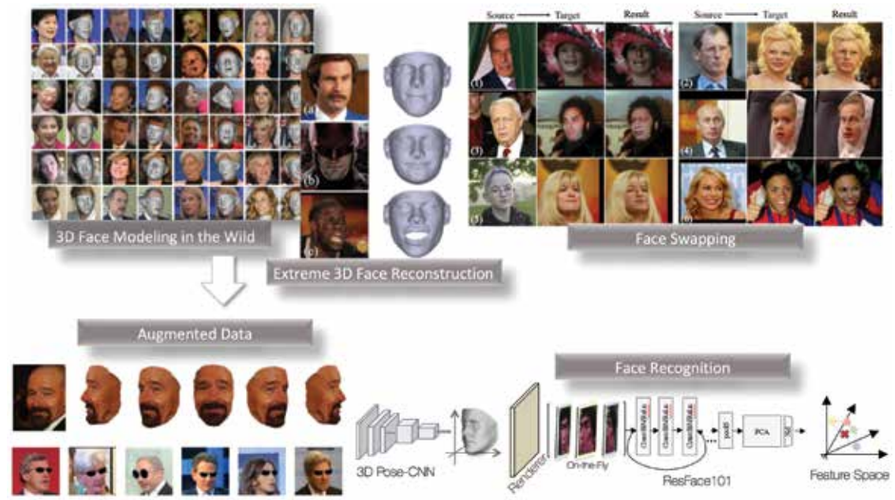
In today's social media rich world, natural or man-made disasters are typically followed by a torrent of tweets and news in the local language which are rich in information that is critical for planning and executing an effective response. While significant advances have been made in machine translation and information extraction in recent years, those advances are predicated on availability of large amounts of training data which is typically only available for languages prevalent in economically advanced regions of the world. The overwhelming majority of the world's 4000+ languages suffer from a lack of automated tools for processing them. Under DARPA's LORELEI program, the $9.6M ELISA project is developing technology for rapidly constructing information extraction, machine translation and topic and sentiment processing capabilities for new languages using limited amounts of training data. The goal of the project is to develop technology that will allow us to quickly field NLP tools for a new language of interest and thereby enable better planned and more effective humanitarian relief operations. The ELISA team includes researchers from ISI, Rensselaer Polytechnic Institute, and ICSI/Berkeley.

GAIA (Generating Alternatives for Interpretation and Analysis)

Initiated in the spring of 2018, the GAIA project has the ambitious agenda of extracting information from multimodal, multilingual data sources, synthesizing richly connected knowledge graphs from the individual information elements, and automatically identifying interesting, scenario-relevant hypotheses for human analysts to consider. Potential data sources include tweets, broadcast news, web videos, podcasts, etc. The $15M GAIA project is led by ISI and includes researchers from Rensselaer Polytechnic Institute, Columbia University, and University of Florida.

GLAIVE (Graphics-Based Learning Approach Integrated with Vision Elements)

Detecting and recognizing faces in diverse collections of consumer quality videos (e.g., those captured by a smartphone) is a challenging problem. Commercially developed systems rely on massive amounts of training data, ranging from tens to hundreds of millions of examples. A significant innovation from the GLAIVE team (sponsored by IARPA at $4.4M) is the development of a graphics-inspired data augmentation approach in which 3D face models for each subject are estimated using available sample images and are then used to generate millions of additional samples for previously unseen poses, occlusions, etc. The 3D model based approach significantly reduces the cost of data acquisition while providing a scalable framework for improving system accuracy as the number of subjects increases.



SARAL (Summarization and Domain-Adaptive Retrieval Across Languages)

Research on the SARAL project is focused on developing domain-adaptive machine translation, cross-lingual information retrieval, and summarization techniques. The SARAL team's approach incorporates a novel hierarchical, multi-lingual embedding technique that provides a unifying, common scaffolding for translation and retrieval algorithms. The $16.8M SARAL project is sponsored by IARPA and is focused on addressing a critical need for automated solutions that can identify information of interest across a multiplicity of domains, languages, and scenarios. The team includes researchers from ISI, University of Massachusetts at Amherst, Northeastern University, Rensselaer Polytechnic Institute, MIT, Idiap Research Institute (Switzerland), and the University of Notre Dame.

COMPUTATIONAL SYSTEMS AND TECHNOLOGY

A Characterization of Workflow Management Systems for Extreme-Scale Applications
Rafael Ferreira da Silva, **Rosa Filgueira**, Ilia Pietri, Ming Jiang, Rizos Sakellariou, and **Ewa Deelman**
Future Generation Computer Systems, vol. 75, 2017

A Comparison of System Performance on a Private OpenStac, Cloud and Amazon EC2
Mikyung Kang, Dong-In Kang, John Paul Walters, and **Stephen Crago**
International Conference on Cloud Computing (Cloud), 2017

Accelerating Circuit Realization via a Collaborative Gateway of Innovations
I. Taylor, A. Brinckman, **Ewa Deelman**, **Rafael Ferreira da Silva**, et al.
9th International Workshop on Science Gateways (IWSG 2017*)*, 2017

Advantages of Unfair Quantum Ground-State Sampling
Brian Zhang, Gene Wagenbreth, Victor Martin-Mayor and **Itay Hen**
Scientific Reports 7, 2017

All-Sky Search for Periodic Gravitational Waves in the O1 LIGO Data
Ewa Deelman (contributing as part of LIGO Scientific Collaboration and Virgo Collaborations)
Physical Review D, 2017

Architectural Refinements for Enhancing Trust and Securing Cyber-Physical Systems
Vivek Venugopalan and C. Patterson
IEEE International Conference on Advanced and Trusted Computing (ATC), 2017

Distributed Workflows for Modeling Experimental Data
V. Lynch, J. Calvo, **Ewa Deelman**, **Rafael Ferreira da Silva**, M. Goswami, Y. Hui, et al.
IEEE High Performance Extreme Computing Conference, 2017

Dynamically Improving Resiliency to Timing Errors for Stream Processing Workloads
Geoffrey Tran, John Paul Walters, and **Stephen Crago**
International Conference on Parallel & Distributed Computing, Applications, & Technologies, 2017

Evaluating Rapid Application Development with Python for Heterogeneous Processor-Based FPGAs
Andrew Schmidt, **Gabriel Weisz**, and **Matthew French**
Proceedings of the International Symposium on Field Programmable and Custom Computing Machines, 2017
(Best Short Paper Award)

First Low-Frequency Einstein@ Home All-Sky Search for Continuous Gravitational Waves in Advanced LIGO Data
Ewa Deelman (contributing as part of the LIGO Scientific Collaboration and Virgo Collaborations)
Physical Review, 2017

GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2
Ewa Deelman (contributing as part of LIGO Scientific Collaboration and Virgo Collaborations)
Physical Review Letters, 2017

Load Balancing for Minimizing Deadline Misses and Total Runtime for Connected Car Systems in Fog Computing
Yu-An Chen, John Paul Walters, and **Stephen Crago**
International Symposium on Parallel and Distributed Processing with Applications (ISPA), 2017

Off-Diagonal Expansion Quantum Monte Carlo
Tameem Albash, Gene Wagenbreth and **Itay Hen**
Phys. Rev. E 96, 063309, 2017

On the Use of Burst Buffers for Accelerating Data-Intensive Scientific Workflows
Rafael Ferreira da Silva, Scott Callaghan, and **Ewa Deelman**
12th Workshop on Workflows in Support of Large-Scale Science (WORKS'17), 2017

OSG-KINC: High-Throughput Gene Co-expression Network Construction Using the Open Science Grid
William Poehlman, **Mats Rynge**, Frank Feltus, et al.
IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2017

PANORAMA: An Approach to Performance Modeling and Diagnosis of Extreme Scale Workflows
Ewa Deelman, Christopher Carothers, Anirban Mandal, et al.
International Journal of High Performance Computing Applications, vol. 31, 2017

Quantum Annealing Correction at Finite Temperature: Ferromagnetic *p*-Spin Models
Shunji Matsuura, Hidetoshi Nishimori, Walter Vinci, **Tameem Albash** and Daniel Lidar
Phys. Rev. A95, 2017

Radiation Hardening by Software Techniques on FPGAs: Flight Experiment Evaluation and Results
Andrew Schmidt, Matthew French and **Thomas Flatley**
Proceedings of the IEEE Aerospace Conference, 2017

Realizable Quantum Adiabatic Search
Itay Hen
Europhysics Letters 118, 2017

Relaxation vs. Adiabatic Quantum Steady State Preparation
Lorenzo Venuti, **Tameem Albash,** Milad Marvian, Daniel Lidar, and Paolo Zanardi
Phys. Rev. A 95, 2017

Reproducibility of Execution Environments in Computational Science Using Semantics and Clouds
Santana-Perez, **Rafael Ferreira da Silva**, **Mats Rynge**, **Ewa Deelman**, et al.
Future Generation Computer Systems, vol. 67, 2017

rvGAHP – Push-Based Job Submission Using Reverse SSH Connections
Scott Callaghan, G. Juve, K. Vahi, P. Maechling, T. Jordan, and **Ewa Deelman**
12th Workshop on Workflows in Support of Large-Scale Science, 2017

Search for Gravitational Waves from Scorpius X-1 in the First Advanced LIGO Observing Run with a Hidden Markov Model
Ewa Deelman (contributing as part of the LIGO Scientific Collaboration and Virgo Collaborations)
Physical Review D, 2017

Search for Intermediate Mass Black Hole Binaries in the First Observing Run of Advanced LIGO
Ewa Deelman (contributing as part of the LIGO Scientific Collaboration and Virgo Collaborations)
Physical Review D 2017

Solving Spin Glasses with Optimized Trees of Clustered Spins
Itay Hen
Phys. Rev. E 96, 2017

SpaceCubeX: A Framework for Evaluating Hybrid Multi-Core CPU/FPGA/DSP Architectures
Andrew Schmidt, Gabriel Weisz, and **Matthew French**
Proceedings of the IEEE Aerospace Conference, 2017

Temperature Scaling Law for Quantum Annealing Optimizers
Tameem Albash, Victor Martin-Mayor and **Itay Hen**
Phys. Rev. Lett. 119, 2017

The Future of Scientific Workflows
Ewa Deelman, Tom Peterka, Ilkay Altintas, et al.
International Journal of High Performance Computing Applications, vol. 32, 2017

The OSG Open Facility: An On-Ramp for Opportunistic Scientific Computing
B. Jayatilaka, T. Levshina, C. Sehgal, R. Gardner, **Mats Rynge**, and F. Wurthwei
Journal of Physics: Conference Series, vol. 898, 2017

Thermalization, Freeze-Out and Noise: Deciphering Experimental Quantum Annealers
Jeffrey Marshall, Eleanor Rieffel and **Itay Hen**
Phys. Rev. Applied 8, 2017

Toward Prioritization of Data Flows for Scientific Workflows Using Virtual Software Defined Exchanges
Anirban Mandal, P. Ruth, I. Baldin, **Rafael Ferreira da Silva**, and **Ewa Deelman**
First International Workshop on Workflow Science (WoWS 2017*)*, 2017

Upper Limits on Gravitational Waves from Scorpius X-1 from a Model-Based Cross-Correlation Search in Advanced LIGO Data
Ewa Deelman (contributing as part of the LIGO Scientific Collaboration and Virgo Collaborations)
The Astrophysical Journal 847.1, 2017

ARTIFICIAL INTELLIGENCE

Abstract, Link, Publish, Exploit: An End-to-End Framework for Workflow Sharing
Daniel Garijo, **Yolanda Gil**, and Oscar Corcho
Future Generation Computer Systems, 75, 2017

Abstraction Improving Publication and Reproducibility of Computational Experiments through Workflow
Yolanda Gil, **Daniel Garijo**, **Margaret Knoblock**, **Alyssa Deng**, Ravali Adusumilli, **Varun Ratnakar**, et al.
Proceedings of the Workshop on Capturing Scientific Knowledge (SciKnow)
(in conjunction with ACM International Conference on Knowledge Capture) 2017

Adaptive Candidate Generation for Scalable Edge-discovery Tasks on Data Graphs
Mayank Kejriwal
MLG Workshop at ACM KDD, 2017

A Distributed Logical Filter for Connected Row Convex Constraints
Satish KumarHoon Xu, Zheng Tang, **Anoop Kumar**, Craig Rogers, and **Craig Knoblock**
Proceedings of 29th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), 2017

A Controlled Crowdsourcing Approach for Practical Ontology Extensions and Metadata Annotation
Yolanda Gil, **Daniel Garijo**, **Varun Ratnakar**, **Deborah Khider**, Julien Emile-Geay, and Nicholas McKay
Proceedings of the Sixteenth International Semantic Web Conference (ISWC), 2017

A Neural Named Entity Recognition Approach to Biological Entity Identification
Emily Sheng, **Scott Miller**, **José Luis Ambite**, and **Prem Natarajan**
BioCreative VI Challenge and Workshop, 2017

A Scalable Data Integration and Analysis Architecture for Sensor Data of Pediatric Asthma
Dimitris Stripelis, **José Luis Ambite**, **Yao-Yi Chiang**, Sandrah Eckel, and Rima Habre
IEEE International Conference on Data Engineering (ICDE), 2017 (Demonstration)

Adversarial Auto-Encoders For Speech-Based Emotion Recognition
Saurabh Sahu, **Rahul Gupta**, Ganesh Sivaraman, Carol Espy-Wilson, and **Wael AbdAlmageed**
InterSpeech, 2017

AI Buzzwords Explained: Scientific Workflows
Daniel Garijo
AI Matters, 3(1), 2017

Analyzing Uber's Ride-Sharing Economy
Farshad Kooti, Mihajlo Grbovic, Luca Maria Aiello, Nemanja Djuric, Vladan Radosavljevic, and **Kristina Lerman**
26th International World Wide Web Conference, 2017

An Investigative Search Engine for the Human Trafficking Domain
Mayank Kejriwal and **Pedro Szekely**
International Semantic Web Conference (In-Use Track), 2017

An Investigation into the Pedagogical Features of Documents
Emily Sheng, **Prem Natarajan**, **Jonathan Gordon**, and **Gully Burns**
Proceedings 12th Workshop on Innovative Use of NLP for Building Educational Applications, 2017

Automatic Alignment of Geographic Features in Contemporary Vector Data and Historical Maps
Weiwei Duan, Yao Chiang, **Craig Knoblock,** Vinl Jain, Dan Feldman, Johannes Uhl, and Stefan Leyk
1st ACM SIGSPATIAL Workshop on Artificial Intelligence and Deep Learning for Geographic Knowledge Discovery, 2017

Automatic Spatio-Temporal Indexing to Integrate and Analyze the Data of an Organization
Craig Knoblock, Aparna Joshi, Abhishek Megotia, Minh Pham, and Chelsea Ursaner
International Workshop on Smart Cities and Urban Analytics, 2017

Automated Data Narratives
Yolanda Gil and **Daniel Garijo**
Proceedings of the Twenty-Second ACM International Conference on Intelligent User Interfaces, 2017

BD2K ERuDIte: The Educational Resource Discovery Index for Data Science
José Luis Ambite, **Kristina Lerman**, **Lily Fierro**, Florian Geigl, **Jonathan Gordon**, and **Gully Burns**
4th WWW Workshop on Big Scholarly Data: Towards the Web of Scholars, 2017

Biomedical Event Extraction Using Abstract Meaning Representation
Sudha Rao, **Daniel Marcu**, **Kevin Knight**, and Hal Daume
Proceedings BioNLP Workshop, ACL, 2017

Combining Convolutional Neural Networks and LSTMs for Segmentation-Free OCR
Stephen Rawls, **Huaigu Cao,** Senthil Kumar, and **Premkumar Natarajan**
IAPR International Conference on Document Analysis and Recognition (ICDAR), 2017

Combining Deep Learning and Language Modeling for Segmentation-Free OCR From Raw Pixels
Stephen Rawls, **Huaigu Cao**, **Ekraam Sabir**, and **Premkumar Natarajan**
1st International Workshop on Arabic Script Analysis and Recognition (ASAR), 2017.

Contagion Dynamics of Extremist Propaganda in Social Networks
Emilio Ferrara
Information Sciences, 2017

Cross-lingual Name Tagging and Linking for 282 Languages
Xiaoman Pan, Boliang Zhang, **Jonathan May**, Joel Nothman, **Kevin Knight** and Heng Ji
Proceedings ACL, 2017

Deciphering Related Languages
Nima Pourdamghani and **Kevin Knight**
Proceedings EMNLP, 2017

Deep Context: A Neural Language Model for Large-Scale Networked Documents
Hao Wu and **Kristina Lerman**
Proceedings of International Joint Conference on AI, 2017

Deep Matching and Validation Network: An End-to-End Solution to Constrained Image Splicing Localization and Detection
Yue Wu, **Wael AbdAlmageed**, and **Premkumar Natarajan**
Proceedings of the ACM on Multimedia Conference, 2017

Designing Hyperchaotic Cat Maps With Any Desired Number of Positive Lyapunov Exponents
Z. Hua, S. Yi, Y. Zhou, C. Li, and **Yue Wu**
IEEE Transactions on Cybernetics, 2017

Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election
Emilio Ferrara
First Monday 22(8), 2017

Dynamics of Content Quality in Collaborative Knowledge Production
Emilio Ferrara, **Nazarin Alipoufard**, **Keith Burghard**t, Chiranth Gopal, and **Kristina Lerman**
Proceedings of 11th AAA International Conference on Web and Social Media, 2017

Early Detection of Promoted Campaigns on Social Media
Onur Varol, Emilio Ferrara, Fillippo Menczer, and **Allesandro Flammini**
EPJ Data Science 6(13), 2017

Early Warnings of Cyber Threats in Online Discussions
Anna Sapienza, Allesandro Bessi, Kristina Lerman, Saranya Damodaran, Paulo Shakarian, and **Emilio Ferrara**
IEEE International Conference on Data Mining Workshops (ICDMW), 2017

E-Cigarette Surveillance With Social Media Data: Social Bots, Emerging Topics, and Trends
Jon-Patrick Allem, **Emilio Ferrara, Sree Uppu, Tess Cruz,** and **Jennifer Unger**
JMIR Public Health and Surveillance, 2017

Effort Mediates Access to Information in Online Social Networks
Jeon Kang and **Kristina Lerman**
ACM Transactions on the Web, 11(1): 3:1, 2017

Engineering Academic Software: Dagstuhl Perspectives Workshop 16252
Alice Allen, Cecilia Aragon, Christoph Becker, Jeffrey Carver, Andrei Chis, Benoit Combemale, Mike Croucher, Kevin Crowston, **Daniel Garijo**, et al.
Dagstuhl Manifestos, 6(1), 2017

EPAT: Euclidean Perturbation Analysis and Transform-An Agnostic Data Adaptation Framework for Improving Facial Landmark Detectors
Yue Wu, **Wael AbdAlmageed**, **Stephen Rawls**, and **Premkumar Natarajan**
12th IEEE International Conference on Automatic Face & Gesture Recognition, 2017

Evaluation of the OntoSoft Ontology for Describing Legacy Hydrologic Modeling Software
Bakinam Essawy, Jonathan Goodall, Hao Xu, Mohamed Morsy, and **Yolanda Gil**
Environmental Modeling Software, 92, 2017

Evidence of Complex Contagion of Information in Social Media: An Experiment Using Twitter Bots
Bjarke Monsted, Piotr Sapiezynski, **Emilio Ferrara,** and Sune Lehmann
Plos One, 12(9), 2017

Experiment Segmentation in Scientific Discourse as Clause-level Structured Prediction Using Recurrent Neural Networks
Predeep Dasigi, **Gully Burns**, Eduard Hovy, and Anita de Waard
CoRR abs/1702.05398. 2017

Extracting Evidence Fragments for Distant Supervision of Molecular Interactions
Gully Burns Pradeep Dasigi, and Eduard Hovy
Proceedings of the First Workshop on Enabling Open Semantic Science (SemSci), 2017

Extracting Human Settlement Footprint from Historical Topograpic Map Series Using Context-Based Machine Learning
Johannes Uhl, Stefan Leyk, Yao Chiang, Weiwei Duan, and **Craig Knoblock**
8th International Conference on Pattern Recognition Systems Vol 8, 2017

Facial Landmark Detection with Tweaked Convolutional Neural Networks
Yue Wu, **Tal Hassner**, KangGeon Kim, Gerad Medioni, and **Premkumar Natarajan**
IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017

Fast Detection of Identity-by-Descent Tracts for Biobank-Scale Inference
Ruhollah Shemirani, G. Belbin, E. Kenny, C. Gignoux, and **José Luis Ambite**
67th Annual Meeting of the American Society of Human Genetics, 2017 (*Abstract + Poster*)

FlagIt: A System for Minimally Supervised Human Trafficking Indicator Mining
Mayank Kejriwal, Jiayuan Ding, Runqi Shao, Anoop Kumar, and **Pedro Szekely**
Workshop on Learning with Limited Labeled Data (LLD) at NIPS, 2017

Graph Filters and the Z-Laplacian
Xiaoran Yan, Brian Sadler, Robert Drost, Paul Yu, and **Kristina Lerman**
IEEE Journal of Selected Topics in Signal Processing 11 (6), 2017

Hafez: an Interactive Poetry Generation System (Best Demo Award)
Marjan Ghazvininejad, **Xing Shi**, **Jay Priyadarshi** and **Kevin Knight**
Proceedings ACL Demo Track, 2017

Humans Outperform Machines at the Bilingual Shannon Game
Marjan Ghazvininejad and **Kevin Knight**
Entropy 19(1), 2017

Implicit Language Model in LSTM for OCR
Ekraam Sabir, **Stephen Rawls**, **Premkumar Natarajan**
International Conference on Document Analysis and Recognition (ICDAR), 2017

Identifying Sentiment of Hookah-Related Posts on Twitter
J. Allem, J. Ramanujam, **Kristina Lerman**, et al.
JMIR Public Health and Surveillance 3 (4), 2017

Incident-Driven Machine Translation and Name Tagging for Low-Resource Languages
Ulf Hermjakob, Quang Li, **Daniel Marcu**, **Jonathan May**, Sebastian Mielke, **Nima Pourdamghani**, **Michael Pust**, Xing Shi, **Kevan Knight**, et al.
Machine Translation, 2017

Information Extraction in Illicit Domains
Mayank Kejriwal and **Pedro Szekely**
ACM World Wide Web Conference, 2017

iPhone Digital Marketplace: Characterizing the Big Spenders
Farshad Kooti, Mihajlo Grbovic, Luca Maria Aiello, Eric Bax, and **Kristina Lerman**
10th International ACM Conference on Web Search and Data Mining, 2017

Knowledge Graphs for Social Good: An Entity-centric Search Engine for the Human Trafficking Domain
Mayank Kejriwal and **Pedro Szekely**
IEEE Transactions on Big Data, Special Call on Knowledge Graphs, 2017

Language, Demographics, and Emotions Shape Online Social Networks
Kristina Lerman, Luciano Marin, Megha Arora, Lucas Costa de Lima, Emilio Ferrara, and **David Garcia**
Journal of Computational Social Science, 2017

Learning with Previously Unseen Features
Yuan Shi and **Craig Knoblock**
Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI), 2017

Learning the Semantics of Structured Data Sources
Mohsen Taheriyan, **Craig Knoblock**, **Pedro Szekely**, and **José Luis Ambite**
Web Semantics: Science, Services and Agents on the World Wide Web, 2016 (not previously reported)

Lessons Learned in Building Linked Data for the American Art Collaborative
Craig Knoblock, **Pedro Szekely**, Eleanor Fink, Duane Degler, et al.
16th International Semantic Web Conference (ISWC), 2017

Liberal Entity Extraction: Rapid Construction of Fine-Grained Entity Typing Systems
Lifu Huang, **Jonathan May**, Xiaoman Pan, Heng Ji, Xiang Ran, et al.
Big Data, Volume 5, 2017

Modified Cheeger and Ratio Cut Methods Using the Ginzburg–Landau Functional for Classification of High-Dimensional Data
Ekaterina Merkurjev, Andrea Bertozzi, Xiaoran Yan, and **Kristina Lerman**
Inverse Problems 33 (7), 2017

Mining Public Datasets for Modeling Intra-City PM2.5 Concentrations at a Fine Spatial Resolution
Yijun Lin, Fan Pan, **Yao-Yi Chiang**, **Dimitris Stripelis**, **José Luis Ambite**, Sandrah Eckel, and Rima Habre
Proceedings of the 25rd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2017

Multi-Lingual Extraction and Integration of Entities, Relations, Events and Sentiments into ColdStart++ KBs with the SAFT System
Hans Chalupsky, Jun Araki, Eduard Hovy, Andrew Hsi, Zhengzhong Liu, et al.

Multimedia Semantic Integrity Assessment Using Joint Embedding of Images and Text
Ayush Jaiswal, **Ekraam Sabir**, **Wael AbdAlmageed**, and **Premkumar Natarajan**
Proceedings of the ACM on Multimedia Conference, 2017
Proceedings of Text Analysis Conference, 2017

Neural Embeddings for Populated Geonames Locations
Mayank Kejriwal and **Pedro Szekely**
International Semantic Web Conference (Resource Track), 2017

Neighbor-Neighbor Correlations Explain Measurement Bias in Networks
Xin-Zeng Wu, Allon Percus, and **Kristina Lerman**
Scientific Reports, 7(5576), 2017

NiW: Converting Notebooks into Workflows to Capture Dataflow and Provenance
Lucas Carvalho, Regina Wang, **Yolanda Gil**, and **Daniel Garijo**
Proceedings of Workshop on Capturing Scientific Knowledge (SciKnow)
(in conjunction with ACM International Conference on Knowledge Capture), 2017

Online Human-Bot Interactions: Detection, Estimation, and Characterization
Onur Varol, Emilio Ferrara, Clayton Davis, Fillippo Menczer, and **Allesandro Flammini**
Proceedings of the Eleventh International AAAI Conference on Web and Social Media, 2017

On Quitting: Performance and Practice in Online Game Play
Tushar Agarwal, **Keith Burghardt**, and **Kristina Lerman**
Proceedings of 11th AAAI International Conference on Web and Social Media, 2017

Performance Dynamics and Success in Online Games
Anna Sapienza, Hao Peng, and **Emilio Ferrara**
2017 *IEEE International Conference on Data Mining Workshops (ICDMW)*, 2017

Performance Dynamics in Collaborative Knowledge Production Systems.
Emilio Ferrara, Nazanin Alipourfard, Keith Burghardt, Chiranth Gopal, and **Kristina Lerman**
Proceedings of the Eleventh International AAAI Conference on Web and Social Media, 2017

Predicting Role Relevance with Minimal Domain Expertise in a Financial Domain
Mayank Kejriwal
DSMM Workshop at ACM SIGMOD, 2017

Requirements for Supporting the Iterative Exploration of Scientific Workflow Variants
Lucas Carvalho, Bakinam Essawy, **Daniel Garijo**, Claudia Bauzer Medeiros, and **Yolanda Gil**
Proceedings of the Workshop on Capturing Scientific Knowledge (SciKnow)
(in conjunction with tACM International Conference on Knowledge Capture), 2017

Scalable Generation of Type Embeddings Using the ABox
Mayank Kejriwal and **Pedro Szekely**
Open Journal of Semantic Web, 2017

Scientific Workflows in Data Analysis: Bridging Expertise Across Multiple Domains
Ricky Sethi and **Yolanda Gil**
Future Generation Computer Systems, *75*, 2017

Self-organized Text Detection with Minimal Post-processing via Border Learning
Yue Wu, and **Premkumar Natarajan**
Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017

Semantic Modeling for Accelerated Immune Epitope Database (IEDB) Biocuration
Gully Burns, Randi Vita, James Overton, Ward Fleri, and Bjoern Peters
Second International Workshop on Capturing Scientific Knowledge, 2017

SemEval-2017 Task 9: Abstract Meaning Representation Parsing and Generation
Jonathan May and **Jay Priyadarshi**
Proceedings SemEval, 2017

Sequential Heterogeneous Attribute Embedding for Item Recommendation
Kuan Liu, **Xing Shi**, **Premkumar Natarajan**
IEEE International Conference on Data Mining Workshops (ICDMW), 2017

Speeding Up Neural Machine Translation Decoding by Shrinking Run-Time Vocabulary
Xing Shi and **Kevin Knight**
Proceedings ACL, 2017

Structured Generation of Technical Reading Lists
Jonathan Gordon, Stephen Aguilar, **Emily Sheng**, and **Gully Burns**
Proceedings 12th Workshop on Innovative Use of NLP for Building Educational Applications, 2017

Supervised Typing of Big Graphs using Semantic Embeddings
Mayank Kejriwal and **Pedro Szekely**
Semantic Big Data (SBD) Workshop at ACM SIGMOD, 2017

Taming the Unpredictability of Cultural Markets with Social Influence
Andres Abeliuk, Gerardo Berbeglia, Pascal Van Hentenryck, Tad Hogg, and **Kristina Lerman**
26th International World Wide Web Conference, 2017

Team ELISA System for DARPA LORELEI Speech Evaluation 2016
Pavlos Papadopoulos, Ruchir Travadi, Colin Vaz, Nikolaos Malandrakis, **Ulf Hermjakob**, **Nima Pourdamghani, Michael Pust**, Boliang Zhang, Xiaoman Pan, Di Lu, Ying Lin, Ondrej Glembek, Murali BaskarL Lukas Burget, Mark Hasegawa-Johnson, Heng Ji, **Jonathan May**, **Kevin Knight**, and **Shrikanth Narayanan**
Proceedings Interspeech, 2017

The BD2K Training Coordinating Center's Data Science Education Platform
José Luis Ambite , Lily Fierro, Florian Geigl, **Jonathan Gordon**, **Gully Burns**, **Kristina Lerman**, et al.
American Medical Informatics Association Annual Symposium, 2017

The DISK Hypothesis Ontology: Capturing Hypothesis Evolution for Automated Discover
Daniel Garijo, **Yolanda Gil**, and **Varun Ratnakar**
Proceedings of Workshop on Capturing Scientific Knowledge (SciKnow)
(in conjunction with ACM International Conference on Knowledge Capture), 2017

The Myopia of Crowds: Cognitive Load and Collective Evaluation of Answers on Stack Exchange
Keith Burghardt, Emanuel Alsina, Michelle Girvan, William Rand, and **Kristina Lerman**
PLOS ONE, *12(3):* e0173610, 2017

Thoughtful Artificial Intelligence: Forging A New Partnership for Data Science and Scientific Discovery
Yolanda Gil
Data Science, *1*, 2017

Towards Automatic Generation of Portions of Scientific Papers for Large Multi-Institutional Collaborations Based on Semantic Metadata
MiHyun Jang, Tejal Patted, **Yolanda Gil**, **Daniel Garijo**, **Varun Ratnakar**, Jie Ji, Prince Wang, et al.
Proceedings of the Workshop on Enabling Open Semantic Science
(co-located with Sixteenth International Semantic Web Conference (ISWC), 2017

Towards Continuous Scientific Data Analysis and Hypothesis Evolution
Yolanda Gil, **Daniel Garijo**, **Varun Ratnakar**, Rajiv Mayani, Ravali Adusumilli, Hunter Boyce, et al.
Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), 2017

Travel Analytics: Understanding How Destination Choice and Business Clusters are Connected Based on Social Media Data
Arthur Huang, Luciano Gallegos, and **Kristina Lerman**
Transportation Research Part C: Emerging Technologies, *77: 245*, 2017

Understanding Short-term Changes in Online Activity Sessions
Farshad Kooti, Karthik Subbian, Winter Mason, Lada Adamic, and **Kristina Lerman**
26th International World Wide Web Conference, 2017

Using Contexts and Constraints for Improved Geotagging of Human Trafficking Webpages
Rahul Kapoor, Mayank Kejriwal, and **Pedro Szekely**
GeoRich Workshop at ACM SIGMOD, 2017

1990 US Census Form Recognition Using CTC Network, WFST Language Model, and Surname Correction
Huaigu Cao, **Stephen Rawls**, **Premkumar Natarajan**
International Conference on Document Analysis and Recognition (ICDAR), 2017

Using Convolutional Encoder-Decoder for Document Image Binarization
Xujun Peng, **Huaigu Cao**, **Premkumar Natarajan**
International Conference on Document Analysis and Recognition (ICDAR), 2017

Web Text-Based Network Industry Classifications: Preliminary Results
Eric Heiden, Gerald Hoberg, **Craig Knoblock**, and **Pedro Szekely**
Proceedings of DSMM: Data Science for Macro-Modeling with Financial and Economic Datasets, 2017

WIDOCO: A Wizard for Documenting Ontologies
Daniel Garijo
Proceedings of the International Semantic Web Conference (ISWC), 2017

NETWORKING AND CYBERSECURITY

A Concurrency-Optimal Binary Search Tree
Vitaly Aksenov, Vincent Gramoli, Petr Kuznetsov, Anna Malova, and **Srivatsan Ravi**
Proceedings Euro-Par 2017*: Parallel Processing, 23rd International Conference on Parallel and Distributed Computing*

Anycast Latency: How Many Sites Are Enough?
Ricardo de O. Schmidt, **John Heidemann** and Jan Harm Kuipers
Proceedings of the Passive and Active Measurement Workshop, 2017 (to appear)

Commoner Privacy and a Study on Network Traces
Xiyue Deng and **Jelena Mirkovic**
Proceedings of 2017 *Annual Computer Security Applications Conference*, 2017

Concurrency and Privacy with Payment-Channel Networks
Giulio Malavolta, Pedro Moreno-Sanchez, Aniket Kate, Matteo Maffei, and **Srivatsan Ravi**
Proceedings of ACM SIGSAC Conference on Computer and Communications Security, 2017

Cost of Concurrency in Hybrid Transactional Memory
Trevor Brown and **Srivatsan Ravi**
31st International Symposium on Distributed Computing, 2017

Detecting ICMP Rate Limiting in the Internet
Hang Guo and **John Heidemann**
Technical Report ISI-TR-717. USC/Information Sciences Institute, 2017

Detecting Malicious Activity with DNS Backscatter Over Time
Kensuke Fukuda, **John Heidemann** and Abdul Qadeer
ACM/IEEE, 2017

Do You See Me Now? Sparsity in Passive Observations of Address Liveness
Jelena Mirkovic, **Genevieve Bartlett**, **John Heidemann**, **Hao Shi** and Xiyue Deng
Proceedings of Traffic Measurement and Analysis Conference (TMA), 2017

Does Anycast Hang up on You?
Lan Wei and **John Heidemann**
IEEE, 2017 *Technical Report ISI-TR-716. USC/Information Sciences Institute*, 2017

Enabling SDN Experimentation in Network Testbeds
Sivaramakrishnan Ramanathan, Pravein Kannan, Chan Mun Choon, **Jelena Mirkovic**, and Keith Sklower
Proceedings of ACM SDNNFV Security Workshop, 2017

Generalized Paxos Made Byzantine (and Less Complex)
Miguel Pires, **Srivatsan Ravi**, Rodrigo Rodrigues
Proceedings Stabilization, Safety, and Security of Distributed Systems: 19th International Symposium, 2017

Grasping the Gap Between Blocking and Non-blocking Transactional Memories
Petr Kuznetsov and **Srivatsan Ravi**
Journal of Parallel and Distrib. Computing, 2017

Handling Anti-Viral Techniques in Malicious Software
Hao Shi, **Jelena Mirkovic,** and Abdulla Alwabel
Proceedings of ACM SAC, 2017

Hiding Debuggers from Malware with Apate
Hao Shi and **Jelena Mirkovic**
Proceedings of CAN SAC, 2017

Inherent Limitations of Hybrid Transactional Memory
Dan Alistarh, Justin Kopinsky, Petr Kuznetsov, **Srivatsan Ravi,** and Nir Shavi
Distributed Computing, 2017

International Networking in Support of Extremely Large Astronomical Data-centric Operations
Jeronimo Bezerra, V. Arcanjo, Julio Ibarra, J. Kantor, R. Lambert, M. Kollross, A. Astudillo, S. Sobhani, D. Jaque, H. Petravick, **Heidi Morgan**, et al.
Astronomical Data Analysis Software and Systems (ADASS XXVII) Conference, 2017

LDplayer: DNS Experimentation at Scale
Liang Zhu and **John Heidemann**
Technical Report 722. USC/Information Sciences Institute, 2017
Proceedings of the SIGCOMM Posters and Demos, 2017 (*poster and abstract*)

Lower Bounds for Transactional Memory
Srivatsan Ravi
Bulletin of the EATCS, 2017

Mitigating the Risks of Supporting Multiple Control Plans in a Production SDN Network: A Use Case
Jeronimo Bezerra, Julio Ibarra, M. Schwarz, H. Freitas, and **Heidi Morgan**
SBRC/WPEIF, 2017

Modeling Aggregate Security with User Agents that Employ Password Memorization Techniques
Christopher Novak, **James Blythe**, Ross Koppel, Vijay Kothari, and Sean Smith
Who Are You?! Adventures in Authentication: Symposium on Usable Privacy and Security, 2017

Novel Network Services for Supporting Big Data Science Research
Joaquin Chung, Sean Donovan, Jeronimo Bezerra, **Heidi Morgan**, Julio Ibarra, et al.
Gateways 2017*, University of Michigan, Ann Arbor*

Password Logbooks and What their Amazon Reviews Reveal About Their Users' Motivations, Beliefs, and Behaviors
Ross Koppel, **James Blythe**, Vijay Kothari, and Sean Smith
2nd European Workshop on Usable Security (EuroUSEC), 2017

Programmable Elasticity for Actor-Based Cloud Applications
Bo Sang, **Srivatsan Ravi**, Gustavo Petri, Mahsa Najafzadeh, et al,
Proceedings of the 9th Workshop on Programming Languages and Operating Systems, 2017

Recursives in the Wild: Engineering Authoritative DNS Servers
Moritz Müller, Giovane C. M. Moura, Ricardo de O. Schmidt and **John Heidemann**
Proceedings of the ACM Internet Measurement Conference, 2017

RESECT: Self-Learning Traffic Filters for IP Spoofing Defense
Jelena Mirkovic, **Erik Kline** and Peter Reiher
Proceedings of 2017 *Annual Computer Security Applications Conference*, 2017

Understanding Malware's Network Behaviors Using Fantasm
Xiyue Deng, **Hao Shi**, and **Jelena Mirkovic**
Proceedings of LASER Workshop, 2017

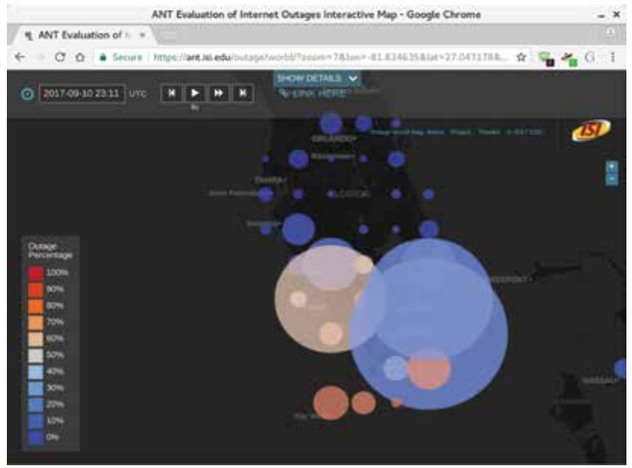
Verfploeter: Broad and Load-Aware Anycast Mapping
Wouter B. de Vries, Ricardo de O. Schmidt, **Wes Hardaker**, **John Heidemann**, Pieter-Tjerk de Boer and Aiko Pras
Proceedings of the ACM Internet Measurement Conference, 2017

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. Natarajan's 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, he 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 2017 two ISI researchers, John Heidemann and Wei-Min Shen, conducted their research under this grant.

JOHN HEIDEMANN | UNDERSTANDING INTERNET OUTAGES

The result of this work was the creation of a new website at https://ant.isi.edu/outage/world/ that supports viewing the Internet outage data that we collect. Our website makes exploring outage data more accessible to researchers and the public by making terabytes of collected Internet outage visible on a world map.



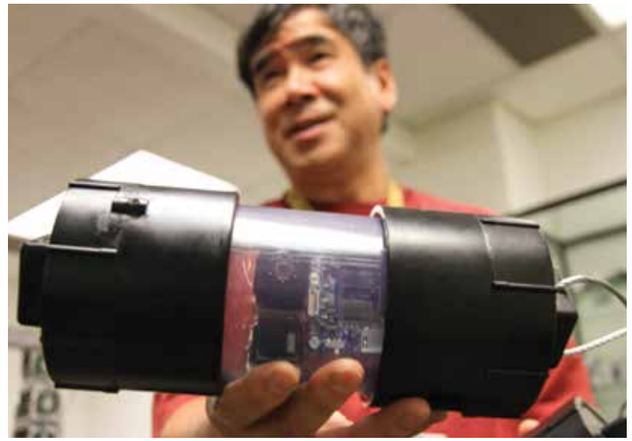
Our website supports browsing more than two years of outage data, organized by geography and time. The map is a google-maps-style world map, with circles on it at even intervals (every 0.5 to 2 degrees of latitude and longitude, depending on the zoom level). Circle sizes show how many /24 network blocks are out in that location; circle colors show the percentage of outages, from blue (only a few percent) to red (approaching 100%). The raw data underlying this website is available on request at https://ant.isi.edu/datasets/outage/index.html).

Our Internet outages website was developed by a collaborative team of ISI and USC researchers. In addition to the Keston grant, this research has also been funded by government agencies such as the Department of Homeland Security.

WEI-MIN SHEN | THE PIPEFISH PROJECT

Led by the researchers from ISI's Polymorphic Robotics Lab, the goal of this research is to develop and test autonomous robots with an inexpensive device that uses sensors to collect data in underground water pipes—thereby enabling us to assess conditions and detect problems. The project is also collaborating with the Los Angeles Department of Water and Power (LADWP).

Far beneath the streets and sidewalks of Los Angeles lies a rarely seen subterranean world—a labyrinthine network of underground pipes extending 7,200 miles across the city, carrying drinking water to more than four million people every day. Like many cities in the U. S., Los Angeles is facing a looming crisis over its aging water infrastructure, and fixing it will be a monumental and expensive task. At least two-thirds of the city's underground water pipes are more than 60 years old, and most will reach the end of their useful lives within 15 years. This can cause a host of problems—from burst pipes and loss of water service to road closures, sinkholes, and even potential water contamination.



Autonomous Pipefish robot with its creator.

PipeFish features a 360-degree camera, lights, sensors, and navigation technology, controlled by an onboard computer. It measures 20 inches to 30 inches in length and 3 inches to 6 inches in diameter. Constructed of plastic and other rigid materials, PipeFish is designed to be fully autonomous and untethered, so it can freely explore the complex network of underground water pipes.

PipeFish enters the water system through existing fire hydrants and follows the path of the pipeline to its final destination, where it is "caught" by a net. The footage captured is then uploaded and analyzed, providing operators with a 360-degree virtual reality interior view, without ever setting foot inside the pipe. Signs of damage inside the pipe, such as cracks, high corrosion or rust are indicators that can help authorities prioritize repairs, without expensive and disruptive excavation or water service interruptions. PipeFish is also a modularized robot; i.e., multiple PipeFish robots can form a chain to handle the complex twists and turns in an underground water network.

In 2017, Shen and his team conducted "dry tests" in pipes at the Los Angeles Department of Water and Power's Sylmar West Facility in the San Fernando Valley. In 2018, the team plans to add water to the system and test the robot in different pipes of various diameters under the streets of Los Angeles.

PipeFish could be equipped with additional sensors to collect more information—including water flow rate, gas space, illegally dumped chemicals and flammable materials. Ultimately, our team hopes that a "school" of PipeFish robots can be programmed to quickly and inexpensively traverse specific paths.

USC COMPUTING FORUM

Information Sciences Institute

USC Viterbi
School of Engineering

EXECUTIVE DIRECTOR/FOUNDER | **PREMKUMAR NATARAJAN**
DIRECTOR OF OPERATIONS | **LORI WEISS**

The USC Computing Forum brings together leading minds in computing—including academic and industry researchers, scientists, venture capitalists, founders of startups and business leaders—to advance innovation, promote thought-sharing, and drive the growth of entrepreneurial activity in Southern California and beyond. Charter members (funding sponsors) of the forum include Northrop-Grumman, Amazon, Lockheed Martin, California Resources Corp., and SAP.

With a focus on fostering interaction between academia, industry, and the U.S. Government, the forum seeks to foster passionate discussions, create networking opportunities, and promote lasting partnerships amongst its constituents. Forum symposia explore emerging trends and progress in cybersecurity, data science, artificial intelligence, quantum computing and other such fields of current and emerging interest. The forum also connects industry leaders with USC students—some of the best and brightest talent in the country—to exchange cutting-edge ideas and translate promising research into real-world innovation.

Founded by ISI's Michael Keston Executive Director Prem Natarajan and located in the heart of silicon beach with links to academia, government and industry, the Forum will continually evolve to meet the interests, needs, and aspirations of USC and our corporate partners.

In May 2018 ISI will host the USC Computing Forum's Inaugural Symposium themed "AI in Life." Speakers will include internationally known academics, business leaders and global thought leaders on artificial intelligence and machine learning, all of whom will share their perspectives on what it means to live and work in an increasingly AI-driven society.

For up-to-date news about the forum's activities, please visit http://usccf.isi.edu/

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