# Computer Science and Artificial Intelligence Laboratory

The Computer Science and Artificial Intelligence Laboratory (CSAIL) takes a computational approach to finding solutions to many of the most challenging problems of our lives, our work, and our world. We employ a long-term lens to engineer innovative solutions in an effort to unlock the secrets of human intelligence, extend the functional capabilities of machines, and explore human/machine interactions.

With approximately 50 research groups working on hundreds of diverse projects, CSAIL researchers focus on finding innovative ways to make systems and machines operate faster, better, safer, easier, and more efficiently for the benefit of humanity. Our projects fall into three areas of inquiry:

- Artificial intelligence—We seek to understand and develop both living and artificial systems capable of intelligent reasoning, perception, and behavior.
- Systems—We seek to discover new principles, models, metrics, and tools of both hardware- and software-based computer systems.
- Theory—We seek to understand the mathematics of computation and its wide-ranging, real-world consequences.

CSAIL has a long history of technological innovations that have impacted how people interact and do business. Current research explores cloud computing, mobile computing, the next generation of laptops, and the application of sensor technology to traffic congestion, medical monitoring, and climate observations. Robotic locomotion and the human-robotic interface are being investigated, as well as medical solutions with image-guided surgery and technical applications to aid clinical decisions. Advancements in biological research are also under way, including developments in the field of computational biology and the application of machine learning to the interpretation of complete genomes and understanding gene regulation.

CSAIL research is sponsored by a large number of diverse sources, from US government contracts to the private sector. US government sponsors include the Air Force (including the Air Force Research Laboratory and the Air Force Office of Scientific Research), the Army Research Office, the Defense Advanced Research Project Agency, Department of Defense Research and Engineering, the US Department of Education, the Department of Energy, the National Institutes of Health, the National Science Foundation (NSF), NASA, and the Navy (including the Office of Naval Research and Naval Air Systems Command). US and international industrial sponsors include Boeing; Clear Wireless LLC; Comcast; Cyberonics, Inc.; DuPont; Ford; Foxconn; IBM; Intel Corporation; Liberty Global, Inc.; Lockheed Martin Advanced Technology Laboratories; Microelectronics Advanced Research Corporation; Nippon Telegraph and Telephone Corporation; Nokia; Northrop Grumman Corporation; Pfizer, Inc.; Quanta Computer, Inc.; SAP; Shell; and Time Warner, Inc. Other organizations sponsoring research include Aarhus University, Defense Science and Technology Agency, Delta Electronics Foundation, DSO National Laboratories, Epoch Foundation, Industrial Technology Research Institute, and the MIT–Singapore Alliance.

## Research Projects

Within CSAIL we have many single- and multi-investigator projects as well as a number of virtual centers and large-scale mega projects. The large-scale projects and collaborations include the following:

### T-Party

Quanta Computer, Inc. extended their support of the T-Party project for an additional five years, raising the total funding to $45.5 million over 10 years. The focus has been expanded and includes four major research areas.

The first area concerns cloud computing technologies. Professor Anant Agarwal is working on new, scalable operating systems for clusters of multi-core computers; professor Frans Kaashoek is developing unmanaged internet architecture to provide easy-to-use, authenticated connectivity; professor Robert Morris is working on scalable data stores supporting high-traffic websites; professor Nickolai Zeldovich is researching security frameworks that limit data dissemination and allow recovery from attacks; professor Samuel Madden is developing new database architectures that increase concurrency of cloud-based implementations; and professor Srinivas Devadas is examining trusted storage built using untrusted commodity cloud storage.

The second research area revolves around human-computer interfaces for interactions between mobile devices and the cloud. Dr. Stephanie Seneff is examining the continued evolution of a framework for natural language understanding and dialogs; Dr. James Glass is developing an embeddable web-based interface for multimodal interactions including speech and gesture; and associate professor Robert Miller is working on creating easy-to-use interfaces using web automation.

The third area is telepresence. Professor Stephen Ward is developing initial prototypes of a presence capture system supporting multiple "presence streams"—video, audio, and annotations either generated in real-time (e.g., audience responses) or via post-production in the cloud (e.g., captions), which can be archived and flexibly recomposed for later playback; and professor John Guttag is researching medical telepresence incorporating multiple video, audio and telemetry channels between ambulance and hospital.

The final research area involves computational photography. Associate professor Frédo Durand is examining the many unique and effective algorithms to manipulate digital images and professor William Freeman is studying algorithms to extract information content or important features from images.

### Agile Robotics for Logistics: Developing a Voice-Commandable Robotic Forklift

MIT, in collaboration with BAE Systems, has formed a team of about 20 faculty, staff and students with expertise in a variety of areas including drive-by-wire actuation, sensing and perception for situational awareness, planning under uncertainty, agile control for mobility and manipulation; and natural human-computer interfaces and interaction to support military logistics operations such as supply chain management.

With funding from Defense Research and Engineering, and intensive engagement with the US Army, we are developing three prototype robotic platforms, each voice-commandable and able to operate in close proximity to people. The first is a robot forklift capable of locating, engaging, lifting, transporting, and depositing pallets throughout an outdoor warehouse facility, all while under voice and gestural command of a human supervisor. The second is a robot "porter," a humanoid robot with a mobile base, two articulated arms and grippers, and a head, capable of lifting, moving, and placing sub-pallet objects (such as boxes, sacks, and cans) on command. The third robot is a small rover that can follow a human through an outdoor environment, interpret a narrated guided tour given by a human, and share its interpretation with the other robots so that they may more easily find places and objects within the environment.

Key technical challenges and innovations in the project include achieving robust and effective situational awareness in a changing, semi-structured, outdoor environment with nearby humans and other vehicles; safe interoperation with humans, including seamless autonomy handoff and return; effective task-level multimodal voice and gesture interface; and safe autonomous engagement, transport, and placement of varying pallet loads over uneven outdoor terrain. The forklift embodies a novel technique called "apparent intent," in which it announces its imminent actions through a variety of visible and audible annunciators before actually acting. We think that this will be a key element in securing human acceptance of a multi-ton robot operating autonomously in their midst.

### Basic Research in Interactive Computing

In January 2008, CSAIL started a three-year collaboration with Foxconn for basic research encompassing several areas of computer science, ranging from networking, human–computer interactions, computer graphics and vision, to theory. Our research is predicated on the belief that computers and information devices are fast becoming interactive; they interact with other computers, with their environments, and, above all, with humans. Each form of interaction adds a new dimension to the challenge of modeling and understanding the behavior of computer systems as well as the task of building and using these systems.

During year two, this effort supported seven projects as shown below:

- Bit-switched wireless networks (professors Hari Balakrishnan and Robert Morris)
- Speech understanding in multimedia environments (Drs. Jim Glass and Stephanie Seneff)
- Computational imaging for future displays and cameras (professors Frédo Durand, Edward Adelson, and Bill Freeman)
- Scene and object recognition (professors Antonio Torralba and Bill Freeman)
- Real-time hand tracking as a user interface (professors Jovan Popovic and Randall Davis)
- Unsupervised audio-visual person identification (Dr. Jim Glass and professor Antonio Torralba)
- Socializing cell phones (professor Randall Davis)

In January 2010, the collaboration will enter into its third and final year and will continue with these projects. In the meantime, negotiations to continue the collaboration are ongoing.

A total of 20 Foxconn engineers in five shifts have visited CSAIL over a 26-month period to facilitate technology transfer and to receive training for advanced research and development in computer science. In addition, six principal investigators visited Foxconn in Taiwan in December 2009 to present preliminary research results and to interact with Foxconn senior management to better understand how our research can achieve maximum impact for our sponsor.

### Research in Cyber Security

In September 2009, CSAIL established a five-year research collaboration with Northrop Grumman Information Technology, Inc. (NGIT) in the research area of cyber security. In so doing, MIT joins Carnegie Mellon University and Purdue University as part of the NGIT Cyber Security Consortium.

During year one, the research collaboration supported three projects as shown below:

- Secure and dependable systems by design (professor Daniel Jackson)
- Secure audit trails (professor Barbara Liskov)
- Metacomputing for security, attribution, and trust (Drs. Thomas Knight, Jr and Howard Shrobe)

In February 2010, Dr. Shrobe curtailed his work as a principal investigator for the third project due to a leave from CSAIL to serve as a program manager at the Defense Advanced Research Projects Agency (DARPA).

### World Wide Web Consortium

The World Wide Web Consortium (W3C) was founded at MIT in 1994 by the inventor of the web, Tim Berners-Lee. W3C is responsible for developing and maintaining the standards that make the web work and ensure the long-term growth of the web. Currently, 360 member organizations, including most of the world's leading technology companies, are working to enhance the capabilities used within web documents and to transform today's web of linked documents to an expanded web of data and services across a wide range of devices enabling everyone on the planet to collaborate and share data and information.

W3C is the leading forum for the technical development and stewardship of the web. The W3C community creates open standards that define how the web works. W3C is hosted by MIT, the European Research Consortium for Informatics and Mathematics, and Keio University. W3C member organizations (including many of the world's leading technology companies), public, and staff are currently enhancing the capabilities of the web as a platform for applications with sophisticated interfaces, driven by integrated data and services, all based on open standards.

W3C's early leadership in linked data has resulted in completion of the basic semantic web standards. W3C recently announced standards for rule-interchange and ontology descriptions. W3C is working closely with diverse communities to help launch these technologies, while addressing interoperability challenges within domains such as e-government, health care, life sciences, and social networking.

W3C has made significant progress on interface and programming technologies, including HTML5, CSS3 (style), and SVG (graphics). These technologies and others—many being standardized at W3C—will bring more interactivity, video, location-based services, and more to end users.

W3C continues to promote international adoption of the second version of the web content accessibility guidelines (WCAG 2.0). These guidelines describe how to make web content and web applications accessible to people with disabilities. Because is essential that several different components of web development and interaction work together in order for the web to be accessible to people with disabilities, W3C''s Web Accessibility Initiative is also developing guidelines to ensure the accessibility of browsers, media players, authoring tools, and other software.

W3C''s Mobile Web Initiative continues to provide guidance on creating websites and web applications that work on mobile phones and other small devices. The success of web access through mobile devices has raised consumer expectations about ubiquitous network access. W3C is organizing a series of workshops related to the web on TV to bring the television industry and other producers of consumer electronics into the discussion about future web platforms. W3C is also working with partner organizations on how mobile devices can help bridge the digital divide and provide some services (for example in health, education, governance, and business) to rural communities and underserved populations in developing countries. Through its work on accessibility, internationalization, and mobile access, W3C works to ensure that the web is available to all.

## Research Highlights

In addition to the large-scale collaborative projects and center research, numerous individual and multi-investigator projects are under way. A sampling of the work is highlighted below.

### A Statistical Model for Lost Language Decipherment

Dozens of lost languages have been deciphered by humans in the last two centuries. In each case, the decipherment has been considered a major intellectual breakthrough, often the culmination of decades of scholarly effort. So far, computers have played no role in this enterprise, even for recently deciphered languages. Skeptics argue that computers do not have the "logic and intuition" required to unravel the mysteries of ancient scripts. Our recent work has demonstrated that much of this logic and intuition can be successfully captured by computational models, enabling the automation of the decipherment process.

Our definition of the computational decipherment task closely follows the setup typically faced by human decipherers. Our input consists of texts in a lost language and a corpus of non-parallel data in a known related language. The decipherment itself involves two related sub-tasks: finding the mapping between alphabets of the known and lost languages, and translating words in the lost language into corresponding cognates of the known language.

While there is no single formula that human decipherers have employed, manual efforts have focused on several guiding principles. A common starting point is to compare letter and word frequencies between the lost and known languages. In the presence of cognates the correct mapping between the languages will reveal similarities in frequency, both at the character and lexical level. In addition, morphological analysis plays a crucial role, as highly frequent morpheme correspondences can be particularly revealing. In fact, these three strands of analysis (character frequency, morphology, and lexical frequency) are intertwined throughout the human decipherment process. Partial knowledge of each drives discovery in the others.

We capture these intuitions in a generative Bayesian model. This model assumes that each word in the lost language is composed of morphemes that were generated with latent counterparts in the known language. We model bilingual morpheme pairs as arising through a series of Dirichlet processes. This allows us to assign probabilities based both on character-level correspondences (using a character-edit base distribution) as well as higher-level morpheme correspondences. In addition, our model carries out an implicit morphological analysis of the lost language, utilizing the known morphological structure of the related language. This model structure allows us to capture the interplay between the character- and morpheme-level correspondences that humans have used in the manual decipherment process.

We applied our model and inference technique to a corpus of Ugaritic, an ancient Semitic language discovered in 1928. Ugaritic was manually deciphered in 1932 using knowledge of Hebrew, a related language. Our model translated 60% of Ugaritic words into their correct Hebrew cognates and correctly mapped 29 out of 30 letters to their Hebrew counterparts.

This research is under the direction of professor Regina Barzilay, with Benjamin Snyder and Kevin Knight.

### Collaborative Micro Aerial Vehicle Exploration of Outdoor Environments

Field personnel such as soldiers, police SWAT teams, and first responders face challenging, dangerous environments, often with little advance knowledge or information about their surroundings. Currently, intelligence, surveillance and reconnaissance (ISR) information is provided by satellite imagery and prior or second-hand experiences. Although satellite imagery is the preferred method for gaining situational awareness about an outdoor environment, it has many shortcomings. The maps are often outdated and, due to shadows and shading, give false impressions of elevations and details of the environment. Critical features of buildings such as doorways and windows are hidden from view. Combined, these flaws often give field personnel a false mental model of their environment.

Since field personnel must perform a primary task (such as finding people and exploring the environment) in addition to surveillance and reconnaissance, an autonomous robot would aid them in performing ISR and improve their situational awareness in real time. Recent efforts have led to the creation of micro aerial vehicles (MAVs), a class of unmanned aerial vehicle that are small and have autonomous capabilities. A few feet in size at most, a MAV can hover in place, perform vertical take-off and landing, and easily rotate with a small sensor payload. The compact size of these vehicles and their maneuvering capabilities make them well suited for performing highly localized ISR missions with a MAV operator working within the same environment as the vehicle. Unfortunately, existing interfaces for MAVs ignore the needs of field operators, requiring bulky equipment and full attention. To be able to collaboratively explore an environment with a MAV, an operator needs a mobile interface that can support his or her divided attention.

To this end, the micro aerial vehicle exploration of an unknown environment (MAV-VUE) interface was designed and implemented. Using MAV-VUE, operators can navigate the MAV using waypoints, which require little attention. When an operator needs more control over a MAV's location and orientation in order to obtain imagery or learn more about an environment, he or she can use the nudge control mode. Nudge control uses perceived first order (PFO) control to allow an operator to effectively "fly" a MAV with no risk to the vehicle. PFO control, which was invented for MAV-VUE, utilizes positional feedback to fly the MAV while presenting rate control feedback to the operator.

A usability study was conducted to evaluate MAV-VUE. Participants were shown a demonstration of the interface and then given only three minutes of training. Next, they were given search and identify objectives with MAV-VUE installed on an iPhone® and an actual MAV to explore a GPS-simulated urban environment. Participants performed well at the task, with 13 out of 14 successfully completing their objectives with no crashes or collisions.

This research is under the direction of professor Missy Cummings and was featured in the American Institute of Aeronautics and Astronautics's *Aerospace America* magazine in December 2009 as one of the year's most substantive advancements in aerospace information systems.

### Computer-Aided Programming and the Future of Programming Tools

Programming requires a combination of clever insights and careful attention to detail at every level of abstraction, from the design of individual algorithms and data structures, to the integration of the thousands of components that make up modern applications. The challenge for programming tools is to relieve programmers from the more mundane aspects of this process while allowing them to express their insights and creativity to produce programs that work. To address this challenge, our group is developing a new generation of computer-aided programming tools that combine new algorithms for automated reasoning with massive amounts of computing power to help turn programmers' insights into efficient and reliable software.

For example, the development of low-level systems code is one of the most demanding forms of programming; it is a domain where programmers are expected to combine deep algorithmic insights with clever implementation tricks to satisfy strict operational requirements. In order to help programmers work in this domain, we have developed a technology called "storyboard programming" that allows programmers to *draw* the key steps of an algorithm and the high-level insights that make it work. From these diagrams, the system is able to derive complete implementations that incorporate the programmer's insights. This technology raises the level of abstraction at which programmers write complex algorithms and data-structure manipulations, but keeps them in control of important low-level details of implementation.

At the other end of the spectrum, we are developing technologies to cope with the challenges of building applications from pre-existing components. Modern software systems increasingly rely on massive application frameworks to provide a number of services to the application. The advent of these frameworks has revolutionized software development, making it possible to write applications with rich functionality with relatively little new code simply by piecing together pre-existing components. But the productivity benefits come at a price—a steep learning curve as programmers struggle to master a complex framework with tens of thousands of components and millions of lines of code. To address this problem, our group is developing a tool called MatchMaker that helps programmers understand the complex interactions between components in a framework. The tool builds a database to record the details of all the different interactions among components in the framework when they are used by different applications and then uses this database to help programmers discover what components to use and how to connect them together to deliver a particular functionality. In this way, the tool can draw from the insight of past users to reduce the learning curve for future users.

Storyboard programming and MatchMaker are two examples of how the application of automated reasoning and massive computing power is enabling a new generation of programming tools to address some of the more pressing challenges in software development.

This research is under the direction of professor Armando Solar-Lezama.

### Programming Cells for Synthetic Biology Applications in Medicine

With recent advances in our understanding of cellular processes and DNA synthesis capabilities, we can now regard cells as "programmable matter." Through advanced genetic engineering methods, we are equipping cells with new capabilities for gene regulation, information processing, cell-to-cell communication, and actuation. These new capabilities serve as catalysts for synthetic biology, an emerging engineering discipline that focuses on programming cell behaviors. An important goal for synthetic biology is to make the construction of novel biological systems into a practical and useful engineering discipline. The key is the development of an engineering methodology based on standardized and well-characterized interchangeable parts.

Towards this end, we have pioneered the use of computer engineering principles of abstraction, composition, and interface specifications to program cells with sensors and actuators precisely controlled by analog and digital logic circuitry. We have demonstrated synthetic gene networks that implement a variety of biochemical logic circuits in bacteria, yeast, and mammalian cells fabricated using genetic logic gates such as "and," "not," and "implies." We built genetic digital logic circuits such as memory units and an ultrasensitive transcriptional cascade whose digital behavior improves significantly with the addition of genetic components. We also implemented analog circuits that perform signal processing to detect specific chemical gradients and generate pulses in response to cell-to-cell communication. We fabricated genetic circuits that integrate digital, analog and communication capacities and coordinate the behavior of cell aggregates for purposes such as pattern formation.

The implementation of these circuits combined with the development of a new biological engineering foundation are now enabling us to pursue novel applications enhanced by our ability to "program" cells in sophisticated ways. The range of applications for synthetic biology is vast, including, but not limited to, medical diagnostics, therapeutics, sensors, environmental remediation, energy production, and a host of other biomolecular and chemical manufacturing outputs. Synthetic biology's most significant contributions may well come in medical applications such as tissue engineering. Recently we achieved preliminary experimental results for obtaining precise spatiotemporal control over stem cell differentiation with the ultimate goal of being able to create tissues by design. For this purpose, we coupled elements for gene regulation, cell fate determination, signal processing, and artificial cell-to-cell communication.

A related project focuses on creating an artificial tissue homeostasis system where genetically engineered stem cells can indefinitely maintain a desired level of pancreatic beta cells despite attacks by the autoimmune response. The system—which relies on artificial cell-to-cell communication, various regulatory network motifs, and programmed stem cell differentiation into beta cells—may one day be useful for the treatment (or cure) of diabetes. In another effort, we are genetically engineering viruses that enter cells. We then compute whether each individual cell has a biomarker profile indicative of cancer and destroy those cells whose profile is matched. For other cells, the virus performs no action and simply degrades. The enabling technology for this new approach is genetic circuitry encoded by a virus that integrates information from multiple biological markers inside an individual cell and activates a therapeutic response only in cancerous cells with a high degree of specificity.

This research is under the direction of professor Ron Weiss.

### Robust Autonomous Explorers and Human Robot Teams

The last decade has seen a revolution in sensor-rich, embedded systems that have fundamentally reshaped how we explore and live within our world. For example, space explorers must achieve increasingly complex missions in harsh environments without human intervention, while industry envisions a future of manufacturing in which humans and robots work fluidly as teammates in unstructured environments. As these systems become more powerful, they become increasingly complex, and this complexity brings with it substantial challenges in understanding and control. The Model-based Embedded and Robotic Systems (MERS) group, led by professor Brian Williams, is pursuing research on model-based autonomous systems that enable these new artifacts to be controlled based on a key principle: robust and powerful control can be achieved through controllers that understand how a device works.

This principle and approach has emerged through two decades of experimentation with and deployment of model-based autonomous systems, including ubiquitous computing, deep space probes, Lunar and Mars rovers, cooperative air vehicles, deep-sea explorers, and teams of humans and robots. Implementation of this principle has four elements: First, to simplify programming, model-based programs specify strategic guidance at the goal level in terms of *what* states a system should achieve over time (e.g., achieve thrust), rather than *how* to achieve those states through specific sensing and control actions (e.g., open valve). Second, model-based executives implement this guidance while achieving substantial robustness by using estimators to determine current system state and to diagnose failures, and a planner to generate actions and recoveries. This enables the executive to detect novel failure combinations and to propose novel courses of action in response. Third, to enable understanding by these executives, probabilistic constraint automata provide a natural representation for modeling and reasoning about elements of systems that exhibit a rich set of discrete, continuous, and stochastic behaviors. Finally, these executives reason about what is optimal by combining logical inference with optimization over discrete and continuous variables. The executives reason in real time by having the optimization algorithms learn from past failures and steer away from these failures in the future, by reasoning incrementally about the differences between candidate solutions, and by automatically decomposing optimization problems into decoupled sub-problems.

During the last year, the MERS group has made significant advances along three fronts: risk-sensitive operation of deep-sea vehicles, fluid coordination of teams of humans and robots, and programming of a lunar rover through demonstration.

In the field of environmental modeling and prediction, remote sensing has enabled a precise understanding of Earth's atmosphere while the sensing (and consequently the understanding) of Earth's oceans remains limited. Autonomous underwater vehicles (AUVs) are expanding our view of the ocean by performing missions ranging from a day to a month without human intervention. However, more ambitious AUV missions can be extremely risky due, for example, to collisions with the sea floor or biofouling. The MERS group is developing risk-sensitive, model-based programming languages that allow operators to specify acceptable risk levels for different mission segments. In addition, based on an approach called iterative risk allocation, MERS is developing risk-sensitive, model-based executives that control AUVs within specified risk levels, but take risks when the mission benefit is the greatest. This approach was demonstrated last August off of Cape Cod using *Odyssey IV*, a hovering AUV from MIT Sea Grant's AUV Laboratory.

Effective human teams perform shared tasks with little communication and work fluidly by constantly adapting to each other. The MERS group developed and demonstrated model-based executives that enable a robot-to-robot team consisting of two whole-arm manipulators to exhibit similar adaptability. In this approach, the two robots acted either as equal partners, in which each partner adapts to the choices of the other partner, or as leader and assistant, in which one robot adapts to the needs of its leader without constraining the leader's course of action. Future research will develop these concepts in the context of manufacturing by human-robot teams.

The next generation of space explorers, such as the robonaut humanoid robot and the all-terrain hex-legged extra-terrestrial explorer (ATHLETE) Lunar Rover have highly sophisticated manipulation and mobility capabilities. For example, ATHLETE is a six-legged rover developed for moving and deploying in lunar habitats with legs that are wheeled and can be used for manipulation as well as walking. Traditional approaches to programming rovers (such as Spirit and Opportunity) are impractical for robots of this complexity. Instead, in collaboration with the California Institute of Technology's Jet Propulsion Laboratory, the MERS group demonstrated a new paradigm for commanding ATHLETE through a combination of commonsense descriptions and demonstration by example.

This research is under the direction of professor Brian Williams.

## Undo Computing

Virtually any computer system can be compromised. New vulnerabilities are discovered and exploited daily. Even if the base system is secure, administrators may make mistakes in specifying security policies, users may choose weak passwords that can be guessed by attackers, or users may unknowingly install malware along with free screensavers or greeting cards they download from the internet. In practice, this means many machines are routinely compromised, leading to a prevalence of large-scale botnets—backdoors that allow attackers to remotely control machines—and other malware. Cleaning up after these inevitable intrusions leads to days of wasted effort by users or system administrators, with no conclusive guarantee that all traces of the attack are gone, or that no legitimate changes are lost.

The key problem in recovering from an intrusion is to preserve all legitimate changes since before an attack while reverting all of the attack's direct and indirect effects. For example, when an attacker compromises a system, he or she may edit the password file to add an account for the attacker or trojan commonly-used programs such as Word or Excel. After this compromise, any login session that uses the password file is suspect because it might be a log in by the attacker possibly using an existing account whose password the attacker changed. Similarly, any documents edited with the trojaned version of Word or Excel are now suspect as well since they may be infected with a virus.

Our research group has been working on a new approach to recovery from intrusions called undo computing. With our system, named Retro, users or administrators specify unwanted actions from the past, such as a network attack or the installation of a trojaned program that they would like to undo, and Retro repairs any direct and indirect effects of those actions while preserving everything else. For example, Retro helps an administrator selectively undo the side effects of his or her mistakes in specifying security policies, or undo the effects of an attacker's break-in. Similarly, Retro enables users to selectively undo the installation of a trojaned screensaver. More broadly, an undo computing system can be useful even in the absence of an attack, such as helping users undo the effects of accidentally changing the wrong setting the week before.

The key research idea behind undo computing is to selectively re-execute affected computations, such as logging into a system or generating a PDF file from a Word document, to determine how that computation would have proceeded in the absence of the unwanted actions. To do this, we have designed a system-wide dependency graph that records dependencies between all computations and their inputs and outputs during normal execution. Because re-execution is expensive and often non-deterministic, we have developed techniques to determine whether the inputs to a computation are semantically equivalent before and after re-execution, in which case re-execution is unnecessary.

Using a prototype of our Retro undo computing system, we have demonstrated recovery from a range of real and synthesized challenge attacks, with our system undoing all of the side effects of the attack while preserving legitimate changes. In future research, we plan to explore programming models that make applications easier to undo, user interfaces that help users pinpoint unwanted actions from the past, and "undoing actions" in distributed systems like the web.

This research is under the direction of professor Nickolai Zeldovich.

## Laboratory-Sponsored Activities

### CSAIL Outreach

#### *Imara*

The overall goal of Imara is to find and implement long-term, sustainable solutions to make educational technology and resources available to domestic and international communities. In concert with partners both inside and outside of the Institute, we hope our work will enable us to help bridge the digital divide—the gap between those who have access to the latest information technologies and those who do not.

#### *CommuniTech*

CommuniTech is a domestic outreach initiative that attempts to heal the digital divide in the local community. The program provides economically disadvantaged adults with the tools they need to gain access to valuable information they can use to better their lives and the lives of their families.

CommuniTech focuses on both teaching basic computer technology skills and providing accessibility to computer hardware. Families Accessing Computer Technology (FACT) is a six-week course taught by MIT student volunteers, designed to teach basic computer skills to adults. Clients gain basic computer proficiency, marketability to prospective employers, and become more connected to the world of information technology at large.

Clients who complete the class receive computers from FACT's partner program, the Used Computer Factory (UCF). UCF refurbishes donated computers by installing new operating systems and productivity software. UCF also plays a small part in greening MIT by recycling unwanted equipment that might otherwise end up in landfills.

### *Lacotec Laare, Kenya*

The Laare Community Technology Centre (LCTC) in Kenya was founded by Eric Mibuari '06 in 2005 with the assistance of MIT's Public Service Center. The aim of this community center is to increase general computer awareness and literacy in the Laare community by providing cheap, local, and accessible training in the use of computers.

LCTC particularly targets high school–educated youth, seeking to equip them with basic computing skills that they can apply in their everyday activities, in industry, and in education. So far, the center has offered various levels of training to more than 500 students. LCTC takes particular note of the economic difficulties of its potential trainees and strives to charge the lowest possible fees. Recently, the center expanded to include a primary school information technology educational initiative that targets elementary school students.

### *Middle East Education through Technology*

CSAIL's support of the Middle East Education through Technology (MEET) program has continued over the past year. MEET is an innovative educational initiative aimed at creating a common professional language between young Israelis and Palestinians. MEET enables its participants to acquire advanced technological and leadership tools while empowering them to create positive social change within their own communities. Many MIT students volunteer to teach MEET summer courses at the Hebrew University in Jerusalem. CSAIL continues to host http://meet.csail.mit.edu/ and provide technical support to the MEET program.

### *TEK*

While the Internet has revolutionized information delivery for most of us, for many communities in the developing world it remains an economic and technological challenge to access online resources. High charges for telephone and internet service provider access can quickly grow unaffordable, and low-bandwidth connections limit the amount of material that can be viewed per session. Because phone lines are limited, it is often difficult to time-share between internet and voice. Furthermore, unreliable network and power infrastructures can sometimes block internet access altogether. Several CSAIL members have been supporting the Time Equals Knowledge (TEK) project, which empowers low-connectivity communities by providing a full Internet experience using email as the transport mechanism.
Compared with direct web access, email can be much cheaper, more reliable and more convenient in developing areas. The TEK client operates as a proxy on the user's machine, enabling users to browse downloaded pages using a standard web browser. New searches are automatically encoded as emails and sent to the TEK server, which queries the web and returns the contents of resulting pages via email. TEK is free software distributed under the GNU Lesser General Public License.

### *OpenCourseWare Outreach Initiative*

Adnan Esmail, a mechanical engineering graduate student, has combined MIT's OpenCourseWare (OCW) initiative with CSAIL's support to bring the Institute's educational resources to the Indian subcontinent. The OCW Outreach initiative consists of mirror sites that make the course information available to those in nations with poor bandwidth infrastructure.

The program has been realized in partnership with Aligarh University in Aligarh, India, and Lahore University of Management Sciences in the capital of Pakistan's Punjab province. It runs on hard drives that have been generously donated by Maxtor for OCW proliferation. This expansion of access gives talented students without the technological resources they need a chance to reach their full educational potential.

### Seminar Series

Three speakers gave presentations during this year's Dertouzos Distinguished Lecture Series:

- Barbara Liskov, MIT, "The Power of Abstraction"
- Oussama Khatib, Stanford University, "Human-Centered Robotics"
- Bernard Chazelle, Princeton University, "The Analytical Challenge of Natural Algorithms"

## Organizational Changes

During 2010, Victor Zue continued to serve as director of CSAIL. The director's duties include developing and implementing strategies designed to keep CSAIL growing and evolving, fund raising, determining laboratory policies, and examining promotion cases. Three laboratory faculty members served as associate directors and assisted the director with his duties: Frans Kaashoek, Christopher Terman, and during the fall semester, Daniela Rus. William Freeman served as interim associate director during the spring semester while Daniela Rus was on leave.

CSAIL has expanded the work and oversight of its executive committee, which met twice per month to review and advise the director on policy, processes, and activities within the laboratory. Members of the executive committee include Regina Barzilay, Jack Costanza, Erik Demaine, John Fisher, David Gifford, John Guttag, Nancy Lynch, Robert Miller, Karen Shirer, and Patrick Winston.

John Guttag continued in the role of space czar, overseeing the space committee and managing the allocation of space within CSAIL. The space committee also implements improvements to the facilities that will increase the quality of the environment for the laboratory's faculty, staff, and students.

Jack Costanza continued as the assistant director for infrastructure, overseeing information technology infrastructure and user support, building operations, and communications. Karen Shirer continued in her role as assistant director for administration, overseeing finance and human resources.

Elizabeth Bruce continued as director of industry partnerships. She oversees the CSAIL Industry Affiliates Program (CSAIL-IAP), a corporate membership program that offers companies the opportunity to access CSAIL's faculty and students through annual conferences, recruiting events, and onsite visits. In addition, Elizabeth Bruce and Chris Terman continued to lead CSAIL's industry strategy working group, which explored policies and processes to improve engagement with our industry partners.

## Awards and Honors

Our faculty and staff won many awards this year, including the following:

Scott Aaronson: 2009 Junior Bose Teaching Award, MIT; 2009 TIBCO Career Development Chair, MIT; 2009 DARPA Young Faculty Award, MIT; 2009 NSF CAREER Award

Regina Barzilay: 2009 Best Paper Awards, ACL with Satchuthanan Branavan, Harr Chen and Luke Zettlemoyer

Erik Demaine: George M. Sprowls Award; Best Doctoral Thesis by Mihai Patrascu

Shafi Goldwasser: 2010 Franklin Medal Recipient; George M. Sprowls Award; Best Doctoral Thesis by Vinod Vaikuntanathan

Polina Golland: Elected to Board of Directors of MICCAI; MICCAI Young Scientist Award with Tammy Riklin-Raviv, Koen Van Leemput, and William Wells III

Berthold Horn: 2010 IEEE Azriel Rosenfeld Award recipient

David Karger: ACM Fellow

Charles Leiserson: 2009 ACM SPAA Best Paper Award

Nancy Lynch: 2010 IEEE Piore Award

Tomaso Poggio: 2009 Okawa Prize

Martin Rinard: ACM, Fellow

Ronald Rivest: 2009 NEC C&C Foundation Award with Adi Shamir & Leonard Adleman

Daniela Rus: IEEE Fellow; 2009 IROS Best Paper Finalist

Madhu Sudan: IEEE Fellow

Russ Tedrake: 2008 NSF Career Award; 2009 DARPA Young Faculty Award; 2009 RSS Best Paper Award

Joshua Tenenbaum: George M. Sprowls Award; Best Doctoral Thesis by Vikash Mansingha

Victor Zue: Electrical and Computer Engineering Department Academy, University of Florida, elected member

## Key Statistics for Academic Year 2010

CSAIL members in AY2009: 814

Faculty: 83 (13% women)

Research staff: 40 (18% women)

Administration, technical, and support staff: 67 (60% women)

Postdocs: 50 (10% women)

Visitors: 121 (12% women)

Paid Undergraduate Research Opportunities
Program participants: 85 (44% women)

Master of engineering students: 45 (29% women)

Graduate students: 325 (27% women)

**Victor Zue**
**Director**
**Delta Electronics Professor of Electrical Engineering and Computer Science**

*More information about the Computer Science and Artificial Intelligence Laboratory can be found at http://www.csail.mit.edu/.*