CONSUMERIZATION OF IT

SECURITY P. 9
USABILITY P. 66
ENTERPRISE IT P. 20

IEEE
IEEE Computer Society

www.computer.org/itpro
Don’t Let Big Data Analytics Bog You Down!

Everyone says this is the era of big data analytics. But what does that mean to you and your business? Can big data analytics help increase profits and satisfy customers? Can you trust it? Find out.

Meet the Rock Star Experts

- Grady Booch, Chief Scientist of Software Engineering, IBM
- Greg Arnold, Data Infrastructure Engineering Director, LinkedIn
- Chris Pouliot, Vice President, Lyft (Former Netflix Director of Analytics)
- Guido Schroeder, Senior Vice President of Products, Splunk
- Matthew Denesuk, Chief Data Science Officer, GE Software
- Mark Davis, Distinguished Engineer, Dell (Former CTO, Kitenga)
- Mike Rosenbaum, CEO, Catalyst IT Services
- Panel: How Far Can We Trust Big Data Analytics?

Be there for the REAL answers on big data analytics.

21 October 2014
San Jose Civic Auditorium
San Jose, CA

REGISTER NOW
Early-Bird Pricing is in Effect!

Reserve Your Place in this Sure-to-be Sold Out Event

computer.org/Data-Analytics
Consumerization of IT

IN THIS ISSUE

16
Guest Editors’ Introduction: From BYOD to BYOA, Phishing, and Botnets
Seth Earley, Robert Harmon, Maria R. Lee, and Sunil Mithas

20
Consumerization in the IT Service Ecosystem
Enrique Castro-Leon

28
Trust and Privacy Exploitation in Online Social Networks
Kaze Wong, Angus Wong, Alan Yeung, Wei Fan, and Su-Kit Tang

34
Low-Cost 3D-Supported Interactive Control
Enrico Masala, Antonio Servetti, and Angelo Raffaele Meo

41
Job Recruitment and Job Seeking Processes: How Technology Can Help
Paolo Montuschi, Valentina Gatteschi, Fabrizio Lamberti, Andrea Sanna, and Claudio Demartini

Most available literature on IT consumerization focuses on current practices. Exploring the driving forces behind consumerization is essential, however, if IT organizations are to make sense of the current dynamics of transformation and disruption and formulate effective strategies in response.

This survey presents the pitfalls of security protection in online social networks and identifies common attack methods. A harmless proof-of-concept malware app demonstrates the vulnerability of online social networks and the significance of the user’s mentality.

Devices for stereoscopic vision are mostly used for entertainment. The prototype device presented here for 3D-video-supported interactive control is based on consumer devices and open source software, paving the way for low-cost general-purpose systems for remote operation, education, training, and surveillance.

This survey of current job search and recruitment tools focuses on applying a computer-based approach to job matchmaking. The authors present a semantic-based software platform developed in the framework of a European project on lifelong learning, highlighting future research directions.
Cinema Cloud: An Enabling Technology for the Movie Industry
Wai-Ming To, Linda S.L. Lai, David W.K. Chung, and Andy W.L. Chung

Deploying a Maritime Cloud
Kleanthis Dellios and Dimitrios Papanikas

With the promise of high definition and advanced special effects, movie theaters worldwide have evolved from traditional film projection to digital cinema projection. The Cyberport Digital Cinema Cloud helps facilitate the development of digital entertainment in Hong Kong.

The authors offer a roadmap for building a Maritime Cloud, offering virtual platforms, tools, and data for the maritime domain. They describe the proposed architecture and delivery models, evaluate advantages and challenges, and suggest future research directions.
Learning analytics is an emergent field of research and practice that aims to use data analysis to inform decisions made on every tier of the educational system. It helps analysts decipher trends and patterns from educational big data—that is, huge sets of student-related data—to further the advancement of personalized, supportive higher education applications.

Learning Analytics Initiatives
Learning analytics is the third wave of developments in instructional technology, which began with the advent of the learning management system (LMS) in 1991. The second wave integrated the LMS into the wider educational enterprise by involving learners on social networks (also known as the Web 2.0 wave). During this third wave, learning analytics as a term has been significantly popularized by the Educause International Conferences on Learning Analytics and Knowledge (LAK), which started in 2011 (https://tekri.athabasca.ca/analytics).

Learning analytics focuses on collecting and analyzing data from a variety of sources to provide information on what works (and what doesn’t) with respect to teaching and learning. This helps educational institutions improve their quality of learning and overall competitiveness. Consequently, many research communities have developed a variety of promising initiatives, models, and applications to improve learner success. For example, Santa Monica College’s Glass Classroom initiative, introduced in December 2012, aims to enhance student and teacher performance by collecting and analyzing large amounts of data. Using real-time feedback of the student’s performance, Glass Classroom adjusts the courseware to meet educational objectives.

Another example is at the University of Wisconsin-Madison. Since May 2012, the university has been working to develop a data-driven “early-warning” system that faculty and advisors can use to support student academic success. The system aims to identify academically at-risk students, using nontraditional indicators that can be gathered early in a student’s career, even at the beginning of a semester. The system aims to intervene early, improve students’ academic success, and bolster the campus’s retention and graduation rates.

Furthermore, many research groups and societies are providing excellent networks for researchers who are exploring the impact of analytics on teaching, learning, training, and development (see the “Related Learning Analytics Research” sidebar). In particular, these groups and societies are promoting several learning analytics models, which have been
developed to identify student risk levels or success factors in real time to increase a student’s likelihood for success and improve learning.

Table 1 provides examples of applications that are using such learning analytics models. Higher education institutions have shown increased interest in learning analytics as they face calls for more transparency and greater scrutiny of their student recruitment and retention practices.

<table>
<thead>
<tr>
<th>Application</th>
<th>Purpose and data</th>
<th>Institution &amp; URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>jPoll</td>
<td>This ubiquitous classroom polling and student-response system engages students using a range of interactive teaching situations. Originally developed as a replacement for clicker-type technologies, it produces data based on student opinions of teaching events.</td>
<td>Griffith University <a href="http://navigator.nmc.org/project/jpoll">http://navigator.nmc.org/project/jpoll</a></td>
</tr>
<tr>
<td>Moodog</td>
<td>This course management system provides a log analysis tool to track students’ online learning activities. It provides instructors with insight about how students interact with online course materials, and it lets students compare their own progress with others in the class. Moodog also provides automatic email reminders to students, encouraging them to view available materials that they haven’t yet accessed.</td>
<td>University of California, Santa Barbara (An article with more information appears elsewhere.)</td>
</tr>
<tr>
<td>Equella</td>
<td>Equella provides evidence of appropriate curriculum coverage, student engagement, and equity while students are on clinical placement (which provides a clinical education at an external facility).</td>
<td>University of Wollongong <a href="http://www.pearsonlearningsolutions.com/equella">www.pearsonlearningsolutions.com/equella</a></td>
</tr>
<tr>
<td>E2Coach</td>
<td>This computer-based coaching system provides a model for an intervention engine, capable of dealing with actionable information for thousands of students. It’s currently being used for a variety of applications in public health—from cessation of smoking to losing weight.</td>
<td>University of Michigan <a href="http://sitemaker.umich.edu/ecoach/about_ecoach">http://sitemaker.umich.edu/ecoach/about_ecoach</a></td>
</tr>
<tr>
<td>Signals</td>
<td>A system that takes course management system (CMS) data (Blackboard data, in this case) into account in predicting whether a student will succeed in a given course.</td>
<td>Purdue University <a href="http://www.itap.purdue.edu/studio/signals">www.itap.purdue.edu/studio/signals</a></td>
</tr>
<tr>
<td>Sherpa</td>
<td>This course recommender system uses a service-oriented architecture to guide students during registration to pick a substitute class if their first choice is full (Phase 1). The second phase provides administrators tools to send “nudges” to students via portal messages, emails, SMS messages, and text-to-speech phone calls. The third phase revamps the student portal with a to-do list, news feed, and calendar, pre-populated with enrollments and important dates.</td>
<td>South Orange County Community College <a href="http://www.socccd.edu/sherpa">www.socccd.edu/sherpa</a></td>
</tr>
</tbody>
</table>

Dealing with Unstructured Data

The applications described in Table 1 apply models of structured data, collected from student interactions with the learning system, to answer the following questions:

- What grade is a student likely to get without intervention?
- What is the best next course for a given student?
- Should a student be referred to a counselor for help?

These learning analytics applications use Web analytics, data warehouses, and other data-management tools, measuring Web traffic to assess and improve the effectiveness of the teaching and learning website. (They use clickstream data to record every page,
segment, or tag requested by the learner.) Then, they extract knowledge from structured data and store it in databases. You define your learning categories up front, so it’s simple to evaluate the status of any given category; track trends in a category; and compare how a category relates to others across time, geography, and so on.

Analyzing unstructured content, however, is different. Unstructured content refers to documents, emails, answers, responses, and other objects (static or dynamic) that are made up of free-flowing text. Analysts, including Gartner and IDC, predict that as much as 80 percent of data is unstructured.9 This data is a treasure chest of valuable information and insight, but it’s notoriously difficult to manage. Gartner defines unstructured data as content that doesn’t conform to a specific, predefined data model. It tends to be the human-generated and people-oriented content that doesn’t fit neatly into database tables.10 Text is the classic and most dominant example of unstructured data.

Capturing and analyzing textual data has changed how decisions are made and resources are allocated in businesses, healthcare, government, and many other fields. However, text analysis hasn’t yet made the impact on education that it has made in other fields, even though learners interact with instructors, other learners, and with the course materials mainly using text. The proliferation of textual data in education is overwhelming, because such data is being constantly generated via learning blogs, emails, educational websites, learning objects repositories, and so on. Although the amount of textual data is increasing rapidly, learning systems’ ability to summarize, understand, and make sense of such data for improved learning remains challenging.

**Textual Analytics in Learning**

A finer level of understanding of student opinions, answers, and concerns can enable educators, researchers, and university leaders to improve teaching and thus learning.11 Text analytics typically involves tasks such as text categorization, text clustering, summarization, and concept extraction. Employing learning analytics based on text analytics helps promote automated intervention and reasoning about textual interaction. It also helps archive, filter, search, and classify text.

Thanks to advancements in textual analysis techniques and other machine-learning technologies, systems like the IBM Watson can defeat top human contestants in Jeopardy (see www.ibm.com/smarterplanet/us/en/ibmwatson). According to Seth Grimes,12 text analytics adds semantics to identify features such as:

* named entities (learning objects, learners, and so on);
* pattern-based entities (such as email addresses);
* concepts (abstractions of entities);
* facts and relationships;
* events;
* concrete and abstract attributes; and
* subjectivity in the form of opinions, sentiments, and emotions.

Learning analytics applications enriched with the power of textual analytics might also perform the following smart capabilities and services to empower the learner and the learning system.

**Annotate Sensitive Information**

In a learning system, you need to communicate clearly and quickly with learners. However, creating screen recordings can be time consuming and frustrating. You need the ability to capture images to get your point across. You also need the ability to add or annotate text and to blur out sensitive information on these images.

**Categorize Groups of Documents**

This would let applications assign unseen text to one of the predefined categories based on a processing algorithm. There are many kinds of models and strategies for text categorizations, but generally, all algorithms could be divided into the following groups:

* supervised categorization uses prior information regarding correct categorization for tested documents;
* unsupervised categorization doesn’t use any prior or external information, so decisions are made internally based on predefined logic or models; and
* semisupervised categorization combines some approaches from the supervised and unsupervised text categorization.

**Check for Relevant Live Feeds**

Live feeds are different types of feeds, which show all messages that are relevant to a specific user or topic.

**Other Text-Based Services**

Other potential text-based services include the following:

* classify entities via manual training and automation,
* filter content by metadata and information regarding core and features,
* generate high-level summaries and detailed reports,
* discover top meta values and related concepts,
* recommend resources,
* analyze feedbacks and opinions,
* build topic models and generate groupings,
* measure and validate results, and
Text analytics applies a variety of natural language processing analysis techniques along with linguistics, statistical, and data-mining techniques to extract concepts and patterns that can be applied to categorize and classify textual documents. It also attempts to transform the unstructured information into data that can be used to support more traditional learning analytics techniques. Finally, it helps identify meaning and relationships in large volumes of information. However, there is no single method appropriate for all text analysis tasks.

- Identify similar documents and responses (possible plagiarism).

Learning analytics approaches must take several different perspectives and accommodate different data sources. The ideal vision for learning analytics is to integrate analytics for both structured and unstructured data (mainly of textual nature). Figure 1 illustrates our bird’s eye view of a comprehensive learning analytics architecture.

The main feature of this architecture is the component that deals with unstructured textual data (and natural language processing). Adding this to the existing components (that is, to the learning management system, Web 2.0, social networking, and the learning analytics for structured data) will strengthen the power of the “meta engines” to reliably assess students’ skills and provide students with formative feedback based on their learning processes (their actual cognitive and intellectual development while performing a learning activity). By adding a component that deals with textual data, the meta engines will be able to capture detailed information about teaching and learning. For example, the predictive engine will be able to record the progress of certain critical-thinking processes and predict the outcome.

Similarly, the content engine will provide more focused information and evaluations using the indexing and content categorizations. The adaptation engine will provide more...
customized content delivery for individual students’ performance and interests. The intervention engine will let instructors and administrators bypass the LMS to directly interact with students. The feedback engine will provide evaluations from other meta engines (such as the predictive engine) as feedback for students, instructors, faculty, and administrators. Finally, the measurement engine will provide measures such as the level of similarity between the student’s solution and that of the instructor.

There are many text analytics tools and APIs that can be used to build the learning analytics component that deals with textual unstructured data (such as LingPipe, RapidMiner, Textanalytics, and OpenText). Incorporating textual analytics into what we have achieved so far in terms of learning analytics research and development will lay the groundwork for redesigning educational institutions to teach 21st century skills.

References
10. “Gartner Predicts 2013: Business Intelligence and Analytics Need to Scale Up to Support Explosive Growth in Data Sources,” Gartner, 2013.

Jinan Fiaidhi is a full professor and graduate coordinator with the Department of Computer Science at Lakehead University, Ontario, Canada. She is also an Adjunct Research Professor with the University of Western Ontario. Her research interests are in collaborative learning, learning analytics, and cloud computing. More information about her can be obtained from her website http://flash.lakeheadu.ca/~jfiaidhi. Contact her at jfiaidhi@lakeheadu.ca.

Selected CS articles and columns are available for free at http://ComputingNow.computer.org.
n the corporate world, Bring Your Own Device (BYOD) is becoming increasingly common, changing how we work. According to a 2012 Intel study of 3,000 IT managers and 1,300 users, productivity is viewed as the biggest benefit of BYOD. An IBM Flexible Workplace Study reported increases in productivity of 20 percent or greater stemming from BYOD practices—the equivalent of an extra day of work per week. BYOD lets employees use their personal device to work seamlessly across their personal user space and enterprise workspace instead of using multiple devices depending on business need, location, and circumstances.

Yet today’s IT departments are concerned with the popularity of BYOD, because mixing personal and enterprise data presents security threats to corporate proprietary information. Enforcing the usage of two different mobile devices—one corporate and one personal—could mitigate this threat, but this strategy faces employee resistance because it’s inconvenient. This creates a need for IT departments to develop company security policies that let employees access sensitive resources using personal devices.

The Challenges
In a survey of 2,100 individuals, conducted by Webroot, 41 percent said they use a personal smartphone or tablet for work purposes. Furthermore, 70 percent of those smartphones and tables used for work had no additional security other than what was installed when the employee first purchased the device. According to another survey, 98 percent of employers claim to have a mobile security policy in place for accessing corporate data, yet BYOD users tend to choose usability over security when it comes to selecting mobile applications. Device security, malware, and enforcement are major security concerns raised by BYOD.

Device Security
Security issues arise at all layers—including the network layer—but the main issues are at the device layer. Enterprise apps on a mobile device can leave company data on that device, presenting a major threat if the device is ever lost or stolen. Furthermore, if the user mixes personal and enterprise data, it could lead to data leakage if company information is accidentally sent to personal contacts.

Malware
Another concern is malware. The total number of known Android malware samples increased more than 10 times between July 2012 (about 45,000 samples) and January 2014 (about 650,000 samples). These malwares tend to steal personal information, issue premium SMSs (which result in a fee for the sender) for financial gain, or engage in denial-of-service attacks. The malware might not specifically target enterprise data, but it creates concerns about backdoor data leakage for BYOD scenarios.

Enforcement
Personal devices used for work are part of the enterprise network, so it’s essential to ensure that all mobile devices comply with enterprise security policies. However, it’s difficult to enforce corporate policies on personal devices. The problem is exacerbated by the large variety of device hardware and fragmentation of the operating system. Moreover, security policies can change from time to time to cope with new security threats. Thus, effective security enforcement requires constant updating of both corporate and personal devices.
Although a virtual private network can be an effective way to maintain corporate security, administrators can also let employees access corporate intranet sites through the AirWatch Browser on employee-owned devices. A single sign-on and app tunneling lets users connect the AirWatch Browser to enterprise intranets and third-party Web filters.

Current Solutions
The following solutions and practices can help address BYOD security issues.

Security Policies
Typically, company BYOD policies include identifying which devices can be used in the company network, listing both allowed and banned apps, and describing classes of data that shouldn’t be stored locally after being used by a mobile app. Companies also implement security policies (such as passwords or screen locks) for all devices and have strategies for lost or stolen devices and for when employees leave the company. Furthermore, although a company might assume it owns the apps and data, the device owner might think differently.

The US White House published a BYOD policy on August 2012 that presents three high-level suggestions for implementing BYOD programs. The first suggestion is to use a BYOD design that presents three high-level suggestions for implementing BYOD programs. This lets administrators manage the devices remotely. It might set customized authentication checkpoints on apps or data or restrict certain device features and settings. Some products (such as AirWatch by VMware) also provide Mobile Application Management (MAM) and Mobile Content Management (MCM) in addition to MDM. They can also wipe the device of enterprise content if necessary. Another technique—secure mobile browsing—is discussed in the related sidebar.

Separation Techniques
Techniques based on virtualization and on the operating system (OS) to separate enterprise space and personal space have shown potential.

In a BYOD scenario, private personal apps and data and important enterprise apps and data are contained and running on the same device. A BYOD design requires that the personal user space in no way compromise the security of the enterprise workspace. On the other hand, the enterprise workspace should in no way compromise the privacy of the personal user space. How to effectively separate the two spaces on the same mobile device becomes crucial for a successful BYOD design. Techniques, including virtualization, dual boot, and recently proposed virtual mobile platforms, can be used to achieve the separation goal. All of these techniques have their own advantages and drawbacks.

Virtualization. With hardware virtualization, hardware resources on a mobile device can be multiplexed to host multiple virtual machines. A type 1 (T1) hypervisor runs directly on the system hardware and provides a virtualized platform to host multiple guest OSs (here, the guest OSs would be the personal OS and the enterprise OS). This provides the best separation between the personal user space and enterprise workspace. However, this solution suffers from great performance degradation due to T1 overhead.

A type 2 (T2) hypervisor can also be used to provide system separation. In this case, the hypervisor runs on top of the original mobile device OS, while the enterprise workspace runs as a VM on the hypervisor. An example of this design is the VMware Mobile Workspace (www.vmware.com/mobile-secure-desktop/overview). This solution provides more flexibility and is more user friendly, because the user can simply install the hypervisor as an app in the original OS. However, the enterprise workspace will be at risk if the original OS is compromised.

To enhance system performance, OS virtualization can be used to provide light-weight separation. In this case, kernel-level device namespaces are created to provide data isolation and hardware resource (device driver) multiplexing, allowing multiple virtual mobile devices to run on a single OS instance. An example of this design is Cells from Columbia University. OS virtualization can improve the system
performance compared to hardware virtualization, but the adopter of this technique should keep in mind that both the personal user space and enterprise workspace can be compromised if an attacker targets the OS kernel.

**Dual boot.** This possible solution for device separation involves installing two OSs in different partitions on the same device. This process is just like performing traditional dual boots on desktops. Canonical provides a dual boot app7 that modifies an Android device to let users switch between different OSs. OS dual boot provides a clean separation, but the long switching time degrades usability.

**Virtual mobile platforms.** The Remotium Virtual Mobile Platform provides another way to separate the personal user space from the enterprise workspace using the idea of remote control (www.remotium.com). This is similar to Microsoft Windows’ remote desktop connection. In the Remotium solution, all enterprise data and apps run in secure virtual machines at Remotium’s or the enterprise’s datacenter. Employees can then set up remote connections to the virtual machines through a Remotium mobile client installed on each employee’s mobile device. This is a simple and effective approach to improve enterprise data security, because no data is stored locally on the mobile device. Note, however, that this solution requires a constant Internet connection, and performance is determined by the Internet speed.

including BYOD security in corporate IT management policies is inevitable, and companies must determine support capabilities, educational needs, and deployment phases. At the same time, they must separate personal and organization data and clearly define device requirements. Technologies that support BYOD must be evaluated based on performance, separation, and usability. Management policies for mobile devices and apps must be inclusive, enforceable, and updated constantly. BYOD might come with a high initial cost, but the payoff should be worth it in the long run.

**References**


**To enhance system performance, OS virtualization can be used to provide light-weight separation.**
IT matters more today than ever before, with companies focusing on how technology can help develop new products and services; find and excite customers; support seamless communication among employees, customers, suppliers, and partners; and provide a secure computing and communications infrastructure. Companies’ dependency on IT will only grow as more business models and processes are digitized. Strategic technologies—social media platforms, data analytics tools, and location-awareness apps—can help find and profile customers, suppliers, and partners. Similarly, cloud computing, biometric security, and virtual computing technologies can provide effective operational infrastructures.

Here, we explore seven operational and strategic technology capabilities and management practices, which are shaping today’s enterprise IT systems and applications.

Rent as Much as Possible

Companies are increasingly renting technology. According to the Gartner Group, “the use of cloud computing is growing, and by 2016 this growth will increase to become the bulk of new IT spending.” The cost equation is more than compelling: it’s generally far less expensive to rent technology from the cloud than to buy and install IT. Cloud delivery is also support-free: companies that lease hardware and software never have to fix things when they break or crash and are no longer in the hardware or software business. Those renting are more than 90 percent satisfied.

Some companies are already leasing everything from their cloud providers, including operational technology—email, word processing, spreadsheets, security, database management, and enterprise resource planning (ERP)—and strategic technology—customer relationship management (CRM), location-based services (LBS), learning management systems (LMS), and predictive analytics. Cloud technologies are maturing, which, coupled with a greater awareness of the potential benefits (and limitations), is driving the accelerated adoption of cloud offerings.

One of the major strengths of cloud computing is the freedom it provides companies to focus on their business models and processes. Instead of worrying about network latency and server maintenance, companies can focus on innovation, sales, and marketing, among other revenue-generating activities.

This focus on revenue generation will be further supported by developments in interoperability between clouds, allowing companies to scale a service across disparate providers, while the service appears to operate as one system. Cloud federation will also support revenue generation by interconnecting cloud services of different providers and from disparate networks. This will let a provider rent computing resources to other providers to balance workloads and handle spikes in demand.

Encourage BYOD

The Bring Your Own Device (BYOD) technology delivery model lets employees use whatever technology they want—especially if it reduces technology costs and increases productivity. Companies are offering annual stipends to help reduce the cost of the technology employees prefer. Over time, many companies will dramatically reduce and then completely eliminate their employee technology budgets.
The objective is the thinnest, most mobile devices possible. Although we still use desktops and laptops, employees are increasingly using tablets, smartphones, and other very “thin clients” to access local area networks, wide area networks, virtual private networks, and hosted applications, as well as applications that run locally on (some of) the devices. The economics are so powerful for thin clients that we can expect the “fat” corporate PC, and even the venerable laptop, to disappear from the list of corporate technology assets. BYOD will accelerate this trend.

There’s also a growing opportunity for businesses to exploit the capabilities of wearables, such as Google Glass, Apple’s iWatch, and Plantronics’ Bluetooth headsets—among countless enabling devices changing the way we search, navigate, transact, and live (see Figure 1). Next will be authentication pills and tattoos, and the use of the human body for data transmission. As the adoption of Internet of Things (IoT) technology accelerates, wearables will integrate, communicate, and automatically transact.

**Exploit Big Data Analytics**
Data is now created by everyone—vendors, customers, suppliers, partners, managers, executives, bloggers, and anyone else who would like to offer insights, solve problems, or purchase products and services. Organizing, analyzing, and managing all this data is an emerging core competency.

The goal is real-time descriptive, explanatory, predictive, and prescriptive data, information, and knowledge about internal and external processes and performance. The torrent of unstructured social media data has overwhelmed conventional collection platforms: Google’s recent unveiling of its Cloud DataFlow technology is one answer to the challenge of real-time analytics—and it’s validation of the role that analytics will play going forward.8 To remain competitive, all companies will invest heavily in analytics and business intelligence (BI).

**Mine Social Data**
Social BI assumes that there’s value in connecting people willing or anxious to collaborate through their affinity with friends, colleagues, associates, places, products, and brands. Customer service and new product releases are especially vulnerable to tweeting. Companies now worry about what’s being said about them on Facebook, Twitter, Tumblr, Instagram, TripAdvisor, and countless other social media sites. Some companies just listen but others extract meaning and purpose from social content. Put another way, companies need to not only know what people are saying about them, but why they’re saying what they’re saying—and the response implications of the conversations on products, services, and strategies. This is social BI.

**Exploit Location Data**
Location awareness fundamentally alters business processes across multiple functional areas, but especially in sales and marketing. The key is to define location-based services and discover what requirements the technology can satisfy.

For example, retailers have a vested interest in knowing the physical location of their customers. They want to know their customers’ movement patterns so they can correlate locations with a variety of attributes, such as time of day, age, gender, wealth, and race, among other variables. The next step is to infer from that data precisely how to engage that customer with the right communications and offers. Location-awareness thus fuels and enables customer analytics.

The relationship between analytics, social media, and location-awareness is clear—and cumulative, especially when location awareness is integrated with BYOD, mobility, and cloud computing. Uber, a rideshare, taxi, and taxi-alternative app, is perhaps the perfect example of this phenomenon (https://www.uber.com). Within minutes, Uber

---

**Figure 1. Opportunities for the business operations sector to exploit wearables. (Source: Beecham Research. See the full chart at www.beechamresearch.com/article.aspx?id=20.)**
connects a customer (using a mobile phone) with the driver of a taxi, private car, or rideshare program in cities around the world.

**Use Crowdsourcing**

Expertise was never confined within the corporate firewall. While companies hope to have the best and the brightest employees solving their problems, there are always other qualified professionals who don’t work at the company. Today, we have the opportunity to look to “the crowd” for help.

Vast amounts of data are being generated by in-house people and devices, but there’s also lots of data, ideas, and solutions available from the crowd. Crowdsourcing is a relatively new problem-solving model. The Internet makes it possible to connect problem-creators with problem-solvers instantly and continuously. Small, medium, and large companies solve problems alongside individual contributors and freelancers. But the definition of “problems” is wide. In fact, there are many crowdsourcing models that companies can use to solve problems, perform specific tasks, and even brainstorm. Innocentive (www.innocentive.com) describes how companies such as Coca-Cola, Unilever, Nokia, Anheuser-Busch, and General Mills are using the crowd to solve important problems.

**Federate**

In the 20th century, technology governance was largely about technology standards and centralized control. As we moved into the 21st century, governance shifted from centralized to federated governance. Control shifted because many of the opportunity areas discussed here—BYOD, crowdsourcing, social BI, and other technologies—are impossible to centrally control.

More importantly, companies realize that trying to control IT from a single desk makes little strategic sense. Although everyone agrees that the management of operational and infrastructure technology should be coordinated and efficient (increasingly by a cloud provider), very few companies with multiple lines of business believe that control should include strategic applications that the businesses need to compete.

This means that there will be “technologists” on all of the business teams. It means that there will be sales technologists, marketing technologists, finance technologists, customer service technologists, innovation technologists, and supply-chain management technologists, among others, who understand both business processes and models and current and emerging digital technology—and the potential opportunities IT presents for business activities. This is the new face of federation.

Operational IT—a company’s technology infrastructure—will move. But unlike what we describe today as “business partners,” infrastructure professionals will move to internal “corporate auditing” offices. They will pursue a three-pronged agenda: architecture, infrastructure, and security. After IT moves to auditing, operational IT—responsible for infrastructure activities such as email, storage, back-up, and recovery—will do what operational IT does best: deliver secure, recoverable basic services as cheaply as possible (again, increasingly with cloud providers), “audited” by external partners.

The auditing department is also the best place to locate cloud and applications service level agreement (SLA) negotiations and management. Because procurement is often part of the larger audit team, it’s a natural place to locate cloud SLA management. The opportunity here is to reorganize technology activities to empower business units through the federation of IT rights and responsibilities.

IT’s changing—again. But this time, it’s not about bottom-up enabling technology that satisfies old and “emerging” business requirements. Today, IT is about the integration of emerging business models and technologies. IT is now an equal partner with the business—and that partnership is now permanent.

Perhaps the most intriguing aspect of IT’s new role is its organization and management. Cloud delivery has fundamentally and permanently changed the relationship between business processes and models and current and emerging technology. It’s also changed the way companies acquire, deliver, support, and measure IT. All seven of the opportunity areas discussed here can be pursued without internal IT bureaucracies or even business unit budget constraints. (That said, there will be some selected applications that traverse the enterprise that will require at least some federated, if not centralized, governance.)

While the seven areas offer exciting possibilities, within five years, there will be a whole new set of possibilities. Technology will continue to enable business models and processes, just as business models and processes will stimulate the development of new technologies. The interplay between business and technology will be more complete and continuous. Of course, concerns about security, privacy, trust, and control must be taken seriously, and regulation must occur at all levels.

**References**


Stephen J. Andriole is the Thomas G. Labreque Professor of Business Technology at the Villanova School of Business at Villanova University, where he teaches courses in strategic IT, emerging technologies, and innovation and entrepreneurialism. Contact him at stephen.andriole@villanova.edu and through LinkedIn.

Irena Bojanova is a professor and program director of Information and Technology Systems at University of Maryland, University College. Read her cloud computing blog at www.computer.org/portal/web/Irena-Bojanova. Contact her at irena.bojanova@umuc.edu and through LinkedIn.

Selected CS articles and columns are available for free at http://ComputingNow.computer.org.
The consumerization of corporate IT systems is a relatively recent development that has hit IT organizations seemingly out of the blue. One day, desktop computers were provisioned for users, and applications were carefully vetted and controlled. The next thing CIOs knew, they were dealing with laptops, smartphones, tablets, and a raft of new tools and technologies that presented significant security, reliability, and IP protection challenges.

Moreover, users have come to expect an information experience on par with that of their personal lives. Google, Facebook, LinkedIn, Twitter, Pinterest, Skype, and Dropbox are examples of applications that have infiltrated the enterprise’s IT infrastructure—whether as an extension of personal productivity applications or as new social-engagement platforms for the business side of the organization. Even for organizations whose policies and firewalls have kept external applications out of the internal information landscape, the game has changed. Users have higher expectations about the ease and speed of information access and the user friendliness of systems required for their jobs.

Clearly, new policies must be enacted and new tools deployed to reduce risks and keep up with growing user expectations, but balancing accessibility and security has proven to be a significant challenge.

Redefining Enterprise IT

This issue of IT Professional touches on a number of these topics. In “Consumerization in the IT Service Ecosystem,” Enrique Castro-Leon observes that the trend has redefined IT in the enterprise: “Consumerization has brought an inversion of roles where now users are driving technology adoption and change.” The consumerization of IT is, in part, being pushed by a younger, more mobile workforce comprising active users of new technologies and applications. Employees expect to use their personal devices—and applications—at work, which relates to the concept of BYOX—Bring Your Own Device, Cloud, App, and Network. Instead of new technology flowing down from the business to the consumer, as it did with the desktop computer, the flow has reversed. The consumer market often gets new technology before it enters (and is fully leveraged by) the enterprise.

This blending of personal and business technology is having a significant impact on
corporate IT departments, which traditionally have governed, deployed, and controlled the technology that employees use to do their jobs. Consequently, IT departments must decide how to protect their networks and manage technology that they perhaps didn’t procure or provision.

Castro-Leon goes on to point out that users and the employee community don’t share this concern: they continue to use their devices as they have in the past, and they expect them to work equally well in corporate settings. Indeed, users want their information experience to be seamless, transparent, and portable, just as they do in the rest of their world. Google, Apple, and Facebook—you’ve spoiled our users forever.

More fundamentally, this drive for BYOD and consumer-focused IT is emblematic of a larger societal trend facilitated by technology. There has been a shift to part-time, transient, and specialized professional workers. Just as a tradesman used to bring his or her tools to the shop floor at the start of the industrial revolution, the independent contractors filling roles in IT shops, as well as other professionals, are now beginning to bring their own laptops and devices.

Castro-Leon thus also explores how our economy has shifted to value networks and highly interrelated trading partners. As organizations provide infrastructure that’s in the cloud, and as information systems transmit data across multiple organizations in a value chain, the boundaries of the enterprise shift and become blurry—especially when software is consumed through cloud-based and hosted application providers. Castro-Leon describes the shift from monolithic to componentized applications, then to Web-services-based applications, service networks, and composite applications, which makes apparent the natural evolution of and connection to BYOD and BYOA (Bring Your Own App), especially as we enter the next evolution of the Internet of Things.

Enforcing Security and Privacy

The flip side of the externalization of software applications and functionality lies in the social component. Services architectures facilitate human-to-machine and human-to-human interactions. The latter allows for opportunities for the attackers to exploit vulnerabilities due to human nature and the complexity of online social networks, as discussed in “Trust and Privacy Exploitation in Online Social Networks,” by Kaze Wong, Angus Wont, Alan Yeung, Wei Fan, and Su-Kit Tang.

In addition to discussing issues around complex privacy settings and the inconvenience caused by enforcing security (which reduces sharing among users), the authors observe that

- the strength of your security level is as weak as that of your friend with the lowest level, and
- your data can be inferred even if you don’t disclose it.

Furthermore, attackers have another significant opportunity to access user data: “An app that looks like an interesting game might be designed with the primary purpose of collecting your data. When installing such an app, you might have given the ‘game’ permission (intentionally or unintentionally) to access your profile, albums, and friends list.” Indeed, how often do we casually give apps permission to access various information sources and systems on our devices?

From malware to spam, scam, and phishing (the authors site evidence that 70 percent of the spam on Facebook leads to phishing websites), to botnets, identity forgery, and excess permissions from application installation, online social networks present a tremendous opportunity for attackers to steal information, spread viruses, and wreak havoc on unsuspecting users.

In case you think this is just theory, the authors present an experiment that illustrates five attack approaches: malware, phishing, botnet, online social networks present a tremendous opportunity for attackers to steal information, spread viruses, and wreak havoc on unsuspecting users.
identity forgery, and excess permission requests. Although the experiment didn’t cause any harm, the approach persuaded 97 percent of users who installed the Facebook app to give the app permission to analyze their newsfeeds. In addition, 27 percent were tricked into a simulated upgrade of a flash plug-in from a fake YouTube site, which was potentially (not actually) malicious software. The excess permissions allowed a social graph containing almost 20,000 users, even though the app was installed by only 276 people (with the added bonus of social network analysis to identify hubs for more effective malware marketing). Because users access their social networks on the same devices that they bring to work, online social networks can easily become a vector for attacks on corporate data and system security breaches.

Increasing yet Controlling Accessibility

This issue also features related Data Analytics and Securing IT departments, both of which emphasize the need for enterprises to raise the bar on usability and information access while simultaneously establishing and monitoring personal device usage, security, and access policies. These goals seem in opposition—on one hand, increasing accessibility of information, and on the other, controlling that accessibility and dissemination. The trick is to use transparent approaches and a light touch, rather than intrusive approaches that will only encourage workarounds, by leveraging hardware monitoring and management technologies to isolate corporate networks from malware and network intrusions.

In addition to policy development, several classes of technology help IT organizations proactively manage mobile device security threats, including virtualization, walled gardens, and limited separation through a variety of mechanisms that allow access of corporate and personal data on the same device. Although these approaches are complex, they provide viable mechanisms for dealing with the inevitable mix of corporate and personal technologies, letting the enterprise reap the productivity benefits offered by BYOD while effectively mitigating risks and threats.

The goal is to give users the information they need, on the devices they use, from wherever they work. That is the consumerization trend driven by the macro forces of societal and technological shifts.

The world has changed. This new world of IT provides interesting opportunities and challenges for organizations and IT professionals to deliver industrial grade “SAP-reliable” services while keeping them “Google fast” and “Apple simple.”

Reference


Seth Earley is CEO of Earley & Associates (www.earley.com). He’s an expert in knowledge processes and customer experience management strategies. His interests include customer experience design, knowledge management, content management systems and strategy, and taxonomy development. Contact him at seth@earley.com.

Robert R. Harmon is a professor of marketing and service innovation and Cameron Research Fellow at Portland State University. His interests include IT-enabled service innovation, technology marketing, and big data marketing. Contact him at harmonr@pdx.edu.

Maria R. Lee is an associate professor in the Department of Information Technology and Management at Shih Chien University, Taiwan. Her research interests include social media analytics, e-commerce, and knowledge management. Contact her at maria.lee@g2.usc.edu.tw.

Sunil Mithas is a professor at the Robert H. Smith School of Business at the University of Maryland. His research focuses on the strategic management and impact of information technology resources. Contact him at smithas@rhsmith.umd.edu.

Selected CS articles and columns are available for free at http://ComputingNow.computer.org.
COMPSAC is the IEEE Signature Conference on Computers, Software, and Applications. It is a major international forum for academia, industry, and government to discuss research results, advancements and future trends in computer and software technologies and applications.

With the rapidly growing trend in making computation and data both mobile and cloud-based, such systems are being designed and deployed worldwide. However, several challenges arise when they are applied to different domains or across domains. COMPSAC 2015 will provide a platform for in-depth discussion of such challenges in emerging application domains such as smart and connected health, wearable computing, internet-of-things, cyber-physical systems, and smart planet.

**Important Dates:**
- January 17, 2015: Paper submissions
- March 15, 2015: Paper notifications
- April 28, 2015: Camera-ready version and registration due

Contact ITiP Symposium organizers at cs_compsac-ITiP-2015@iastate.edu

**CALL FOR PAPERS**

July 1-5, 2015
Tunghai University
Taichung, Taiwan

This new COMPSAC symposium on IT in Practice (ITiP) is directed towards the interests and needs of IT practitioners, application developers, IT managers, and CXOs from business, industry, and government, as well as academics and researchers. The symposium, co-sponsored by IT Professional, is an interactive forum for sharing and gaining valuable insights and experiences and for active engagement among technical professionals around the globe. It aims to inspire professionals, executives, and researchers to focus on diverse elements necessary to successfully architect, design, engineer, configure, and field a range of IT applications in several domains. Topics of interest include the following:

- Innovative & challenging IT applications
- Deploying and managing large scale IT systems
- Real world issues, concerns & challenges
- Development & operational experiences
- IT application design & validation
- Emerging trends in enterprise IT
- Strategic alignment of IT
- IT workforce & IT workplace
- Managing IT in practice
- Opportunities & challenges of C-level positions
- Operational issues & insights
- Crowd sourcing problems & solutions
- The research-practice nexus
- IT-enabled business innovation
- Case studies
- Future perspectives

Authors are invited to submit original, unpublished research work and novel computer applications in full-paper format. Simultaneous submission to other publication venues is not permitted. The review and selection process for submissions is designed to identify papers that break new ground and provide substantial support for their results and conclusions as significant contributions to the field. Submissions will be selected that represent a major advancement in the subject of the symposia to which they are submitted. Authors of submissions with a limited contribution or scope may be asked to revise their submissions into a more succinct camera-ready format; e.g., workshop paper, fast abstract, or poster to the main conference.

Contact ITiP Symposium organizers at cs_compsac-ITiP-2015@iastate.edu

**COMPSAC Sponsors:**

- IEEE
- IEEE Computer Society
- IEEE Cloud Computing
- Information Processing Society of Japan
- Korean Institute of Information Scientists and Engineers
- Tsinghua University

**COMPSAC Technical Co-Sponsors:**

Information Processing Society of Japan
Korean Institute of Information Scientists and Engineers

**COMPSAC 2015 Local Host:**

Tunghai University
Taichung, Taiwan
Consumerization in the IT Service Ecosystem

Enrique Castro-Leon, Intel

Most available literature on IT consumerization focuses on current practices. Exploring the driving forces behind consumerization is essential, however, if IT organizations are to make sense of the current dynamics of transformation and disruption and formulate effective strategies in response.

IT consumerization—also known as bring your own device (BYOD)—refers to the practice of employees using mobile devices (such as smartphones, tablets, or laptops) for work activities in the workplace, at home, or while traveling. BYOD can also involve bring your own app (BYOA), and both practices are transforming IT by redefining its role in the enterprise. IT, once a centralized technology purveyor, assessor, and gatekeeper of information, is now becoming a facilitator in a more federated environment. Consumerization has brought an inversion of roles, where now users are driving technology adoption and change.

This transformation hasn’t come without challenges, but it presents new opportunities for IT departments to address requirements for increased efficiency and nimbleness. BYOD practices can also help organizations adapt to cultural change and address the needs of the millennial generation, which expects devices to work and deliver value in a corporate setting just as in a personal setting. IT consumerization is thus a topic of interest to CIOs, CTOs, and solution architects alike.

However, most available literature focuses on current practices—operational aspects or the disruptive effects of BYOD policies. Exploring the driving forces behind consumerization is an essential exercise to make sense of the current transformation and disruption dynamics and ensure that IT continues delivering value to the enterprise. Here, I briefly review two such forces—the changing relationship between workers and employers and the redefining of enterprise
boundaries, both of which have been amply documented. Then, I delve further to identify a third, more fundamental force behind this transformation—the move to a service-oriented economy.

Revising the Worker-Employer Relationship

The first force driving BYOD policies is the changing relationship between workers and employers. The enterprise workforce has become more diverse, especially in the wake of the financial sector’s 2008 crisis. In the last six years, there’s been an increase in the number of transient staff—that is, contractors from workforce augmentation agencies and service providers, such as Volt and Kelly Services, as well as independent professionals and part-time employees. The number of full-time employees—the traditional audience for IT services—is getting proportionately smaller, aided by the parallel trend of cloud computing, which is the outsourcing of infrastructure.

Members of this more transient workforce bring their own tools. Relatively short tenures mean that it’s not practical for these professionals to adopt the company’s existing tools and processes. Temporary teams might bring tools they’re comfortable with—such as tools for project tracking and management or customer-relations management (CRM)—and use them outside the IT department’s purview. Because establishing a local infrastructure wouldn’t be practical, such tools are usually based on software as a service (SaaS) and accessed through mobile devices.

Even when corporate applications and data are involved, employees entering the workforce, especially temporary employees, place less value on corporate-issued tools and would rather use their own devices—particularly if they’re temporary employees. This puts an enormous amount of pressure on IT departments to address these requests and support personal devices.

Redefining Enterprise Boundaries

The second force driving the consumerization of IT is the rapidly shifting definition of enterprise boundaries. Today, large corporations, such as aircraft and automobile manufacturers, have effectively become gigantic systems integrators managing complex supply chains. Smaller companies follow the same dynamic to a lesser degree or participate in various supply networks. A highly diverse community thus routinely accesses corporate applications and data. When application components or whole applications, such as email messaging, get outsourced to the cloud, the notions of what’s inside or outside the enterprise become blurry, with corporate processes using resources beyond the traditional enterprise perimeter.

In this environment, the diverse user community isn’t concerned about distinguishing between internal and external resources; users expect to use the devices, data, and tools in the same way they always have. Effectively, IT no longer has control over devices that connect with corporate data and applications. A heavy-handed approach to managing BYOD practices will likely alienate users, who will ultimately sidestep IT altogether. This quandary has no simple solutions, but it’s important in understanding the dynamics behind this move to consumerization.

Transitioning to a Service-Oriented Economy

Both of the forces just described are in fact being driven by a third, more fundamental transformation in the economy—the long-term and apparently irreversible transformation of the industry moving away from a manufacturing- and product-centric focus. With globalization, the world economy has instead become service-oriented. For all advanced economies, the economic value from financial, travel, legal, healthcare, and other services exceeds that of manufacturing, agriculture, and extractive industries.

If we compare the GDP service component with all other sectors as a metric, many countries made
CONSUMERIZATION OF IT

Invented in 1989,12 the Web worked because of three essential technology components: HTML, URIs/URLs, and HTTP, which allows information exchange across the Web. Five years later, the Web had been adopted across the world. The Web’s value is in providing a universal, compatible human-to-machine interface that allows access to data and applications across the Internet regardless of differences in machine architecture.

The next step in technology evolution occurred with the technical community’s realization that the Web was good not only for human-to-machine communication, but also for interoperable machine-to-machine communication. This realization led to the development of Web services technology in the early 2000s. The services came in two flavors: Simple Object Access Protocol (SOAP) and Representational State Transfer (REST).

Finally, around 2007, service networks emerged, enabled by the ubiquitous adoption of the cloud. BYOD is an example of a service network application.

The gold ovals in Figure 1 represent successive milestones in technology evolution, with enabling hardware technologies shown in the blue boxes on the left and key actors that played a role in the service ecosystem’s evolution.

**The Beginnings of Service-Oriented IT**

The IT industry hasn’t been exempted from the transition to a service economy, but the realignment started relatively late. The transformation to service-oriented IT, out of which consumerization is a prime manifestation, can be traced to three historical technology transitions:

- the invention of the World Wide Web,
- the development of Web services technology, and
- the emergence of service networks.

Figure 1. The evolution of the service ecosystem, with main milestones captured in the gold ovals. The left side shows the enabling hardware in blue, while the right shows in green the key actors that played a role in the service ecosystem’s evolution.
significant context and configuration before any information exchange was possible, introducing considerable deployment overhead and restrictions. In order for these systems to work it was necessary to compile all the source code together or at least require that communicating entities within the system use a narrow range of software versions, possibly with components from the same vendor. This state of affairs was obviously unrealistic for Internet applications.

Systems that support late binding are said to be **loosely coupled**. The traditional, strongly coupled application architectures mentioned earlier must be assembled using a set of compatible components. A new revision or a customization of an application might require a different set of components, which might be functionally equivalent but unable to function in both the old and the updated environments. It’s also possible that one component can support only one instance of the application, and therefore must be replicated for each application instance. This approach created unnecessary replication, bloat, and complexity due to incompatible versions. The availability of loosely coupled Web services allowed developers to factor out common functions, which run as self-standing services. An application needing the capability simply invokes that particular service.

In the initial state, in which strongly coupled architecture was the norm—until early in the decade of 2000—applications ran in stovepipes. The left side of Figure 2 shows this initial state, with three applications (A1, A2, and A3) that each run a separate copy of the application on dedicated hardware. With loosely coupled Web services, capabilities common across multiple stack instances are abstracted out in the form of a service, implemented once, and accessed as a service. The duplicated stacks characterizing tightly coupled, stovepiped applications morph into a graph, with each node representing a coalesced capability delivered as a service. The stovepipe “forest” morphs into a service network (the right side of Figure 2). Also, the enterprise perimeter shrinks with fewer application instances, but it also encompasses resources deployed in the cloud. Finally, the re-architected application must support multimodal access.

Each network node is a service component, or *servicelet*, which is a self-standing application component accessible through a URI and discoverable.
CONSUMERIZATION OF IT

through the Universal Description Discovery and Integration (UDDI) registry using the Web Services Inspection Language (WSDL).

As Figure 2 shows, the three stack layers have been re-architected as generic internal services S2, S3, and S4 that can be sourced externally (shown as S6, S8, and S9). Application A1 has stringent security requirements and gets deployed with servicelets running in an internal (private) cloud. Application A3 carries more relaxed requirements and gets implemented as a software-as-a-service (SaaS) application using third-party sourced servicelets. Application A2 is somewhere in between.

Servicelets are intended to be components and thus might not be fully formed applications. Today, servicelets typically run in a virtualized environment hosted in a pool of servers. Operators can rebalance workloads across this pool by managing how they’re launched or even by moving virtual machines around according to specific performance policies.

**Service Component Externalization and Composite Applications**

Applications created from service components are called composite applications; the architectural style for building these applications is known as a service-oriented architecture (SOA). As defined in the mid-2000s, SOA’s primary goal is to radically reduce costs in IT delivery to the enterprise by eliminating duplicate or redundant functions. Specific functions are implemented just once as a service, instead of instantiating a new copy for every application. An example is Intel IT’s Common Directory Information System (CDIS), which is a repository of employee data. Applications that use employee data—such as payroll, expense reporting, and the employee directory—link to the CDIS.

Large organizations were the primary SOA adopters. Because URIs are truly universal handles, it matters little whether the resource lives inside or outside of the organization. IT departments thus realized that it wasn’t always desirable or cost efficient to host generic services in house, especially when such services were unrelated to the organization’s line of business (as with email, messaging, and even storage services).

The cloud’s emergence around 2007 was timed perfectly to allow developers to realize the benefits of service networks. A little-documented side effect of the transition to composite applications is that, while service components might have begun as in-sourced service components within enterprise boundaries, in reality, there were no fundamental reasons why they couldn’t be outsourced. Applications such as payroll, expense reporting, and job postings could be conveniently outsourced to third parties who could take advantage of economies of scale to run the application at a lower cost than if it were hosted in house.5

**From Service Networks to Consumerization**

The externalization drivers just described were purely economic: third parties offered servicelets at a lower cost than the in-sourced equivalent due to two factors. First, as hinted at earlier, providers specializing in a specific type of service could develop skills and knowledge to build and deliver these services more efficiently than corporate IT departments. Second, a virtualized cloud infrastructure let IT amortize physical assets over a larger customer base, which allowed economies of scale well beyond that of a corporation’s few internal clients. However, the involvement of multiple parties has security and quality of service implications; solutions to such issues are still works in progress (and beyond this article’s scope). Still, the economics of composite applications are so compelling that these issues haven’t deterred their deployment in areas in which the risks are manageable.

The synergy between cloud and service networks manifested in multiple back and forth movements in offered capabilities. The first, as I described earlier, was the externalization of servicelets. A second manifestation was in third parties offering fully formed applications. Pioneering examples here include Web-based mail applications (such as Google Mail and Yahoo! Mail) and Salesforce.com’s CRM application. Figure 1 (center) shows how legacy applications for large IT or corporate IT were re-engineered as composite applications; initially, the applications used internal servicelets, but they eventually included third-party servicelets as well when economics justified the replacement. As the industry has experienced the learning curve with composite applications, economies of scale have been attained that have extended composite applications’ reach to audiences without the technical
savy or deep pockets of large corporations. This has gradually delivered service networks’ value to small and medium businesses (SMBs) and individual consumers, ensuring their applicability for small-scale IT and personal IT and leading to IT’s democratization, as Figure 1 shows.

Servicelets have reached individual consumers in the form of the thousands of applications available through Google Play, the Apple App Store, and portals such as Dropbox. In this environment, sales managers familiar with the Salesforce.com CRM continue using their own accounts with their teams, and team members use their personal tablet smart devices to link to their manager’s account. Assuming the IT department allows such practices, BYOA and BYOD is at this point achieved.

As Figure 1 shows, the appearance of API manager portals facilitated this transition to BYOD and BYOA. The underlying dynamic is reverse externalization, in which in addition to using external servicelets, organizations realize that exposing certain corporate applications externally—both as servicelets and through self-contained Web browser access—allows them to activate additional revenue streams or increase their marketplace presence. Examples include travel reservation systems, supply chain networks, and shopping networks such as Amazon or eBay, which have become the marketplace for a vast network of sellers and resellers.

Such capabilities are accessible to third-party developers through API manager portals that make it relatively easy to build mobile front ends to cloud capabilities. Front-end services have also gone through a notable evolution, in which traditional PC Web access has been augmented to enable application access through mobile devices. API managers become new integration points of formerly internal functions hosted through cloud service providers and hosting software from independent software vendors. Mobile app developers become the new integrators.

“Native” servicelets are intended to support cloud-based composite applications. In practice, however, it’s possible to “servitize” legacy applications, retrofitting them as servicelets by slapping an API on top of the application. Conversely, software designed as a servicelet can be made to behave as a self-standing application through a thin GUI that calls the servicelet’s API (see the right side of Figure 2).

The bottom line, as Figure 2 shows, is that the re-engineered application must support multimodal access—that is,

- traditional access by corporate employees,
- BYOD and BYOA access by the general public, and
- app developers interacting through the company’s published interface in an API portal.

Once servicelets “escaped” into the wild as third-party offerings available to any developer, new business opportunities surfaced and developers started building applications that previously would have been created inside the enterprise. Because the ecosystem involved is essentially the whole computer industry, including corporate and individual consumers, the rate of evolution has accelerated quickly. Most new applications don’t remain in use, but the ones that survive can develop a customer base of hundreds of thousands in just a few months.

Consumerization developed in this context, with a profound economic effect: expensive data centers costing hundreds of millions to billions of dollars delivered applications accessible to consumers for at most a few dollars per month. The audience for these applications grew from a single corporate customer wanting a complete solution stack built in-house to a few hundred large corporations, which then grew to a few tens of thousands of small and medium businesses (SMBs) and then tens of millions of individual users. Under the consumerization paradigm, individual users became familiar with these applications and brought them back to the enterprise, completing a full circle.

A strong contributing factor here is the preferences of the millennial generation, which is just entering the workforce and grew up in this environment. From this perspective, it makes perfect sense to continue this modality at work, using personal devices—whether a mobile device or a PC client—to connect to the cloud-based applications behind them. This is what brings us to the notion of BYOD and BYOA today.

What about tomorrow? BYOD is primarily (and once again) human-to-machine interaction. There are inklings of a new circle starting with machine-to-machine interactions in the...
CONSUMERIZATION OF IT

form of an Internet of Things—that is, devices such as printers, Web cams, and drones acting as network nodes. Outsourced printers are becoming ubiquitous, with printers deployed on premises, managed by a service provider as a service, with the customer paying for the page count.

Changes brought by IT consumerization will likely redefine IT’s role. What we see now is actually the proverbial tip of the iceberg, driven by long-running trends of economics and technology evolution. The transition to a service economy in advanced economies is fundamentally changing relationships between organizations and workers and, in the process, transforming how IT is delivered to constituents. Consumerization also reflects the blurring of enterprise boundaries, in part due to these changing relationships.

So, do IT departments risk irrelevance, becoming the caretaker of an ever-shrinking core of enterprise capabilities while business units meet their own technology needs through third-party services? A risk, yes. A predestined outcome? Not necessarily, given the appropriate strategy. Outcomes from prior crises have demonstrated that IT can transform itself and deliver better business value to the enterprise than a fee-for-all alternative. This occurred in the early 1990s, when the glacial rate of development and deployment of mainframe-based applications led business units to started running spreadsheets on shrink-wrapped software outside of IT’s purview. At that point, the industry adopted client-server and three-tier architectures on volume servers. A second crisis was the crash in demand after the Y2K crisis ended. IT’s response was higher efficiencies through a restructuring of applications behind the service paradigm and taking advantage of cloud technology. Consumerization is a reflection of these economic efficiencies now reaching individual consumers, who might also be employees.

Finally, consumerization must be recognized as a strategic trend; looking at it solely from a tactical approach—as end points to be managed and regulated, for example—would be ineffective. Organizations have an opportunity to rethink current processes and look for new value vectors and partnerships with the business. For example, with increasing process maturity, BYOD and BYOA instances can be integrated into current Business Process Model and Notation (BPMN) and Business Process Execution Language (BPEL) activities to determine how they dovetail into existing workflows, providing objective criteria to distinguish useful applications from the frivolous. Today’s state of the art doesn’t allow an implementation of this capability.

Rethinking current technology adoption strategies and putting retraining plans in effect won’t be an easy path, but it will allow IT departments to continue delivering value to the enterprise—and therefore ensure its continued relevance.

References


Enrique Castro-Leon is an enterprise architect and technology strategist with the Intel Data Center Engineering Group. His research interests include enterprise IT solution integration, cloud and service engineering, and internal processes for organizational transformation under emerging service paradigms enabled by cloud computing. He is the lead author of The Business Value of Virtual Service Grids: Strategic Insights for Enterprise Decision Makers (Intel Press, 2008) and Creating the Infrastructure for Cloud Computing: An Essential Handbook for IT Professionals (Intel Press, 2011). Castro-Leon has a PhD in electrical engineering from Purdue University. Contact him at enrique.g.castro-leon@intel.com.
This survey presents the pitfalls of security protection in online social networks and identifies common attack methods. A harmless proof-of-concept malware app demonstrates the vulnerability of online social networks and the significance of the user’s mentality.

Online social networks let users keep in touch with families, friends, and colleagues in an easy and convenient way, but they also present various security risks. Large-sale online social networks have become very popular owing to advances in network and mobile technologies.\textsuperscript{1,2} Their popularity has, in turn, drawn the attention of attackers and intruders, who want to exploit this large population to spread malware or steal personal information. Some online social networks, such as Facebook, LinkedIn, and Google+, require users’ real identity. Online social networks also preserve a great amount of users’ confidential and financial information, including credit card information.

There are two major types of attacks in online social networks. The first is accomplished by exploiting security loopholes in the networks, and the second is accomplished by abusing trust among users. The second type can sometimes be easier to accomplish, because humans aren’t always cautious. We focus here on this type of attack, discussing the pitfalls of the security in social networks and popular attack methods that exploit the user’s trust. We also present a proof-of-concept Facebook app to demonstrate the risks in social networks.

**Security Pitfalls in Social Networks**

Online social networks involve many services, stakeholders, user behaviors, and business

Kaze Wong, Macao Polytechnic Institute
Angus Wong, The Open University of Hong Kong
Alan Yeung and Wei Fan, City University of Hong Kong
Su-Kit Tang, Macao Polytechnic Institute
considerations, resulting in a complicated security paradigm with the following pitfalls.

Confusing and Complicated Settings
Social network companies are looking for ways to make money. They’re introducing new products and services, such as video advertising, multimedia streaming, and automated photo tagging, and are trying to expand payment services. They’re also attempting to debut mobile tools and apps to exploit the fast-growing mobile market.

Such endeavors might make the social network look more versatile and state of the art, but they might also complicate the network and make it more difficult to ensure there’s a unified strategy for controlling privacy and security.

Convenience over Security
Enforcing security measures can result in higher operational costs and lower performance, but social network providers are likely even more concerned with any inconveniences the security measures cause in terms of users sharing content or developers distributing apps. Such inconveniences can hinder attempts to obtain a bigger market share.

Consequently, the default privacy settings in social networks are commonly set to a level favorable for sharing (and thus less secure).

Friends’ Influence on Privacy Control
Your data’s security also depends on the security settings of the friends with whom you share content. For example, even if you have a high security setting, if you share your content with friends, and one of your friends is hacked, your data will be exposed to the attacker. Thus, the strength of your security level is as weak as that of your friend with the lowest level of security.

Inferred Information
Even if you don’t disclose your information, it can be inferred by analyzing your and your friends’ shared content. For example, suppose that you haven’t entered any personal information in your profile. However, if your friends tell Facebook that they are from college ABC, and in many of your group photos, you’re with those people at the college, then you’re probably from the same college. Ensuring privacy in social networks can sometimes require more than just a technical solution.

New Incentives for Attackers
User data is essential for many business opportunities, which presents incentives for various attacking or intrusion activities. The data—user profiles, posted text messages, shared pictures, or “like” actions—can be harnessed to help advertisers better market their products.

Various apps and people thus use different approaches to try to collect as much data from you as possible. One way to achieve this legitimately is to have your permission. In fact, an app that looks like an interesting game might be designed with the primary purpose of collecting your data. When installing such an app, you might have given the “game” permission (intentionally or unintentionally) to access your profile, albums, and friends list.

Common Attack Methods
Here, we introduce possible attack methods for online social networks, which we used in our proof-of-concept Facebook app.

Related Work in IT Pro

The strength of your security level is as weak as that of your friend with the lowest level of security.
Malware
Malware is malicious software that appears in the form of codes/scripts or other kinds of software. Hackers can use it to collect sensitive user information. Hackers can spread malware in online social networks by sharing malicious URLs that direct users to third-party websites, where the users are asked to download a (malware) file. If a user downloads and executes the file, he or she will become infected and will in turn send the same malicious URLs to his or her friends. For example, a malware named “Ice IX” has stolen credit card information from many Facebook users.3

Identity Forgery
Identity forgery happens when an attacker accesses someone’s private information and uses it to pretend to be that person. It occurs not only in real world but also in online social networks.9

Excess Permissions
People can access online social networks using smartphone apps. Before installation, these apps ask users for certain permissions, such as access to contacts or photos stored on the phone, although some permissions aren’t actually necessary for the provided function.

Another form of excess permission is requested by third-party apps on online social networks. For example, people can use third-party apps on Facebook to play games or share photos and videos. Some apps request access to the personal data of the user’s friends or request permission to post messages on user’s behalf. This can be dangerous if the apps use the data in unlawful ways, and it can lead to privacy leakage.

Our Proof-of-Concept Facebook App
In our experiment, we first created a proof-of-concept Facebook app, which is a harmless malware. Our app is a social game bonus collector, which automatically collects game bonuses shared by friends of the user. By default, after installation, a Facebook app can access a user’s name, gender, networks, and friend list. Our app requests three extended permissions from the user:

- **email**—permission to obtain the user’s email address;
- **read_stream**—permission to read all posts (status updates, pictures, or links) in the user’s newsfeed; and
- **publish_actions**—permission to, on the user’s behalf, publish posts on the user’s wall, comment on others’ posts, and “like” posts published by others.

(See [http://developers.facebook.com/docs/reference/login/extended-permissions](http://developers.facebook.com/docs/reference/login/extended-permissions) for more information.)

The app seemingly has two main functions: analyzing the newsfeed and present a tutorial. By reading the user’s newsfeed, the app can filter game bonus posts and extract the links to claim the bonuses au-
automatically. After claiming the game bonuses, the app posts to the user's wall to advertise the app to others. A video tutorial, which redirects users to a video-sharing website, teaches people how to use the app.

However, in reality, the app is implicitly performing additional tasks. It accesses the user's friend list to construct a social graph and analyze the various relationships. It also reads the user's newsfeed to analyze not only the user's posts but also his or her friends' activities. Finally, the “watch tutorial” function redirects users to a bogus website that looks similar to a well-known corporate website, where users are asked to upgrade a browser plug-in before watching the video. The download link then points to a malicious file.

Figure 1 shows the system architecture of our proof-of-concept program, in which we’re acting as an attacker. The server is fully controlled by us (the attacker). The Web server (www.ni-hacking.com) is used to host a fake website that can do phishing and spread malware. The application server is used to host our Facebook app program, which we refer to as the Facebook app engine.

Operations
Here, we outline the operations of our proof-of-concept experiment.

First, we post a link to the app on our Facebook walls, recommending the app to our friends (see Figure 2a). When our friends see it and click on the link, the installation process starts. During the process, some access permissions are requested (see Figure 2b).

In our app, users can choose the “analyze newsfeed” function, as shown in Figure 2c. The function will be provided to users, but meanwhile, our app can secretly reads all posts from the users and stores them in our (attacker) server for data analysis.

Or, if users choose the “watch tutorials” function, they’re redirected to a fake YouTube site, where they’re asked to upgrade a browser plug-in to watch the video tutorials (see Figure 2d). When the users click on the download link, the system links to a malicious malware. (In our system, we just link to the download page of Adobe Flash so that users aren’t actually attacked.)

In our proof-of-concept Facebook app, the five attacking methods discussed earlier were or could be used. The malicious app, with excess permissions granted by users, can

- obtain users’ email address so attackers can spam them,
- post to users’ wall on their behalf so the app can attract other users, and
- read all posts in users’ newsfeed for information analysis.

In addition, the fake YouTube site can lead to phishing and botnet attacks.

Results and Discussions
Our experiment shows how privacy can easily be violated:

- user feeds can be obtained and harnessed;
- friends lists can be obtained to perform social network analysis or to build a larger pool of potential victims;
- phishing can be done by redirecting to a bogus site to further steal users’ data; and
- malware can be downloaded to affect users’ computers.

As of this writing, 276 people have installed our app. Of those, 97.1 percent have selected the “analyze newsfeed” function, and 27.20 percent have downloaded the potentially malicious software. Furthermore, based on the friends lists secretly taken by our app, we could build a social graph containing 19,763 people, representing a large pool of potential victims. More importantly, we could analyze...
this social network to identify the hub (the most important person), allowing for better marketing or more targeted attacks.

Our experiments have shown that people grant permissions without always understanding what an app can do with them. Users granted permission to our app mainly because they thought the app recommended by us should be trustworthy. Clearly, trust can be abused, creating opportunities for attackers. What happens if someone steals your identity and posts a malicious app to your wall?

By taking a few countermeasures, users can avoid attacks such as the one outlined in our experiment. Before installing an app, users should carefully check what permissions the app requests. If unnecessary or excessive permissions are requested, users shouldn’t install the app.

Furthermore, users should

- be mindful when the app is trying to redirect them to another site, because bogus sites can be well disguised;
- validate the URL to avoid accessing phishing sites (in our app, the URL of the redirected video site is www.ni-hacking.com, so even though the website looks like a YouTube site, users should be able to identify it);
- review their privacy and security control settings periodically, because social networks often add new features and services that can introduce loopholes and opportunities for attackers; and
- periodically check the information that third parties are accessing.

Users tend to trust the content of requests in social networks, due to their inherent trust of their friends and the social network.

**Figure 2.** Screenshots of our proof-of-concept Facebook app: (a) posting an app recommendation to the wall, (b) the malicious app requesting some permissions, (c) the app reading all posts in the users’ newsfeeds while “analyzing” the newsfeed, and (d) a user being redirected to a fake YouTube site.
platform. However, users need to understand that attackers can easily exploit this trust, despite the technical aspects of their security settings.

Acknowledgments
This work was supported by the Macau Science and Technology Development Fund 039/2010/A.

References

Kaze Wong is a researcher at the Macao Polytechnic Institute and a graduate student in electronic and information engineering at the City University of Hong Kong. He has finished a number of large-scale government-funded projects on Internet security. His research interests include network security, complex networks, queuing systems, Internet computing, and mobile computing. Contact him at kaze.c.y.wong@gmail.com.

Su-Kit Tang is a lecturer at the Macao Polytechnic Institute, Macau. He is also affiliated with the IPv6 Network Research Laboratory in the institute. His research interests include the security of IPv6 network protocol operations, trust and privacy issues in online social networks, and performance improvement in ad hoc routing using network coding. Tang has a PhD in computer science from Sun Yat-Sen University, China. Contact him at sktang@ipm.edu.mo.

Wei Fan is a doctoral student in the Department of Electronic Engineering at the City University of Hong Kong. Her research interests include complex networks and social network analysis. Fan has an MPhil in electronic engineering from the City University of Hong Kong. Contact her at fanwei.fw@gmail.com.

Alan Yeung is an associate professor in the Department of Electronic Engineering, City University of Hong Kong, as well as a consultant in computer networking and communication systems for government and corporate clients. His research interests include network infrastructure security, mobile communication systems, and Internet caching systems. Yeung is also active in providing consultancy services to local industry in the areas of computer networking and communication systems. His clients include government agencies and listed companies. Contact him at eeyeung@cityu.edu.hk.

Kaze Wong is a researcher at the Macao Polytechnic Institute and a graduate student in electronic and information engineering at the City University of Hong Kong. He has finished a number of large-scale government-funded projects on Internet security. His research interests include network security, complex networks, queuing systems, Internet computing, and mobile computing. Contact him at kaze.c.y.wong@gmail.com.

Angus Wong is the program leader of the Engineering Sciences team in the Open University of Hong Kong. He has undertaken a number of network-related projects for the government. His research interests include Internet systems, network infrastructure security, and social network analysis. Wong has a PhD in IT from the City University of Hong Kong. Contact him at akywong@ouhk.edu.hk.

Selected CS articles and columns are available for free at http://ComputingNow.computer.org.
For decades, researchers have been experimenting with real-time multimedia services with 3D technology support. Such services have been used in specialized contexts—such as in the entertainment industry, for teleoperation, and in critical applications—but have mostly relied on special-purpose hardware. Given recent advances in 3D technology and the increasing diffusion of low-cost 3D devices in the consumer market—including 3D TVs, mobile phones, and gaming devices—wouldn’t it be great if we could use these ordinary devices to build low-cost 3D communication systems?

We started this line of research in 2012, motivated by the need to add a low-cost stereoscopic remote-control module in a research project developing an open source framework for real-time teleoperation. In fact, researchers have demonstrated that teleoperation is an application area that stands to benefit greatly from stereoscopic vision (see the “Related Research in 3D Remote Control” sidebar). 3D images can enable better perception of depth characteristics in the environment—especially in terms of the relative object distance—thus enhancing precision and reducing the time needed to complete tasks involving remote robot piloting and manipulation. Of course, for each type of application, it’s necessary to verify that the achievable quality-of-experience—which might be bounded by the limited capabilities of consumer-grade hardware—meets the application constraints and user expectations.

Our goal was to study and design a 3D teleoperation prototype offering good performance at a low cost.
Related Research in 3D Remote Control

The use of 3D vision for teleoperation is a very active research field, because it has been shown that 3D viewing technologies might provide users with higher depth perception. As expected, some tasks benefit more than others from a better comprehension of distance, such as teleguide and obstacle localization. However, even in demanding situations such as in emergency applications, teleoperation with 3D vision typically improves remote-control performance.

John McIntire, Paul Havig, and Eric Geiselman have reviewed the literature investigating human factors that have implications on task performance when stereoscopic 3D displays are used. The tasks include judging absolute and relative distances, finding and identifying objects, navigating and manipulating objects in terms of position, and performing orientation and tracking.

Researchers have also investigated acquisition and visualization technologies. For example, work has shown that, using efficient algorithms, correctly calibrated stereoscopic images provide a guaranteed depth perception. Other work has assessed the performance achieved using different viewing technologies during remote-control operations, providing results in terms of usability while comparing 3D and 2D viewing, 3D and virtual-reality displays, and robot sensors.

Studies have also addressed the problem using simulation. For example, researchers performed a pick-and-place task in a simulated virtual environment using three different systems: multiple 2D displays, a 3D perspective display, and a 3D stereoscopic display. Although the gain in performing simple tasks was limited, introducing a stereoscopic display resulted in better performances than the perspective display for highly difficult tasks.

Finally, Salvatore Livatino and Giovanni Muscato were interested in the effect of video feedback latency, so they performed usability studies that substituted real video images with synthetic images from robot laser data. The authors observed that by using a technique to minimize the amount of bandwidth (and delay) required for the transmission, they could significantly reduce the amount of time spent performing a given task, thus improving system usability.

References


COTS and Open Source Solutions

To reach our goals, we focused our attention on consumer off-the-shelf (COTS) hardware and open source software. COTS components cost much less than dedicated hardware, but they can have (slightly) lower performance.

The scientific literature and open source domain offer many algorithms and reference implementations for problems in the fields of communication, visualization, stereoscopy, and so on. The downside of open source versus commercial software again relates to performance. However, with suitable modifications and component selection, open source software offers...
good-performance, low-cost solutions that the final installer can customize to address the specific implementation and its goals.

Furthermore, while designing our system, we focused on simplicity of use and customization. We also wanted to create an open-platform-based design and use widespread standards for coding and communication. Although our current solution is only a prototype, we hope it stimulates further research and investments in this research area.

Architecture Design
A generic structure of a teleoperation system based on real-time 3D video feedback has two main parts: a mobile streaming node equipped with a stereoscopic camera, and a client application for 3D visualization (see Figure 1).

The transmitter node is located on the teleoperated machine so that the capture device can record the vision of the area from the device’s point of view. The system’s core must process the captured images so that a timely robust transmission can occur through a wireless channel to the remote operation side. On the receiver side, the software reconstructs and renders the 3D image from the received data on a graphic display device in a format suitable for stereoscopic vision. The key point for successful system operation is to guarantee low end-to-end latency, which is a fundamental requisite to achieve a good real-time interaction experience.1

Issues with a Consumer-Level Solution
The deployment of consumer-level solutions for remote-control 3D video systems is made difficult by constraints that relate to the high cost of industrial-grade components and to the use of proprietary, closed systems. As noted earlier, to enable consumer-level solutions at a reduced cost, we investigated the feasibility of using COTS hardware components, integrated using open source software frameworks. In doing so, we prioritized the following issues.

First, we wanted to find a mobile device with suitable resolution and highly responsive output of the captured images. For this reason, we excluded still-picture cameras. We also wanted to reduce the video encoding latency using appropriate encoding algorithms.

Next, we wanted to make the system sufficiently robust to some packet losses—an event possible in any communication scenario that can severely affect the user’s perceived video quality. As we will see later, suitable choices of encoding algorithms and related configurations can help address this issue.

Finally, we wanted to render the stereoscopic video using a technology widely supported in an open source environment.
Component Selection and Evaluation

The previous considerations led us to develop the building blocks of our proposed teleoperation system based on real-time 3D video feedback (see Figure 2 and the related video at http://youtu.be/5zApDO3wSq4, which also shows the components).

Our prototype includes the following five key components.

**A stereoscopic camera.** We used a commercially available 3D mobile phone (HTC Evo 3D) that let us achieve, at a moderate cost, stereoscopic vision, portability, a battery-operated power supply, and wireless communication. In field experiments, we experienced a latency between 150 and 200 ms in the best conditions. This is slightly higher than the ideal requirement (less than 150 ms), but it’s probably the most efficient solution that the COTS market can currently provide. In fact, this loss of performance is the price to pay for keeping the overall system cost low.

We also evaluated other solutions, such as 3D cameras, but the majority of them either didn’t let us transmit video in real-time, or the format (such as the Digital Video format) required further processing and thus increased system cost, complexity, and latency.

**Encoding and transmission software.** We developed a custom-built open source encoding and transmission software specifically for the HTC Evo 3D phone to minimize encoding and transmission latency.

**Networking hardware.** We used standard networking hardware—a commercially available 802.11 access point for transmission over a standard Wi-Fi channel.

**Receiver and decoder software.** We custom built open source receiver and decoder software on top of the Linux OS and openGL libraries.

**Stereoscopic hardware support.** We used a standard desktop computer with stereoscopic hardware support, including a graphics board, a monitor, and an infrared emitter device for the active glasses. The identified solution for supporting hardware openGL is based on an Nvidia Quadro 4000 graphics board, which provides a direct connection from the video card to the infrared emitter device without the need for specific driver support in Linux (see the specifications at www.nvidia.com/object/product-quadro-4000-us.html). This greatly simplified the management of stereoscopic issues in the receiver but also increased the final system’s cost.

---

**Figure 2.** System architecture and components. A mobile phone with stereoscopic capabilities wirelessly transmits to a computer that can render 3D video using active glasses. The total cost of the system is approximately US$1,500.

**Bill of materials:**

1. HTC Evo 3D: $480
2. Wi-Fi router: $60
3. Nvidia Quadro 4000: $700
4. Nvidia 3D vision kit: $90
5. Stereoscopic monitor: $200

Total: $1,530
Alternative Solution
Because Nvidia doesn’t officially support 3D visualization for the Linux environment, we could have implemented a cheaper solution (for approximately $900) using an experimental driver that’s publicly available on the Internet (http://sourceforge.net/projects/libnvstusb). The driver lets you work with much cheaper video cards by synchronizing the infrared emitter using the USB connector. However, we didn’t pursue this solution, because it doesn’t currently provide stable performance.

System Implementation
Much of our time was spent developing a video transmission application that could achieve performance comparable to more expensive professional solutions while relying only on low-cost COTS devices. The application (available at http://media.polito.it/itpro) consists of both a transmitter and a receiver part. We implemented the video transmission on top of the Android platform (www.android.com) using both Java code and our optimized routines for video compression, compiled in native code to improve performance. To achieve maximum robustness against data loss and minimum latency, we implemented the video codec, which segments images in independently decodable subareas, to allow pipeline processing and better error resilience.

Due to the strict latency constraints, it’s not possible to retransmit data; however, encoding and packetization is designed so that the content of each single packet that has been successfully received can be decoded and presented to the user, because no packet depends on other ones for decoding. Missing data is estimated using a frame-copy concealment technique. Data is sent using the standard RTP protocol with a header extension that simplifies recovery in case of packet losses.

On the receiver side, the software features a multithreaded architecture that decouples the packet reception and decoding from data visualization so that the interaction with the device-rendering engine doesn’t affect the time required to process the input data. For the visualization part, we used the openGL library (www.opengl.org/sdk), including its extension for stereoscopic visualization. Therefore, the developer needs to draw data only on the correct frame buffer (left or right) and the library automatically handles all visualization issues, including alternating left and right views on the screen at the correct frame rate and synchronizing with the shutter glasses.

Our choice was to rely on open source software at the underlying level, but the whole system can be easily ported and extended to any platform. The transmitter is written partly in Java and partly in C (compression routines), while the receiver is standard C code using sockets, threads, and openGL—a set of libraries available on the vast majority of platforms.

Results
We tested the system in a practical scenario designed to investigate a set of factors that could affect the operator’s performance in remote teleoperation—in particular, the latency of the video feedback for effective control and the quality of the stereoscopic images for effective 3D perception.

The chosen components and system design already take latency and image size into account. Nevertheless, because these two factors conflict with each other, care should be taken to correctly balance the two components. For example, the more we increase image quality, the more we burden system latency and thus reduce usability.

Figure 3a shows the total amount of delay introduced by the different components of the communication chain: acquisition, processing, transmission, and rendering. First of all, an intrinsic lower bound in the overall delay is given by the sum of the camera acquisition latency, the rendering screen refresh rate, and the network transmission time. In our setup, this delay accounted for approximately 130 ms and was the same for any system configuration we tested. Second, we tested whether the time required for image compression—which depends on the device’s computational capabilities—heavily influences the maximum frame rate that can be achieved.

The latency measurements with the different image-quality configurations showed that the request for high quality strongly affects the time required for the image-compression process. This behavior limits the achievable frame rate and, in turn, increases the total end-to-end delay. For example, with a frame resolution of $416 \times 240$ pixels, we were able to encode video at 20 to 22 frames per second, while with a frame resolution of
of $640 \times 352$ pixels, we achieved 14 to 18 fps; with $1,280 \times 720$ pixels, only 3 to 5 fps was possible.

We designed an experiment with the smartphone camera mounted on a radio-controlled toy car and asked users to perform an alignment task using the 3D vision system and the car remote controller (see Figure 3b and the related video at http://youtu.be/1Olg1oV2fIs). User feedback confirmed that the frame rate allowed by the maximum resolution configuration was too low to effectively drive the car, because its overall latency (more than 300 ms) highly impaired user reactivity to the car movements. On the contrary, users preferred the first two setups—even if the image quality was lower—because their latency was within 200 ms—a value still considered satisfactory for interactive communications. In particular, because the difference in latency between the first two setups wasn’t really perceivable, almost all users preferred the second setup ($640 \times 352$). However, recent findings suggest that further quality reductions might not significantly impair the ability to perform given tasks. This could pave the way for adaptively choosing codec and coding parameters, thus varying quality if necessary, depending on the network bandwidth availability.

Moreover, we also considered the case of transmission over the wireless Internet. Experiments show that the system can work with the video qualities shown in Figure 3a using between 800 and 1,700 kbit/s. These rates have been experienced using a 3G connection with a packet loss rate lower than 1 percent. However, latency is approximately 80 ms higher compared to Wi-Fi, and packet-delay variation (jitter) was up to 30 ms, which delays the visualization of some frames. Latency has an impact mainly on the system’s perceived responsiveness, while jitter makes the video feedback subject to very short freezes that can be annoying.

In general, the results of the evaluation process indicate that the system can be applied in a context of remote teleoperation where real-time video feedback requirements can be somewhat relaxed by trading off costs with latency. With this setup, we expect that low-cost 3D visualization systems will become popular in myriad new contexts, from domestic and disability assistance to gaming and entertainment. Further studies are in development, with the aim of increasing system timeliness, even for the high-quality video setup.

Future work might include additional optimizations to enhance the user experience. For example, dropping the open source requirement in favor of proprietary software libraries will let us reduce processing time by exploiting specific hardware support for video acceleration. At the same time, the system’s total cost can be significantly reduced using much cheaper 3D video cards not yet supported by open source software.

References


Figure 3. System prototype: (a) the quality versus latency tradeoff in the test scenario and (b) the 3D vision prototype performing the test.
FEATURE: ADVANCED VIDEO TELEOPERATION


Enrico Masala is an assistant professor at Politecnico di Torino, Italy. His main research interests include the development of algorithms for quality optimization of communications over packet networks, wireline and wireless, with a particular emphasis on multimedia and specific contexts such as 3D video and dynamic adaptive HTTP streaming. Contact him at enrico.masala@polito.it.

Antonio Servetti is an assistant professor at Politecnico di Torino, Italy. His research focuses on speech/audio processing, and real-time multimedia communications. With the advent of video and audio support in HTML5, his interests include also multimedia Web applications and HTTP adaptive streaming. Contact him at antonio.servetti@polito.it.

Angelo Raffaele Meo is an emeritus professor at Politecnico di Torino, Italy. He has been involved in numerous national projects in computer engineering, investigating a broad range of topics, including switching theory, hardware design, signal processing, speech analysis and synthesis, and pattern recognition. He also served as the president of the Academy of Sciences of Torino and as the president of the Italian government commission entrusted with the task of promoting open source in the Italian Public Administration. Contact him at meo@polito.it.

Selected CS articles and columns are available for free at http://ComputingNow.computer.org.

ADVERTISER INFORMATION

Advertising Personnel
Marian Anderson: Sr. Advertising Coordinator
Email: manderson@computer.org
Phone: +1 714 816 2139 | Fax: +1 714 821 4010

Email sbrown@computer.org
Phone: +1 714 816 2144 | Fax: +1 714 821 4010

Advertising Sales Representatives (display)
Central, Northwest, Far East:
Eric Kincaid
Email: e.kincaid@computer.org
Phone: +1 214 673 3742
Fax: +1 888 886 8599

Northeast, Midwest, Europe, Middle East:
Ann & David Schissler
Email: a.schissler@computer.org, d.schissler@computer.org
Phone: +1 508 394 4026
Fax: +1 508 394 1707

Southwest, California:
Mike Hughes
Email: mikHughes@computer.org
Phone: +1 805 529 6790

Southeast:
Heather Buonadies
Email: h.buonadies@computer.org
Phone: +1 973 304 4123
Fax: +1 973 585 7071

Advertising Sales Representatives (Classified Line)
Heather Buonadies
Email: h.buonadies@computer.org
Phone: +1 973 304 4123
Fax: +1 973 585 7071

Advertising Sales Representatives (Jobs Board)
Heather Buonadies
Email: h.buonadies@computer.org
Phone: +1 973 304 4123
Fax: +1 973 585 7071
Job Recruitment and Job Seeking Processes: How Technology Can Help

Paolo Montuschi, Valentina Gatteschi, Fabrizio Lamberti, Andrea Sanna, and Claudio Demartini, Politecnico di Torino, Italy

This survey of current job search and recruitment tools focuses on applying a computer-based approach to job matchmaking. The authors present a semantic-based software platform developed in the framework of a European project on lifelong learning, highlighting future research directions.

Job seeking and recruiting processes have drastically changed during the past decade. Today’s companies are exploiting online technology (job portals, corporate websites, and so on) to make job advertisements reach an ever-growing audience. However, this advantage can create a higher post-processing burden for recruiters, who must sort through the huge amount of résumés and curricula vitae received, often expressed in different languages and formats. Similarly, job seekers spend considerable time filtering job offers and restructuring their résumés to effectively communicate their strong points and address the job requirements.

Consequently, job recruiters and seekers often use various special-purpose tools, such as job aggregators (including www.jobrapido.com and www.indeed.com) and social networks (including www.linkedin.com, www.glassdoor.com, and www.jackalopejobs.com). To further optimize selection processes with respect to processing time and accuracy, job portals such as Monster (www.monster.com) and Jobnet (www.jobnetchannel.com) have started to develop advanced search engines to automatically sort résumés based on job offer requirements. These approaches could exploit, among others, supervised and unsupervised learning, software agents, and genetic algorithms. Nonetheless, creating such tools is a complex task that requires identifying which variables influence the user’s final choice and defining ad-hoc ranking algorithms.
Computer-Supported Job Matchmaking Techniques

Computersupported job matchmaking has been explored using different methods. Supervised Learning-Based Techniques These methods exploit a set of training data to identify an inferred function that can predict the output value for each input element. Decision trees. Predictive models that represent the output (the leaves) as the result of a combination of input variables (the internal nodes). This approach has been used to predict which job offers published online are relevant for a job seeker. Neural networks. These data modeling tools can represent complex input/output relationships by mimicking the human brain’s behavior. This technique has been applied to matchmaking. Unsupervised Learning-Based Techniques These methods use unlabeled data to distinguish between and explain key features. Ant colony optimization. This metaheuristic reproduces the behavior of ants, who deposit an evaporating pheromone while searching for food to communicate with the colony. Other ants won’t cover previously explored long paths, thus avoiding the convergence to locally optimal solutions. This method is used to identify the variables with the strongest impact on the selection process. Cluster analysis. This statistics approach aims to create clusters of objects by maximizing similarities within a group and dissimilarities among objects belonging to different clusters. It has been used to support personnel assessment (in terms of identifying lateral moves or promotions, for example) by identifying “families” of employees within a company. Genetic Algorithms These heuristics were inspired by Darwin’s theory about natural evolution, where a starting population (individuals) evolves by changing its properties (chromosomes). This approach could be used by a company that aims to fill some job positions, starting with internal or pre-screened external candidates. In this case, job candidates (chromosomes) are recombined to find better chains (individuals) optimizing the allocation of human resources to job positions. Software Agents These (semi)autonomous entities can accomplish (automated) tasks for users, such as collecting job offers or résumés. Semantic Approaches These methods target the creation of self-descriptive and machine-understandable contents on the Web by adding appropriate knowledge on complex networks of relations among domain concepts. Semantic approaches let machines perform automatic processing over big and heterogeneous data by mimicking human reasoning processes.

References

Furthermore, solving the job matchmaking problem—that is, finding the best match between job offer and demand—isn’t only a matter of processes and algorithms but also of the software system’s usability, especially for those with limited computer skills. In this respect, semantic technology can help, because the computer-based job matchmaking process can be configured to mimic human-like interaction and reasoning. In this article, we present a semantic-based software tool used for job matchmaking (see also our introductory video at http://youtu.be/yi9gNd_9g8E). First, however, we
**Technology Solutions for the Enterprise**

**Semantic Technology**

Semantics—the study of the meaning of words and sentences—is often associated with disciplines such as philosophy, communication, and semiotics. For IT, the concept of semantics started to be extensively used in 2001 with the coining of term the *Semantic Web*, which represented the revolutionary passage from a Web of documents to a Web of data.¹

Today, semantics is exploited in different contexts, ranging from Web searches to various knowledge-intensive scenarios, such as natural language processing, bioengineering, and forensics. The aim is to produce more accurate answers/solutions to questions/problems. In fact, traditional search engines could sometimes fail to return the results expected because of the logical operators used in the query or the broad or ambiguous meaning of words chosen. For example, while searching for *java photos*, results might contain websites hosting pictures of the “Java island,” as well as resources to manage images in the “Java programming language.” Semantic technology aims to address the dilemma by labeling data with information related to its meaning and context, such as the specification that within webpages or in the user’s query, the word “Java” refers to the Indonesian island and not to the object-oriented programming language.

Moreover, with semantics, relations among concepts can be specified. A semantic-based search engine might manage incomplete queries and show pictures of other islands related to Java because they’re located in the same archipelago.

The knowledge management capabilities of semantic technology rely on the existence of suitable models, such as taxonomies and ontologies, to express the meaning behind the information handled. The models self-differentiate the relations managed. For example, taxonomies usually represent hierarchical classifications of concepts using trees of “is a” relations (“Java is an island,” for example). Within semantic-based processing, taxonomies are often replaced by ontologies. An ontology, defined as a formal, explicit specification of a shared conceptualization,² is meant to model a variably shaped network of relations by also recording properties about concepts (such as “Java is the most populated island of Indonesia” or “Java is located between Sumatra and Bali”).

**References**


review the role of technology in today’s job recruitment and job seeking processes, showing how job matchmaking is an application area that stands to benefit from semantic technology.

**Semantic Job Matchmaking**

By leveraging machine-understandable models of the target domain, semantic-based systems can handle a variety of both structured and unstructured content, expressed in many languages, with different degrees of completeness, and in varying levels of detail. A semantic-based job matchmaking system can improve job matchmaking in numerous ways.

First, it can help job recruiters express job requirements in flexible ways, ranging from template-based forms to natural-language descriptions. Similarly, job seekers can flexibly express skills in their résumés.

Second, a context-aware semantic-based system can improve the user experience through better human-computer interaction. By applying contextual information, the system can help recruiters revise their position postings to better advertise their requirements, and it can help job seekers revise their résumés to better promote their qualifications.

Third, job offers and demands can be automatically matched—and a ranking of positions and candidates obtained—by considering concepts that aren’t explicitly mentioned in the job offers and résumés but that are semantically related to the terms that are used.

Finally, systems can suggest ways to improve such rankings—for example, through company-wide training or individual certification. (For more information, see the “Computer-Supported Job Matchmaking Techniques” and “Semantic Technology” sidebars. You can also see a related video at http://youtu.be/OGQ26dM_KjQ.)

There are a variety of approaches you can use to develop such a semantic job matchmaking system (see Figure 1a). However, regardless of the approach used, development should occur in three major phases (see Figure 1b).⁹¹¹² The first phase should be devoted to creating a shared data model (a sort of “common language”) that defines standards for representing the information relevant for job recruitment and job seeking contexts in the

computer.org/ITPro 43
semantic knowledge base. This resource-intensive step usually involves domain experts (“knowledge engineers”) and results in the creation of domain-specific taxonomies or ontologies (see Figure 1c).

A second phase targets annotation, where information provided by users (in job advertisements and résumés) is “augmented” with meta-information derived from the taxonomical and ontological models defined in the previous step. The annotation can be carried out automatically, semi-automatically, or manually. Different investigations can be done to find a reasonable mix of automation and user involvement, though the effort is often comparable to that of constructing the model itself.9

A third phase is devoted to defining the algorithms for measuring the semantic distance among concepts and actually computing the match. At this stage, the relations among concepts used for the annotation of résumés and job offers are navigated and weighted to evaluate similarities among competences possessed and required.

The final result produced by such a system depends on the matchmaking algorithm exploited, the quality of the model defined, and the accuracy of annotations made. Currently, the research is almost equally focused on the three phases, with the aim of finding effective solutions for an improved formalization, annotation, and computation of the match, while at the same time reducing the human workload and user expertise required.

**Use Case: The LO-MATCH Platform**
The Learning Outcome (LO)-MATCH platform (www.lo-match.polito.it) is a practical example of a semantic-based matchmaking system (see a demo of the system at http://youtu.be/IT7-L_SgFVU). It was developed as part of the MATCH project (http://match.cpv.org) and co-funded by the European Commission under the Lifelong Learning Program (http://eacea.ec.europa.eu/llp), which aims to support learning opportunities during all stages of life, from early childhood through to old age. With LO-MATCH, recruiters can advertise open positions and match them with résumés posted by job seekers to find potential candidates to interview. Similarly, job seekers can insert information related to previous education or work experiences, match the information with job offers published by recruiters, and receive a ranked list of job positions.

In our platform, the match between offer and demand is computed by considering the words in user achievements (or requirements) as well as the associated semantics—that is, the meta-information derived from suitable taxonomical...
and ontological models. This way, dependencies on how information in résumés and job postings has been written are relaxed. In fact, a semantic system can operate beyond pure keyword-based comparisons, possibly enabling the exploration of partially matching contents expressed in different ways—for example, in another language, with different words, or with varying levels of detail.

**System Design**

While designing and implementing LO-MATCH, we paid particular attention to the following requirements:

- simplify the interactions and make the technological details transparent to users, and
- be flexible and provide valorization and acknowledgment of all experiences listed by job seekers or requested by recruiters.

We also made two key choices. First, we constructed the semantic model using the core concept of a learning outcome, which is what the European Commission is using to create a common language to easily translate across different countries, systems, and sectors what an individual knows, understands, and has accomplished in completing a learning process. We thus modeled résumés and job postings as collections of knowledge, skills, and competence elements, either acquired during formal or informal learning experiences or required for a given position. LO-MATCH is one of the first attempts to put into practice European strategies for lifelong learning in the context of job matchmaking.

Second, we relaxed the dependencies on the existence of an a priori strict language to formalize the domain. Although the richness of taxonomical and ontological models contributes to determining the effectiveness of semantic approaches, their design and maintenance costs can be very high. Moreover, sometimes defining complex relations among model elements can be even counterproductive. For example, what is the actual relation between programming and debugging abilities? Is it always correct to assume that if an individual can write a software program, then he or she can also debug it (or vice versa)?

**Ontology Development**

To create the LO-MATCH ontology, project partners started by identifying domain experts who could populate the LO-MATCH repository with learning-outcome-based statements belonging to professional profiles possibly owned or searched by job seekers and recruiters.

Next, this information was automatically related to the WordNet semantic thesaurus (http://wordnet.princeton.edu), where words were grouped in sets of synonyms, or “synsets,” with each one expressing a distinct concept, linked by both lexical and semantic relations.

A further ontological layer was overlapped in the platform repository by identifying key components—such as action verbs, knowledge objects, and context elements—and linking them to WordNet synsets.

Finally, the domain experts reviewed and fine-tuned this model.

**System Deployment**

We deployed a simplified user interface, targeting real-world users, enabling both the elicitation of requirements or requirements and the computation of the match. Figure 2 shows some screenshots of how a job seeker can use the platform.

When a new résumé or a job offer is entered into LO-MATCH, the system complements it with additional information derived from the underlying knowledge base. Among possible operations in this and later phases, job recruiters and seekers can browse annotated profiles or perform a free text search and receive suggestions to better define required expertise (for the job posting) or personal experience (for the résumé). This feature strongly relies on the semantic-similarity properties among learning outcomes and it is aimed at filling possible information gaps and asymmetries throughout the process.

In our design, semantic similarity plays a central role in computing the match between job offers and demands:

- Starting from WordNet and the profiles knowledge base, semantic similarity is determined from concepts found and their network of relations.
- Job seekers and recruiters can flexibly tune the weights of the relations influencing the match, moving from pure keyword-based search to the full exploitation of semantic relations potential.
- The final outcome of the matchmaking is a list of companies and candidates that better match candidates’ characteristics and recruiters’ requests,
FEATURE: SEMANTIC TECHNOLOGY

Each linked to a semantic similarity-based score as well as to its distance from the corresponding expectations and needs.

The last feature is particularly useful for both user groups:

- Job seekers can get hints about missing competences to increase their chances of being hired by a given company.
- Recruiters can quickly and easily examine the different résumés, as well as identify candidates’ weak points, needing monitoring and reinforcement once hired.

Figure 3 shows a simplified example of how learning outcomes are compared in LO-MATCH. Learning outcomes possessed by job seekers are annotated (dotted line) according to the WordNet semantic thesaurus (central sphere), containing concepts (keyword boxes) and relations among them (solid lines among boxes).

Assuming that a recruiter is looking for an individual who can “autonomously exploit current...”

Figure 2. The LO-MATCH platform interface. The job seeker chooses learning outcomes that better describe him- or herself. The platform presents a ranked list of matching job offers by showing possessed learning outcomes that match up with those required by the position.
apparel software applications,” the platform could suggest a candidate who can “draw dresses manually” (because of the relation between “apparel” and “dresses”) as well as someone who can “use, under supervision, CAD applications for the fashion industry” or “develop technical drawings using a PC program” (because of the relations between “software,” “PC program,” and “CAD,” among others). Once heterogeneously shaped contents have been brought to a common understanding, semantic distance between concepts can be exploited to provide the recruiter with ranked results so he or she can make a final decision.

Testing and User Feedback

In the MATCH project, the LO-MATCH platform represents the technological tool supporting a broad job placement methodology aimed at fostering the growth of cross-national and -cultural occupational opportunities in Europe. A piloting phase was carried out, focused on migrants. Although migrants are often qualified for available positions, their previous learning in their home country might be only partially acknowledged abroad. Furthermore, migrants might have difficulty interacting with job centers and hiring staff in the language of the receiving country.

Consequently, during piloting, the role of the platform was twofold. First, it enabled the development of an accreditation procedure for recognizing, validating, and certifying prior learning—that is, for constructing the migrant’s professional dossier (including the résumé) and issuing a corresponding vocational certificate. Second, it was used to implement the matchmaking by comparing certified learning outcomes with annotated job offers.

The methodology was implemented in “one-stop shops”—that is, platform-enabled job centers managed by the Chambers of Commerce and training institutions of six project countries. There, from February to December 2012, qualified operators provided guidance and coaching to more than 100 migrants. Almost all of them shared the same professional experiences—mainly as house or hotel cleaners, cooking assistants, health and social care operators, or shop assistants. Most of them also had previous formal learning at the secondary or vocational level. The platform let operators link both formal and nonformal learning to professional profiles recognized in the receiving country. For migrants without any previous experience, the platform indicated ways to fill the education and training gaps by showing learning outcomes most frequently requested for the available positions.

Résumés were then matched against job offers published by 54 employers and job agencies from the various participating countries. In almost all cases, the one-stop shop experience was followed by an interview at the employer premises (with the hiring staff confirming the validity of the platform-based selection process) or with the suggestion of (further) education or training actions.

During piloting, several issues were raised about platform usability that should be addressed before large-scale exploitation. One issue was related to the requirement to operate either in English or in
the profile original language (because of the underlying ontology). Also, some employers complained about the need to insert data into “yet another platform.” A further problem related to the constrained set of reference profiles, which limited the expressive power of résumés. The last issue was the lack of a way to record evidence of previous formal and nonformal learning.

The project formally ended in March 2013, and there are plans to further address these issues and prepare the platform for “prime time” use. Such plans include introducing computer-supported translation functionalities, extending the ontology, and implementing suitable data integration strategies to automatically feed into the platform data coming from external repositories of national and sectorial profiles and job offers.

Job matchmaking is an important issue in today’s global, distributed, and heterogeneous job market. Future research should be tailored to reduce the dependency on experts and users by developing automatic strategies for a seamless integration of existing profiles, résumés, and job advertisement repositories with matchmaking platforms. Additional research and implementations could also be targeted to further strengthen system flexibility, identify suitable approaches for fully exploiting existing standards and classifications, and define and implement validation methodologies to evaluate and check the effectiveness and performance of automatic job matchmaking techniques.

Acknowledgments
This work has been partially supported by the MATCH “Informal and non-formal competences matching devise for migrants employability and active citizenship” project (510739-LLP-1-2010-1-IT-GRUNDTVIG-GMP). This project has been funded with support from the European Commission. This document reflects only the authors’ views; the Commission can’t be held responsible for any use that might be made of the information contained herein. We thank the anonymous reviewers for their comments.

References

Paolo Montuschi is a professor of computer engineering at Politecnico di Torino, Italy. His research interests are in computer arithmetic, architectures, graphics, electronic publications, and semantics and education. He is an Associate Editor-in-Chief of IEEE Transactions on Computers and
CALL FOR ARTICLES

IT Professional seeks original submissions on technology solutions for the enterprise. Topics include:

- emerging technologies,
- cloud computing,
- Web 2.0 and services,
- cybersecurity,
- mobile computing,
- green IT,
- RFID,
- social software,
- data management and mining,
- systems integration,
- communication networks,
- data center operations,
- IT asset management, and
- health information technology.

We welcome articles accompanied by Web-based demos. For more information, see our author guidelines at www.computer.org/itpro/author.htm.

WWW.COMPUTER.ORG/ITPRO

Valentina Gatteschi is a research assistant at Politecnico di Torino, Italy. Her research interests are in semantic processing. Gatteschi has a PhD in computer engineering from Politecnico di Torino. Contact her at valentina.gatteschi@polito.it.

Fabrizio Lamberti is an assistant professor at Politecnico di Torino, Italy. His research interests include computational intelligence, semantic processing, distributed computing, human-computer interaction, computer graphics, and visualization. Lamberti has a PhD in computer engineering from Politecnico di Torino. He is an IEEE Computer Society member and IEEE senior member. Please visit his personal page at http://staff.polito.it/fabrizio.lamberti and contact him at fabrizio.lamberti@polito.it.

Andrea Sanna is an associate professor at Politecnico di Torino. His research interests include computer graphics, virtual reality, parallel and distributed computing, scientific visualization, and computational geometry. Sanna has a PhD in computer engineering from Politecnico di Torino. He is a senior member of ACM. Please visit his personal page at http://sanna.polito.it and contact him at andrea.sanna@polito.it.

Claudio Demartini is a professor at Politecnico di Torino, Italy, where he teaches information systems and innovation and product development. His research interests include software engineering, architectures, and Web semantics. He is a senior member of the IEEE and a member of the IEEE Computer Society and the Academic Senate of Politecnico di Torino. Please visit his personal page at http://staff.polito.it/claudio.demartini and contact him at claudio.demartini@polito.it.

Andrea Sanna is an associate professor at Politecnico di Torino. His research interests include computer graphics, virtual reality, parallel and distributed computing, scientific visualization, and computational geometry. Sanna has a PhD in computer engineering from Politecnico di Torino. He is a senior member of ACM. Please visit his personal page at http://sanna.polito.it and contact him at andrea.sanna@polito.it.

Claudio Demartini is a professor at Politecnico di Torino, Italy, where he teaches information systems and innovation and product development. His research interests include software engineering, architectures, and Web semantics. He is a senior member of the IEEE and a member of the IEEE Computer Society and the Academic Senate of Politecnico di Torino. Please visit his personal page at http://staff.polito.it/claudio.demartini and contact him at claudio.demartini@polito.it.

Andrea Sanna is an associate professor at Politecnico di Torino. His research interests include computer graphics, virtual reality, parallel and distributed computing, scientific visualization, and computational geometry. Sanna has a PhD in computer engineering from Politecnico di Torino. He is a senior member of ACM. Please visit his personal page at http://sanna.polito.it and contact him at andrea.sanna@polito.it.

Claudio Demartini is a professor at Politecnico di Torino, Italy, where he teaches information systems and innovation and product development. His research interests include software engineering, architectures, and Web semantics. He is a senior member of the IEEE and a member of the IEEE Computer Society and the Academic Senate of Politecnico di Torino. Please visit his personal page at http://staff.polito.it/claudio.demartini and contact him at claudio.demartini@polito.it.
With the promise of high definition and advanced special effects, movie theaters worldwide have evolved from traditional film projection to digital cinema projection. The Cyberport Digital Cinema Cloud helps facilitate the development of digital entertainment in Hong Kong.

Although the movie industry invests heavily in developing sophisticated tools and techniques for producing special effects, paradoxically, movie distribution continues to rely on the traditional mode of physical distribution. In the past, studios produced authorized copies in the form of roll film. Currently, studios produce authorized copies in digital form that are stored on hard disks. Copies are distributed to authorized movie theaters through a secure delivery system, and the keys are sent via email. The key is a coded file made specifically for each theater that controls the digital projector regarding which version of the movie will be played within a specific period. The entire physical distribution process involves duplication, transportation, and validation—which can take days to weeks to complete.

Furthermore, digital movies are demanding on computer systems. In a digital movie, projection resolution is represented by the horizontal pixel count, usually in the 2K (2,048 × 1,080 = 2.2 megapixels) or 4K format (4,096 × 2,160 = 8.8 megapixels). Thus, even in the 2K format, a full-length, 90-minute movie with subtitles requires a storage capacity of 75 to 150 Gbytes, whereas in the 4K format, the same movie requires a storage capacity of 360 to 900 Gbytes. At 24 frames per second, the bandwidth for real-time playback of a video is approximately 21 to 42 megabits per second for the 2K format and 100 to 250 Mbps for the 4K format. Hence, companies such as Barco, Christie, Kinoton, NEC, and Sony have developed special digital cinema projectors. In addition, some companies, such as Barco, Dolby, Doremi (partnering with NEC), GDC, Qube, and...
Sony, have supplied servers that are responsible for the control and automation of an entire multiplex.

Nevertheless, the cost and technical knowledge for installing, operating, and maintaining such a complex system are high. The future cost of upgrading the system could be even higher. Deploying cloud computing can help minimize the investment from movie theater operators while enhancing the workflow. In particular, using cloud computing as a framework that enables convenient, on-demand network access to a shared pool of configurable computing resources (for example, wired and wireless networks, servers, storage, applications, and services) can offer greater flexibility and reliability to users—that is, movie distributors and theater operators.

With funding provided by the Film Development Council of Hong Kong, in 2010, Hong Kong Cyberport Management initiated and managed a pilot project called the Digital Cinema Exchange. The project aimed to establish Cyberport’s Digital Cinema Cloud (DCC) and deploy its services to all movie theaters in Hong Kong and possibly parts of mainland China.

**Configuring a Cinema Cloud**

Cyberport’s DCC is a platform as a service that operates in a private cloud environment. Figure 1 shows the system architecture of Cyberport’s DCC.

Figure 1 illustrates that the DCC receives digital cinema packages and alternative content (see the related sidebar) via a satellite link or an international network and then distributes the content to local digital movie theaters. Cyberport is equipped with a 10-gigabits-per-second backbone network, whereas most Hong Kong digital theaters subscribe to a 100 Mbps high-speed broadband connection to Cyberport’s DCC. As of February 2014, 48 local theaters with 206 screens are streaming movies from the DCC, up from 11 local theaters with 60 screens in 2010.

Cyberport established a centralized digital-asset-management system to manage different media files in a secure environment. Nevertheless, transferring and replicating large-size digital files over WANs presents challenges using traditional TCP-based replication tools. The latency and packet loss created over long distances disrupt replication over TCP. Based on the trial tests, TCP’s throughput dropped to less than 10 percent of the maximum amount possible, and the round-trip time (RTT) increased to 300 milliseconds, representing a 5 percent packet loss.

In the DCC, we adopted heterogeneous networks and various network protocol technologies, such as forward error correction and multicasting via IPv6, to overcome these problems and restrictions. By connecting DCC and multiple theaters as a virtual private network, the system achieved a low RTT, with less than 5 ms latency. The use of IPv6 multicast applications also greatly reduced network resources and traffic loading. In terms of communication bandwidth, DCC can easily handle 100 to 350 Gbytes in the JPEG2K format for a movie, and up to 50 Mbps for stereoscopic 3D content in MPEG-2 or MPEG-4 format for live movie signals to multiple theaters.

**Digital Theaters in China**

The film and movie theater industries have enjoyed significant growth in China, even under the threat of the rise of home entertainment via the Internet. In 2013, China’s box office takings surged by 27.5 percent to ¥21.8 billion. Out of
Alternative Content

Alternative content refers to events and programs other than movies—such as operas, musicals, stage performances, rock concerts, and sports—that can be projected in digital cinemas, particularly in stereoscopic 3D format. This type of content represents a new source of revenue for the cinema industry. According to Chris McGurk, chairman and CEO of US Cinedigm Digital Cinema, alternative content in digital movie cinemas could be a US$1 billion business worldwide.

In Hong Kong, Cyberport’s DCC has broadcasted live events to cinemas, including more than 20 sports games in stereoscopic 3D format and 30 Japanese pop concerts in 2D format in the past two years. At one time, two live pop concerts were transmitted simultaneously to theaters via the multicast networks—that is, multiple-content to multiple-location screening. Although the revenue generated from alternative content is low compared with the traditional movie box office, the growth potential is tremendous, because people who don’t have the time or desire to travel might be interested in watching overseas live events on big screens, preferably in stereoscopic 3D format.

References


Figure 2. Number of movie theater screens in China. The number of digital screens (both 2D and 3D) increased from 700 in 2007 to 4,100 in 2010, then to 18,000 in 2013.

In 2013, China had 18,195 movie theater screens, around 99 percent of which were digital. The total number of screens was estimated to be 20,285 by the end of April 2014, implying that, on average, 17 movie theater screens are opening each day. Figure 2 shows the total number of movie theater screens grew in China from 2007 to 2013. The figure indicates that the number of digital screens (2D and 3D) increased from 700 in 2007 to 4,100 in 2010, then to 18,000 in 2013.

Figure 3 shows the number of movie theater screens in the US. The total number of screens increased slightly from 38,975 in 2007 to 39,783 in 2013, whereas most digital screens (3D and 2D) were used to replace analog screens from 2007 to 2013. The number of digital screens accounted for 92.5 percent of all US screens in 2013.

In comparing the information presented in Figures 2 and 3, the number of screens in China is approximately 45.7 percent of that in the US. On the other hand, China’s population is approximately 1.35 billion, whereas that of the US is 0.31 billion. Thus, the number of people per screen in the US is 8,000, whereas this number in China is 74,200. The number of movie theater screens in China is expected to have annual
double-digit percentage growth over the next decade as it follows a logistic growth curve.\textsuperscript{4}

In realizing the rapid development of digital movies and digital theaters in mainland China, Cyberport’s DCC is actively testing the reliability, flexibility, and scalability of cloud infrastructure. Moreover, Cyberport has met with China’s content providers to discuss streaming alternative content to Hong Kong and other regions or streaming movies and alternative content to mainland China.

The Digital Cinema Cloud
Using cloud technology in the movie industry presents clear advantages, as illustrated in Table 1.

Advantages
One obvious advantage of the DCC is that it lowers the cost and difficulty of film animation sequences, which can be particularly expensive due to the amount of computing power they typically require.\textsuperscript{8} Using cloud computing servers for film animation is justified, given the periodic nature of animation production as well as the occasional need to procure an extensive amount of power, which, as Qi Zhang, Lu Cheng, and Raouf Boutaba noted, is one of the main strengths of cloud computing.\textsuperscript{8}

In addition to these technological advantages, accessing and licensing creative productions, such as movies and songs, will become easier. Daniel Gervais noted that cloud computing offers the advantage of enabling significantly easier distribution of works directly to theaters and allows for better tracking of the distribution of these works.\textsuperscript{9} Furthermore, he argued that film studios must embrace this technology rather than try to maintain control of distribution through physical devices, as has been attempted in recent years, because consumers are increasingly unwilling to accept this type of restriction on their media use.\textsuperscript{9} Gervais states that using cloud-based distribution approaches can enable better control of media, such as films, and offer fewer distribution points than are currently used; as such, the potential for loss of property rights will be restricted, and the extent of intellectual property loss will be controlled.

Concerns
Cloud technology isn’t without its drawbacks for the movie industry. One of the most important concerns is the management and control of intellectual property. Intellectual property is one of the main issues in contractual relationships for cloud computing, because cloud computing is inherently more open and more prone to jurisdictional and other legal issues (although not necessarily more prone to data security breaches) than self-hosted or on-site computing services. The potential for an intellectual property breach should not be overstated. As one report noted, “75 to 80 percent of intellectual property breaches are a result of attacks made inside the company.”\textsuperscript{10} Based on this figure, the greatest potential problem for intellectual property will actually not be from the cloud computing services.

Additionally, auditability and privacy can be added to cloud services, and highly sensitive data can be retained in-house, to prevent breaches. However, privacy and protection of intellectual property remain pertinent concerns. Other concerns must...
be considered for cloud computing when applied to the movie industry. Some of the problems generally include cost of services, service quality and reliability, and the potential that the provider could go out of business (which could constrain the availability of existing data). These concerns grow even larger with customer lock-in, which constrains the ability of the user to move out of the provider’s services.

**Cinema Cloud Opportunities**

One of the major opportunities for theaters using digital screens and cloud distribution is that cloud computing enables efficient storage and streaming of ultra-high-definition (UHD) media formats. Okung-Dike Ntofon, Dimitra Simeonidou, and David Hunter discussed the emergence of future media technologies such as 4K and 8K technologies, which can be classified as UHD. They noted that these formats are significantly higher definition, but also have significantly higher size and quality of service requirements. To overcome these requirements, Ntofon, Simeonidou, and Hunter proposed the implementation of an Advanced Media Services Cloud, which can be deployed over optimal networks. The development of this cloud will enable digital and cloud media producers to deliver higher quality media to consumers in their homes or other locations, although success is dependent on the implementation of appropriate optical networks and the adoption of UHD technology at the consumer level.

However, theaters using digital screens and cloud computing also offer opportunities to challenge the current industry model. Inexpensive cloud computing means that independent filmmakers can extensively use CGI and special effects in their films—an opportunity previously only afforded to big-budget studio films. Gregory Graham has also noted that film production and distribution has been substantially eased for independent filmmakers since the introduction of cloud computing, offering them a much easier avenue to reach viewers as well as theaters. Thus, cloud computing offers an opportunity to change the fundamental nature of the film industry by realigning the productive power of major studios and independent outfits.

Although digital transmission of the movie content has been available for more than 10 years, the large-scale deployment of digital screens has only occurred recently due to 3D blockbusters. In China, a number of movie production companies are currently working on 3D movie projects, and Cyberport’s DCC will help them present their movies to audiences in a more secure and effective manner. The movie industry presents opportunities for IT growth—in particular, through the deployment of cloud-based services, ranging from digital archiving to high intensity processing, B2B monetization, and business analytics. Indeed,
Hewlett-Packard has launched the HP Public Cloud to explore opportunities brought by the rapid growth of IT use in the entertainment industry, including movies.15

Acknowledgment
A grant (RP/ESCE-01/2013) from the Macao Polytechnic Institute supported this research.

References

Wai-Ming To is a professor in the School of Business at the Macao Polytechnic Institute, China. His research interests include environmental management and total quality management. To has a PhD in structural dynamics from Imperial College London, UK. Contact him at wmto@ipm.edu.mo.

Linda S.I. Lai is an associate professor in the School of Business at the Macao Polytechnic Institute, China. Her research interests include knowledge management and electronic commerce. Lai has a PhD in information management from Lancaster University, UK. Contact her at silaol@ipm.edu.mo.

David W.K. Chung is the chief technology officer of Hong Kong Cyberport Management, where he is responsible for IT&T services operations and technology center management. His research interests include cloud computing and knowledge management. Chung has a PhD in engineering management from the City University of Hong Kong. Contact him at davidchung@cyberport.hk.

Andy W.L. Chung is the founder of XNT, a company specializing in creative digital applications, phone games, and cloud applications. His research interests include environmental modeling and game design. Chung has an MSc in environmental engineering and management from the Hong Kong University of Science & Technology. Contact him at info@xnt.hk.

Selected CS articles and columns are available for free at http://ComputingNow.computer.org.
The authors offer a roadmap for building a Maritime Cloud, offering virtual platforms, tools, and data for the maritime domain. They describe the proposed architecture and delivery models, evaluate advantages and challenges, and suggest future research directions.

Society depends on physical, cybernetic, and network infrastructures to provide information and communications technology (ICT) services and systems and to interact with numerous complex and heterogeneous IT systems. Nowadays, the need for modern ICT systems is particularly apparent in the transportation sector, including the maritime domain, which includes a complex digital environment for coordinating numerous entities, such as ports, ships, port authorities, maritime and insurance companies, and government ministries.

The EU Maritime Transport Strategy recognizes the critical role of ICTs for increasing productivity, but the involved parties—from participating enterprises to governments and standardization bodies—face numerous challenges when it comes to ensuring the standardization, interoperability, and availability of information. Because the maritime environment comprises numerous heterogeneous IT systems hosting a wide range of e-services, managing unlimited I/O information and data, cloud computing technology offers a possible solution. Here, we explore how traditional maritime IT systems could be expanded to develop and deploy a harmonized maritime cybernetic environment—a Maritime Cloud that is able to host cross-border and trustworthy e-maritime services.

The Maritime Domain
We can group the various maritime entities and assets as follows:

- **the physical environment**—port infrastructure, facilities, authorities; maritime and insurance companies; shipping and cargo industry,
manufacturers, and suppliers; government ministries; the financial and banking domain; linked transportation infrastructure; and human resources; and

- the cybernetic environment—infrastructure (buildings, ships), platforms (servers, networking terminals, databases), telecommunication systems (networking terminals, satellites, geographic information systems), software and manuals (information and data), e-services (applications, frameworks, test environments), other equipment (such as fire alarm and extinguisher systems and surveillance systems), internal (administrators, personnel) and external users (port authorities, maritime companies), and objects (ships, cargo).

Because the maritime environment consists of such a heterogeneous variety of hardware and software without any middleware solution for managing big data, adopting new technologies is becoming a major burden for the maritime domain. Although cloud computing technology is still evolving, it could help manage big data and integrate new technologies. Most users are already interacting with cloud-enabled services (that is, email, social media, online gaming, and mobile applications), and new datacenters are continuously being built to improve IT agility, resulting in reduced energy consumption and carbon footprints.

In fact, the US Navy successfully adopted the cloud computing concept in 2011, creating a private cyber environment, exploiting land, marine, and aerial technological resources as well as dynamic, static, and collaborative infrastructures and shared computer systems and resources. Building a cloud model for the commercial and enterprise maritime environment to help manage the complexity and heterogeneity of distributed and decentralized infrastructures and integrate existing services would reduce management effort via geographically remote networks. It would also deliver services and provide access to data, enhancing operational capabilities and improving the way in which IT supports the modern e-maritime domain.

**Constructing the Maritime Cloud**

The Maritime Cloud stack is based on the proposed architecture of unified cloud entities, as described by Lamia Youseff and widely adopted by IBM. Figure 1 depicts the proposed Maritime

![Figure 1. IT architecture design of the Maritime Cloud stack.](image-url)
Cloud architecture and stack, comprising the following five layers.

**Hardware Layer**
The hardware layer is the physical hardware and firmware, including the maritime IT infrastructure and physical facilities forming the backbone of the Maritime Cloud stack. It can be used in cases of government organizations or ministries investing in datacenters and IT infrastructure. It includes servers and storage solutions.

For example, storage solid state drives (SSDs) provide a reliable, high-performance serial attached SCSI (SAS) storage solution. The SAS interface with SSDs is used for I/O-intensive, demanding enterprise environments such as the maritime domain, high-performance computing, video on demand, and large-scale datacenters for other mission-critical applications. A virtual switch can also be included in this layer as part of the server firmware. Datacenter management and power consumption optimizations are also included.

**Kernel Layer**
The kernel layer is responsible for memory management, process management, task scheduling, and disk management. Yet kernels are often overlooked, despite being one of the most critical pieces when creating a cloud architecture. This layer includes the software interface for hosting three major components: the hardware, guest systems, and console operating system. It adds a management interface for the interaction of the core server system and the virtual machines (VMs) running on the server. It provides the software kernel (to be implemented as the OS kernel), a hypervisor (virtualization manager), VM monitoring, and cluster middleware.

**Infrastructure Layer**
This fundamental layer provides resources to the other layers of the stack, including applications. The Maritime Cloud services of this layer can be categorized into

- *computational resources* (infrastructure as a service, or IaaS), including virtualization concepts (that is, paravirtualization and hardware-assisted virtualization) for providing computational resources to Maritime Cloud users;

- *data storage as a service* (DaaS), letting Maritime Cloud users store their data and access them anytime from anywhere; and

- *scheduled and reliable communication capability as a service* (CaaS), including traffic isolation, dedicated bandwidth, communication encryption, and network monitoring.

The data storage helps administrators develop virtual storage class data services, offer scalability, and present a choice of storage protocols, all with better throughput to meet storage requirements. A distributed storage manager enables VM load balancing based on storage I/O and capacity within a cluster VM, with the VM workload balance based on CPU and memory resource utilization. The virtualized high-performance cluster system provides storage virtualization optimized for VMs and efficiently stores the entire VM state in a central location that supports virtualization-based distributed infrastructure services. Tiered storage allows the entire system to consolidate and deploy different types of storage within the same network-attached storage-based system.

The communication capability includes the network interface controller (NIC) acting as the base of the entire selected network protocol stack, allowing communication between the physical layer and data link layer among groups of the same LAN and large-scale network communications through routable protocols (that is, IP). NIC virtualization provides information on NIC functionality for the virtualized 10-Gb Ethernet (GbE) channel. It allows sharing a single 10-GbE network port, and its virtual Media Access Control provides statistics per server on transmitted and received (Tx/Rx) traffic. The Link Aggregation Control Protocol for gigabit interfaces configures the GbE port channels, allowing bundling of multiple GbE links into a single logical interface on a router. It is defined in the IEEE 802.3ad specification as aggregation of multiple link segments.

The virtual switch is the software program that allows one VM to communicate with another; it actually inspects the data packets before passing them forward. The hypervisor or the VM monitor (VMM) acts as the software-based virtual datacenter for delivering, providing, and consuming services, and it offers automated storage services.
Platform Layer

This layer uses IT automation to integrate service provisioning and allow administrators and architects to develop complex automation tasks, also delivering service-level agreements (SLAs) for all applications. It supplies the Maritime Cloud stack's software developers with programming framework environments and APIs to develop and deploy Maritime Cloud interactive and scalable applications.

Application Layer

This is the top layer of the Maritime Cloud stack. Here, users access the deployed services provided through the Maritime Cloud portals.

Delivery Models

The following Maritime Cloud delivery models are based on National Institute of Standards and Technology definitions.

Infrastructure as a Service

Maritime IaaS is the bare infrastructure on which maritime IT requirements will be deployed, as previously described, in the Maritime Cloud hardware, kernel, and infrastructure layer of the Maritime Cloud stack. The M-IaaS delivery model is responsible for providing server, storage, networking, and load-balancing services, and system management.

The Maritime Cloud infrastructure architecture can be designed using the existing maritime IT without excluding the hardware components of other organizations or companies related to the maritime sector. Existing maritime IT infrastructure and systems can be refurbished and fully utilized.

Platform as a Service

Maritime platform as a service (M-PaaS) is the premium infrastructure and united platform that provides tools and facilities for designing, developing, testing, deploying, and hosting maritime application solutions and services. Developers can create fully customizable e-maritime platforms while integrating existing systems and following business models and standards. Furthermore, M-PaaS enables database integration, security, scalability, storage, and management.

The crucial modules in a Maritime Cloud environment are the operating systems and platforms, the databases, and the development tools (for example, Java runtimes and webservers). Because not all cloud models are appropriate for the e-maritime ecosystem, and because maritime entities have different requirements and expectations, developers need to capture and consider requirements before designing, creating, or using a Maritime Cloud platform model. For example, e-maritime environments usually host, store, and exchange information to provide maritime participants with e-services, so interoperability and security are major challenges at all levels of an appropriate delivery of a Maritime Cloud platform model design.

Software as a Service

Maritime software as a service (M-SaaS) is the delivery model where all maritime software, applications, and services are provided on demand. As the final component of the Maritime Cloud delivery model, it can lead to the creation of new cross-border and trustworthy e-maritime services. M-SaaS is replacing the NIST-defined SaaS model in the Maritime Cloud delivery model, so that all involved maritime entities can collaborate under the same Maritime Cloud and use the same proposed Maritime Cloud architecture and stack.

Numerous e-maritime services can be included in the M-SaaS layer, including:

- vessel services that deliver e-information and vessel authentication and monitoring (via RFID and geographic information systems);
- cargo services that deliver e-documentation and cargo authentication (again, via RFID and geographic information systems);
- logistics services that provide logistics operations such as e-orders, e-invoicing, e-payments, and e-reservations;
- multichannel communication (Wi-Fi, VPN) services that communicate with all maritime entities; and
- integration services that deliver customs declarations and controls (for example, taxes and penalties) and offer e-health and e-tourism support.

Each one of these Maritime Cloud delivery models and services can also be parameterized to meet the maritime domain's needs.
Cyber Ocean Deployment

Cloud computing technology can be exploited for building various types of Maritime Cloud models in order to deploy them in the maritime domain and offer e-maritime cloud services worldwide.

Private Maritime Clouds can be limited to specific purposes, serving only the Maritime Cloud architects and developers at a regional level, and within the limits (and identity access management restrictions) of the Maritime Cloud platform’s users.

Developers can also create community Maritime Clouds by expanding the capabilities of private Maritime Clouds and allowing them to be used by other parties with similar interests, target groups, and policies (that is, security, privacy, certification, and accreditation). A community Maritime Cloud can be deployed at the national, regional, or worldwide level, offering cross-border e-maritime services accessed by the same entities as in the private Maritime Cloud but also by suppliers and end-user partners. Moreover, Maritime Clouds can provide unlimited access through public Maritime Cloud portals for the general public to access maritime services. The public Maritime Cloud can be deployed at the national, regional, or worldwide level, offering cross-border e-maritime services to clients and customers.

Finally, a hybrid cloud is a combination of the previous types. In a future maritime hybrid cloud, models can be deployed to offer e-maritime services to the general public but also consolidate legacy applications and data designed for all sides of the maritime spectrum.

Benefits and Risks

Of course, any decision about diving into the cyber ocean and using the Maritime Cloud must be based on a detailed analysis of potential benefits and risks.

If properly designed and implemented in the maritime domain, cloud computing offers numerous benefits. A Maritime Cloud can provide increased operational management, improved asset utilization (that is, servers), and improved productivity, especially in application development. It can achieve energy efficiency and proper human resource management, and reduce the time and cost of maintenance and management of IT resources. Users get rapid access to Maritime Cloud services.

A Maritime Cloud also offers efficient optimization of valuable storage, enabling continuous improvement and agility. In addition, by enhancing maritime mobility with network scalability, it can provide an efficient network load balance. It can contribute to a stable e-maritime environment, scale up the number of supported end users, and satisfy their requirements. Emerging technological trends like BYOD (bring your own device) can be combined, utilized, and implemented into standardized processes and methods of certification and accreditation of a Maritime Cloud environment.

Green IT (energy-efficient computing) is part of the IT life cycle and environment, both physical and cybernetic. By efficiently managing virtualized resource pools and optimizing application performance, a Maritime Cloud can minimize energy consumption in datacenters. Environmental protections can help users adopt eco-friendly approaches and technologies. By reducing the IT infrastructure, adopting a Maritime Cloud architecture, or recycling the life of already existing maritime IT infrastructures, users can also reduce energy costs and system footprints (that is, servers and datacenters) inside a maritime environment.

Having fewer application licensing costs, better server utilization, less infrastructure costs, more efficient data storage, less energy use, easy and free application development and testing, and instant e-maritime service provision makes it possible for Maritime Cloud developers to reduce costs and create immediate profit as a vital benefit and advantage for the maritime domain at any level—regional, national, and global.

However, despite the initial success of cloud computing, a significant number of risks exist, because cloud infrastructures and platforms in their current form don’t employ standardized methods of storing user data and applications. Issues to be addressed include the so-called CIA triad—confidentiality, integrity, and availability—regarding users’ privacy, data security, legacy system performance, and disaster recovery, and the transition issues that arise when establishing, maintaining, sustaining, and managing a new Maritime Cloud environment.

The security, privacy, and trust issues involving Maritime Cloud data migration, management, portability, and interoperability, along with secure data and application storage, all affect the
Technology Solutions for the Enterprise

The adoption of cloud computing technology can revolutionize the maritime domain, enabling a collaborative environment offering cross-border e-maritime services to all participants and users, and providing reliable information and immediate processing. IT professionals and experts can take advantage of highly efficient cloud infrastructures while increasing agility. On-demand, ubiquitous network access and location-independent resource pooling are exactly what the maritime domain needs. In addition, large-scale infrastructures can lower energy use and emissions by using cloud computing technology. The maritime deployment models we’ve presented can be implemented at any level to deliver a variety of innovative e-maritime services.

However, it’s essential to develop a standardized migration strategy for moving to the Maritime Cloud. This will require generating a new set of e-maritime cloud services to solve interoperability issues among cross-border maritime standardization bodies, organizations, and working groups—such as the International Maritime Organization (www.imo.org), the European Maritime Safety Agency (www.emsa.europa.eu), and the European Commission’s Innovation and Networks Executive Agency (http://tenta.europa.eu/en/about_us/organisation)—leading to a universal maritime policy. Topics such as security, data interoperability, SLAs, and mobility are open issues for future research work.

References


Kleanthis Dellios is a PhD candidate in the Department of Informatics at the University of Piraeus. He received an MSc in advanced information systems from the same university. His research interests include computer security, biometric technology, e-government, and intelligent transport systems. Dellios is a certified ISMS Auditor (ISO/IEC 27001:2005). Contact him at kdelios@upi.gr.

Dimitrios Papanikas is a PhD candidate in the Department of Informatics at the University of Piraeus. He received an MSc in advanced information systems from the same university. His research focuses on software engineering, information systems security, vulnerability assessment, and penetration testing technology. Papanikas is a certified ISMS Auditor (ISO/IEC 27001:2005). Contact him at papanik@upi.gr.

Selected CS articles and columns are available for free at http://ComputingNow.computer.org.
Paraguay, with its population of approximately 6.8 million, has one of the lowest gross domestic products (GDPs) per capita of Latin America. Furthermore, IT penetration in the country is quite low. As of 2012, only 27.08 percent of individuals used the Internet—the lowest percentage in South America—and only 29.48 percent of households had a computer.

The private IT market, independently from the government’s efforts to promote IT, has grown remarkably in the last two decades. It first grew in the mobile telecommunication industry, then the growth moved into the financial market, and it’s now affecting other industries, including agriculture.

Academia is also supporting IT growth. Currently, 34 universities offer 46 computer science programs and a few professional-oriented master’s programs in IT. There are also some promising research groups (two universities offer doctoral programs in IT). However, in the absence of an effective national quality-control system, the quality of university programs varies.

Therefore, although there are some highly skilled IT professionals, it’s still difficult to find enough local personnel to cover the demand of the growing market.

In the last few years, the central government has put IT on its policy agendas. The goal has been to introduce IT-based solutions to improve public administration processes; deliver quality public services; and improve transparency, citizen participation, and accountability. In April 2012, Paraguay joined the Open Government Partnership (OGP; www.opengovpartnership.org), which helps countries make their governments more “open, accountable, and responsive to citizens.” Here, we outline Paraguay’s journey toward open government (or open-government) and the technical and organizational challenges it faces.

Moving Beyond E-Government
In 1997, Paraguay implemented a metropolitan network to connect 105 government agencies via a national Internet-based virtual private network (VPN). However, the communication infrastructure wasn’t enough to improve processes and services, because the accompanying IT solutions were developed in isolation, often by foreign companies. On several occasions, high-quality software solutions, deployed by foreign teams, failed due to the lack of local support (the solutions weren’t sustainable). Moreover, significant effort was required to make these isolated solutions interoperable.

In 2010, the government decided to design a “master plan” for information and communication technology (ICT) and, two years later, it created an ICT secretary position. Part of its master plan was to emphasize the use of free software. Yet despite advances in the infrastructure and the creation of an ICT master plan to modernize government processes and better empower citizens, challenges still remained. In particular, the government struggled to do the following:

- apply citizen-centric approaches to system development,
- improve local capabilities for technology development and support, and
- improve information availability and accessibility for all stakeholders.
The latter introduces the challenge of open data, which enables transparency and is the foundation of open government.

**Open Government**
The o-government philosophy fosters a more horizontal and democratic way of governing, which benefits citizen participation, promotes transparency and strengthens technology with the aim of improving the quality of public services. The key components to any o-government initiative are transparency, participation and collaboration, and accountability.

**Transparency**
Transparency refers to openness of information about government activities and decisions in a way that is comprehensive, timely, and accurate, following open data standards. Transparency is the key enabler for citizens to understand the workings of their government, and it's important in preventing and detecting corruption and inefficiencies in government services.

**Participation and Collaboration**
Participation and collaboration lead to citizen engagement and empowerment, which are essential for effective governance and policy making. This allows public opinion to influence how the government works, which helps improve the delivery of services.

**Accountability**
Accountability is the process by which citizens can hold the government responsible for its management of public goods.

**Open Data**
Open data, which is the base of transparency, refers to data or content that “is open for anyone to freely use, reuse, and redistribute… subject only, at most, to the requirement to attribute and/or share-alike” (http://opendefinition.org). Interoperability and re-use is achieved by following standards such as the 5-Stars Open Data Standard, defined by Tim Berners-Lee (see Table 1). The higher the number of stars, the more re-usable and interoperable the data will become; however, the technical effort needed will also be higher. Considering that different institutions in the government have different technical capabilities, there should be a tradeoff between high-level standards and available resources when opening data.

There are several benefits to open data. First of all, in terms of government, open data is a key enabler for transparency, citizen participation, and accountability. Second, it fosters innovation, the discovery of new knowledge, and improved services. Third, some estimate that open data could have an economic benefit of up to US$3 trillion, resulting from increased efficiency, the development of new products and services, and cost savings from higher quality products.4

Furthermore, the costs of developing government data visualization can be outsourced to interested stakeholders. Projects such as WhereDoesMyMoneyGo.org and OpenSpending.org are examples of how the government, by publishing budgets in a reusable format, enabled effective visualization tools by third parties.

**O-Government Challenges**
The challenges for an optimal implementation of o-government in Paraguay, starting from transparency, span several dimensions, but for the most part, success will depend more on political willingness than technical capabilities. Fortunately, citizen- and government-led initiatives aim to address the following challenges.

**Legal**
Despite an Article in the Paraguayan National Constitution that states that all individuals have the right to receive truthful, responsible, and unbiased information, and that all sources of public-sector information should be freely available, the most important legal challenge for o-government in Paraguay is to create a law that regulates access to public-sector information. Furthermore, there are no legal definitions of public-sector information, data, or datasets, or of open data.

Recent developments, such as the Supreme Court decision to grant access to information regarding the salaries of public officials, and the partial approval of a Public Sector Information Access Law, show there’s a high degree of interest by both the lawmakers and the public in overcoming current legal challenges.

**Organizational**
There’s also a need for proper organizational tools and procedures for releasing public data. The Administrative Law of Paraguay only allows public officials to do what’s explicitly permitted; anything else is forbidden and subject to legal consequences. This implies that
in the absence of organizational procedures for releasing public sector information, public officials aren’t allowed to publish data.

Another challenge is the insufficient number of experts available to work on the technical and legal aspects of opening data. The final challenge refers to the lack of knowledge about o-government principles that sometimes creates a resistance to change by public officials.

**Technical**
The government must select open data standards that foster transparency and help create technological tools, such as forums and feedback mechanisms for collaborative consultation processes that support citizen engagement and community management. However, it should also provide an IT infrastructure that can guarantee continuous availability of such tools.

Furthermore, Paraguay needs to create tools such as catalogs, indexes, and search and navigation mechanisms that help citizens find and use open data. Such data should include metadata that specifies the spatial and temporal validity to make the semantics of each piece of information explicit. Interoperability must also be ensured at the vocabulary, schema, metadata, and data levels.

**Adoption**
Adoption deals with creating awareness of o-government and of how citizens and public officials can use open information. The challenge is to create and maintain an informed community that can actively participate in o-government initiatives using the available legal and technological tools.

Considering this, Tim Davies used Berners-Lee’s standard to create a set of principles, called 5-Stars Open Data Engagement, that guide the process of opening data:

1. be demand driven (*),
2. provide context (***),
3. support conversation (****),
4. build capacity and skills (*****), and
5. collaborate with the community (******).

The idea of this principle is that an informed community can guide the process of opening technically correct data, maximizing the adoption of standards and open data.

**References**

Luca Cernuzzi is a professor at the Universidad Católica Nuestra Señora de la Asunción, Paraguay. Contact him at lcernuzz@uca.edu.py.

Juan Pane is an open data consultant at the Centro de Estudios Ambientales y Sociales (CEAMSO), Paraguay, and a lecturer at the Universidad Nacional de Asunción. Contact him at jpane@pol.una.py.

Selected CS articles and columns are available for free at http://ComputingNow.computer.org.
January/February
**IT Security**
Richard Kuhn, US National Institute of Standards and Technology
Morris Chang, Iowa State University
Tim Weil, Coalfire Systems

March/April
**IT-Enabled Business Innovation**
Robert Harmon, Portland State University
Haluk Demirkan, University of Washington

May/June
**Internet of Everything**
Irena Bojanova, UMUC
George Hurlburt, STEM Corp
Jeff Voas, US National Institute of Standards and Technology

July/August
**Data Analytics**
Seth Earley, Earley & Associates
Jinan Fiaidhi, Lakehead University, Canada
Vladimir Dimitrov, University of Sofia
Maria R. Lee, Shih Chien University

September/October
**Wearable Computing**
Irena Bojanova, UMUC
Maria R. Lee, Shih Chien University

November/December
**Smart Systems**
Fulvio Corno, Politecnico di Torino
Robert Harmon, Portland State University

For more information, see www.computer.org/itpro/cfps
Usability for Internal Systems: What’s the Payoff?

Seth Earley, Earley & Associates

Consumer applications are raising the bar when it comes to usability expectations, and such expectations are being carried over into corporate environments. IT organizations thus should be paying close attention to elements that make internal systems and applications intuitive and easy to use. But what’s the justification for doing so? Easier access to information doesn’t provide business value unless there’s a corresponding improvement in efficiency or a new capability. Organizations are often hard pressed to justify the significant investments required to address enterprise-wide applications when the outcome is uncertain and the payoff unclear.

For external applications, the stakes are far more clear. A user experience arms race is taking place in the marketplace—with multibillion dollar rewards for those who get it right. For the ones left behind, failing to measure up can mean lost opportunities and market share or, in more extreme circumstances, irrelevance in the eyes of customers.

Focusing on External Apps

Usability and the user experience are key differentiators for pure-play Internet businesses. Although they sound similar, there’s a significant difference between the two concepts. The user experience entails everything that a user sees, feels, and thinks about an organization. Usability is more about the mechanics of a site or application. Can users get to the information they need with minimal friction? Can they accomplish their tasks easily and elegantly? Is the site intuitive?

Measuring Usability

Usability is measured by articulating tasks for target users to accomplish and letting them execute those tasks in the application or website—whether on a desktop, laptop, tablet, or smartphone. The success of those tasks and the ease with which they’re executed (along with some other metrics such as number of false paths or restarts) become aggregated across user groups and multiple scenarios to develop usability measures.

Complexities arise when considering that different users have different levels of fluency with technology or might have different educational and cultural backgrounds. Therefore, an important part of usability is selecting the correct demographic for a user group. For Internet retailers, for example, the target users are defined through market research and by capturing data and indicators about current customers and extrapolating to understand desired customers. User groups are screened as part of the study design and set up.

Personas—archetypical customer composites representing characteristics of a target user—are used to get into the mind of the users to imagine how they might react. Personas are typically given a name for easy reference—“Bob” might represent a 25-year old tech-savvy college graduate who lives alone and is interested in martial arts. The seemingly trivial or unrelated details help round out the type of person that system designers think “Bob” is. This way, they can make design choices they believe would suit him.

Some of this conjecture is more art than science, but most
enterprises dealing with a broad range of user types on their public commerce-focused websites find personas to be extremely valuable. The point is that a deep, rich understanding of the customer can be represented in model development. Taking this a step further, usability specialists, marketers, and business analysts build libraries of scenarios for personas that describe a “day in the life”—the tasks and activities that these “personas” engage in while going about their business. Use cases can be defined or derived from scenarios and provide yet another level of detail around user understanding. These use cases can be more along the lines of task descriptions than the strict use cases developed for system testing. Formal test use cases for system functionality are derived from task descriptions.

When placed into the online context, each step of a task typically requires interaction with data and content or the ability to locate a particular piece of information. Therefore, this process of understanding users—from the macro level down to the detailed task and use case level—defines content access, information management, and functional requirements.

Understanding the User Experience
The user experience is a broader concept. It typically begins with awareness and brand perceptions. Once a user becomes interested in a product or service—whether through word of mouth, social media, a newspaper article, an email communication, or a Web search—engagement can begin in any number of ways.

If the company has a physical presence, the user (now a potential consumer) might walk into a store. He or she might do more research online and get more information from a third-party site, a friend on Facebook, or the company site itself. Each one of these mechanisms creates a perception on the part of users that impacts what they do next. At some point, they might decide to make a purchase. If this takes place online, the architecture of the website and its ease of use have a direct impact on whether the individual follows through with the purchase. Conversion is an important metric for online retailers; however, it is merely the final step in a potentially long chain of events.

Imagine a scenario in which an organization has spent millions of dollars on its website and tested the usability until it was optimized across all of the target audiences. The organization has spent all of its resources on the site and didn’t leave enough money for call center training and support systems. A new customer who has a problem with shipping or billing that can’t be resolved online might call the company, get placed on hold or transferred multiple times, become disconnected, or reach a poorly trained offshore representative who can’t solve the problem. News of this experience reaches discussion boards, Twitter, Facebook, and the company’s community of users. Clearly, the site’s excellent usability can’t overcome this terrible user experience.

Although exaggerated, this type of scenario is happening across thousands of organizations. In this scenario, parts of the user experience are broken. Legacy systems that are stitched together have evolved over time and are poorly integrated. The call center isn’t aware of marketing’s new offer. Billing hasn’t heard about and adjusted for the upcoming service outage. Customer support doesn’t have the latest fix for the problem that cropped up with the newest model. Online offers are out of stock with items in the store because of delays in getting all of the systems updated with the correct information. The customer experience, also referred to as the “customer journey,” is then taking place over rutted roads with detours and snarled intersections, poor and confusing signage, and dead ends.

Linking Internal and External Performance
Analytics and big data play a crucial role in the customer experience. We now have data streams that are emitted as “exhaust” from every process. Internal systems can measure the performance of the various departments supporting each stage of the customer journey—learning about the product or service (marketing), purchasing (sales) and receiving (shipping and logistics) the product, enjoying the defect-free product (service), handling the billing and payment component (finance), and providing additional post purchase repairs or services (support). This model might vary somewhat based on the industry, product, and service, but for the most part, it’s fairly universal. At each step, internal department and process performance can be measured via financial measures, key performance indicators, and dashboards.
DATA ANALYTICS

An external set of metrics referred to as the voice of the customer (VOC) consists of surveys, point-of-sale feedback, website comments, service calls, complaints, feedback to call centers, social media commentary, and discussions. This input is mostly in the form of unstructured data, which requires text analytic methodologies to derive insights. User reactions and feedback can also be derived from online behaviors—for example, clickstream analysis can track when users are attempting to locate information or are interacting with specific Web content.

Because website and e-commerce investments have obvious links to sales and conversions, content that directly supports the customer experience receives a great deal of attention and resources. However, because the customer experience consists of a chain of interactions, poor internal user experiences and the consequent inefficiencies in internal systems can’t help but have a downstream impact on the customer experience. This chain of interactions is increasingly taking place on mobile devices. Many retailers view mobile devices as being part of a behavior pattern (rather than as a marketing channel) and as a way to bridge the physical and virtual worlds. As such, the rise of “omnichannel” communication, engagement, and commerce requires that the value chain of interactions—the customer journey—be smooth and consistent.

Omnichannel marketing thus means understanding that customers might start on one device and then move to another or end up at a physical location. Furthermore, content and communications should be the same whether in a newspaper, on a radio ad, on a mobile application, on a website, or in the store itself. Customer service systems and floor sales associates all should have the same information and key content for supporting the customer. Internal systems and processes need to be integrated and continuous so that the customer receives consistent messaging, offers, and functionality, creating an overall positive, friction-free experience.

Dashboards can be designed using text analytics to correlate VOC data with internal process performance. Dashboards containing operational analytics from financial systems and business intelligence applications are nothing new. Associating external user data harvested from various sources with internal operational metrics can provide a real-time scorecard of how internal enterprise processes affect external user perceptions. Broken and inefficient internal processes have an impact on downstream and external processes.

Making the Investment

The challenge is to improve internal processes across the enterprise to create greater efficiencies that will then enhance the customer journey.

Implement Master Data Management

Programs such as master data management (MDM) and semantic integration can greatly improve information flows and ease of integration across departments, systems, and processes. MDM requires a significant investment of time, money, and attention to fix data inconsistencies at their source so that downstream applications and stakeholders can trust data quality and not go through costly data cleansing activities. Semantic integration helps different stakeholders in the enterprise “speak the same language” and use consistent terminology.

All organizational environments change over time. Systems and processes grow organically to solve problems, applications might be deployed from various vendors who develop their tools using different mental models, and modern corporations grow through acquisition. Change is ongoing, so this problem never goes away. The only way to address it is to put into place organizational structures whose sole purpose is to continually rein in the chaos and invest the energy needed to reverse information and system entropy.

Tame Unstructured Data

Users’ higher expectations for greater levels of quality and usability in corporate information systems require that IT organizations overinvest where they have traditionally underinvested. IDC estimates that 90 percent of information is unstructured.1 Yet unstructured information management remains the stepchild of IT organizations, because the impact on critical processes isn’t directly observed. Poor information hygiene for unstructured information is insidious, because it grinds down internal processes slowly over time, reducing the ability to quickly react and adapt.

Process improvement begins by examining those systems that directly and indirectly support the customer and assessing places where terminology is inconsistent or causes friction points due to incorrect or outdated architectures. Content, process, and application audits, along with domain modeling, can reveal where clashes in terminology, ambiguity, poor data quality, and manual integration choke points cause the greatest pain and problems with customer-facing processes. Interventions can then be prioritized to remedy these problems using enterprise information architecture development approaches and validating design by testing scenarios,
tasks, and use cases with the usability methodologies previously described.

**Build the Infrastructure**
Achieving meaningful results requires that this process be conducted at the enterprise level. For many IT organizations, that doesn’t sound feasible. After all, how can we document scenarios and use cases for thousands of activities? The answer is that you don’t document all of them—only critical processes that allow for extrapolation of organizing principles to validate an enterprise framework.

Much of this work is infrastructure—part of the cost of doing business. Foundational work of this type is required for higher value functionality. It needs to be justified by articulating the vision of adaptability, agility, and greater customer intimacy.

At a recent conference on retail systems and processes, the presenter asked “Who has completely integrated their back-end information?” No hands were raised. “Who has clean data for personalization?” Again, no one raised their hand. “We are all in this together,” she concluded. CIOs might think this is a retail and e-commerce story, but it’s not. Every organization will be transforming at least some of its processes through digital capabilities. We are, indeed, all in this together.

**Reference**

**Seth Earley** is CEO of Earley & Associates (www.earley.com). He’s an expert in knowledge processes and customer experience management strategies. His interests include customer experience design, knowledge management, content management systems and strategy, and taxonomy development. Contact him at seth@earley.com.

Selected CS articles and columns are available for free at http://ComputingNow.computer.org.
Some algorithms make for “better” programs than others—that is, programs that execute in less time or require less memory. How can we quantify differences to determine which algorithms are better? No one has done more to answer this question than Don Knuth, who has been called the “father of the analysis of algorithms.”

In his precollege days in Milwaukee, Knuth already showed signs of having a special talent, and his performance as an undergraduate at Case Institute of Technology was so outstanding that the faculty awarded him a concurrent master’s degree. He undertook graduate study in mathematics at Cal Tech and received his PhD in 1963. Knuth remained on the Cal Tech faculty until 1968, when he moved to Stanford University, where he remains to this day as an Emeritus Professor.

In 1962, while still a mathematics graduate student, Knuth decided to write a book on programming languages and compilers—subjects for which he was already an important contributor. The first eight chapters would cover computing “preliminaries,” and the last four chapters would deal with his intended subjects. However, as he gathered material, he noticed that the book was becoming very long. He decided that a thorough treatment was better than a brief one, so the “preliminaries” became a series of books that defined the analysis of algorithms. The series was called The Art of Computer Programming, because Knuth viewed computer programming as an art as well as a science. Three volumes containing the first six chapters were published between 1968 and 1974 (the latest editions of these volumes were published in 1997 and 1998).Remarkably, the first part of volume 4 wasn’t published until 2011! Here, I describe some of the content of the first three volumes, indicate one reason for the publishing hiatus, and report on Knuth’s intentions for the future of the series.

Volume 1: Fundamental Algorithms
The basic concepts covered in Chapter 1 begin with assembling the mathematics and statistics used throughout the volumes to analyze algorithms. The analyses include both the memory requirements and the execution time of the algorithm. Regarding execution time, Knuth tackles both the worst-case time, which depends on worst-case input, and the average-case time, which depends on the statistical distribution of possible inputs.

After laying out the math, Knuth defines a hypothetical computer called MIX and then discusses assembly language programming in MIX. Even in the late 1960s, assembly language programming was starting to go out of fashion, but he defended the choice as required for a careful analysis of run time. He also gave a high-level description of each algorithm to be analyzed. Finally, he defined an MIX simulator, so that MIX programs could be run as well as written.

The information structures described in Chapter 2 proved to be the basis for many of the following texts on algorithms and data structures. The basic information structures presented are linear lists and trees. The implementation
techniques considered are **sequential** and **linked** allocation. Among other names, a sequentially allocated linear list is also called an **array**. Trees are often implemented by linking the subtrees to the parent nodes by pointers. Various combinations of these structures and implementation techniques provide the foundation for constructing complex data structures. However, the emphasis needs to be on the operations to be applied to a given data structure. That is, the best way to implement a data structure depends on the mix of operations expected to be used in the application.

Chapter 2 ends with a discussion of storage allocation methods: techniques for carving out a bit of memory in which to create a data structure. Unfortunately, even today, one of the notorious mistakes programmers make is failing to check for “buffer overflow” when allocating space for new data (that is, allocating space beyond that reserved for the purpose). Nefarious hackers often use this flaw to insert malicious code into a computer.

**Volume 2: Seminumerical Algorithms**

Random number generation is the subject of Chapter 3. Knuth made clear he was discussing **pseudo-random** sequences: deterministic but with characteristics of randomness. Determinism was important, because programs could then be debugged and otherwise rerun with the same (pseudo)random sequence as input. Because modeling and simulating physical systems were—and are—such important computing applications, and because such simulations often involve random selections from a distribution of values, doing the randomization correctly can be the difference between a good simulation and garbage.

Knuth analyses the **linear congruential method**, \( X_{i+1} = (a \times X_i + b) \mod c \). That is, the next random number, \( X_{i+1} \), is computed from the previous random number, \( X_i \), and is the remainder of dividing the expression \( (a \times X_i + b) \) by \( c \). If \( a \), \( b \), and \( c \) are carefully chosen, this simple technique returns a pseudorandom sequence of numbers uniformly distributed between zero and one. He then shows how such a sequence can be turned into a random sequence from any number of other probability distributions.

Computer arithmetic is the subject of Chapter 4. One clever reviewer of Volume 2 said, “Oh, no! I don’t want to know that much about computer arithmetic.” However, some computer programmers need to understand how to connect abstract numerical operations with real computers (hence the designation as seminumerical). For example, how should exponentiation be implemented? If one naively assumes that computing \( x^{64} \) requires 63 multiplications (that’s the way it’s defined), then reading this chapter would be instructive. If the variable \( x \) is replaced by the value \( x^2 \) six times, the result is \( x^{64} \), so it actually requires only six multiplications. A slight generalization of this squaring process works for raising a number to any integer power. If there are \( n \) bits in the exponent, the algorithm requires at most \( 2n \) multiplications.

**Volume 3: Sorting and Searching**

Sorting, the subject of Chapter 5, is divided into internal and external sorting. Internal sorting is when the file to be sorted fits into the computer’s main memory. External sorting is required when disks or tapes are necessary to hold the file. Knuth says that sorting algorithms were among the first to be developed for the earliest computers. He suggests that the study of sorting algorithms can teach us much more than sorting. For example, it can teach us how to analyze algorithms and how to approach a computing problem from many directions. One of the slowest internal sorting algorithms described is called **bubble sort**, and one of the fastest is called **quick sort**. Knuth’s analysis shows that quick sort would, on average, require approximately six minutes on a one-gigap computer to sort a list of a billion numbers. Bubble sort, on the same computer, would require approximately 192 years!

Searching is the subject of Chapter 6, and it’s divided into three techniques: searching by **comparison of keys**, **digital searching**, and **hashing**. Searching by comparison of keys in an ordered file of \( n \) keys never takes more than \( \log_2 n \) searches. In digital searching, the search key is considered to be a sequence of small numbers, which act as indices into arrays that contain pointers to the part of the file where the key might be found. In hashing, the key is considered as a single number. In one hashing method, the key is divided by the length of the array that contains the keys. The array length should be chosen as a prime number, and the remainder of the division then produces a random index into the array where the key will be located, with no searching required.

One of the notorious mistakes programmes make is failing to check for “buffer overflow” when allocating space for new data.
Two keys occasionally hash to the same index, called a collision, and Knuth discusses collision resolution techniques. Each of these search methods can be used for both internal and external searching applications. As with sorting, a good algorithm will underlie a speedy search, which Google and other search engines can achieve.

TeX and Metafont
One reason for the long publishing pause between Volume 3 and the first part of Volume 4 was that Knuth got interested in improving the technology to typeset his books. Type setting had started to go digital in the 70s, but was immature. Finding no digital system up to his standards, he decided to create one. Perhaps the fact that his father had owned a small printing business gave him a leg up for this project. He called it TeX, taken from the Greek letters tau epsilon chi and pronounced “tech.”

One enthusiastic observer called it the greatest advance in printing since Gutenberg, and it was honored by the printing, publishing, and scholarly communities alike.

In a related development, he created Metafont, a system for developing digital fonts, and he used it to define a font he called Computer Modern for use in his books (and elsewhere). TeX was under development for several decades and has now been stabilized in a final version. A new packaging of TeX, called TeXLive, provides an XML front end to the TeX system.

Knuth retired early (in 1992 at the age of 54), because he thought that he would need about 20 years of full-time work to complete The Art of Computer Programming. He now projects that Volume 4 (with chapters on combinatorial searching and recursion) and Volume 5 (with chapters on lexical scanning and syntactic techniques) will be completed by 2020 (when he will be 82). He states the following on his webpage (http://www-cs-faculty.stanford.edu/~uno):

After Volume 5 has been completed, I will revise Volumes 1–3 again to bring them up to date.... Then I will publish a “reader’s digest” edition of Volumes 1–5, condensing the most important material into a single book. And after Volumes 1–5 are done, God willing, I plan to publish Volume 6 (on the theory of context-free languages) and Volume 7 (on compiler techniques), but only if the things I want to say about those topics are still relevant and still haven’t been said. Volumes 1–5 represent the central core of computer programming for sequential machines; the subjects of Volumes 6 and 7 are important but more specialized.

Regardless of its incomplete status, many people have judged The Art of Computer Programming to be the most monumental series of books on computer programming. For example, at the end of 1999, it was named one of the best 12 physical-science monographs of the century by American Scientist. Knuth has also received numerous honors, including the Turing Award (1974) and the National Medal of Science (1979).

I’ll leave the final word to Bill Gates, who is quoted on the jacket of the third edition of Volume 1 (Addison–Wesley Professional, 1997) as saying, “If you think you’re a really good programmer... read [The Art of Computer Programming]... You should definitely send me a résumé if you can read the whole thing.”

References

George Strawn is director of the National Coordination Office for the Networking and Information Technology Research and Development Program (NITRD). Contact him at gostrawn@gmail.com.
2015

Richard E. Merwin
Distinguished Service Award

CALL FOR AWARD NOMINATIONS
Deadline 15 October 2014

ABOUT THE MERWIN AWARD
The highest level volunteer service award of the Computer Society for outstanding service to the profession at large, including significant service to the Computer Society or its predecessor organizations.

ABOUT DICK MERWIN
Dick Merwin was a pioneer in digital computer engineering who participated in the development of the ENIAC, MANIAC and STRETCH computers. Despite a busy and productive technical career, Merwin found time to be active in professional societies, including the Computer Society, ACM and AFIPS. His generosity of spirit and genuine helpfulness was an important element in the progress of the computer profession.

AWARD
A bronze medal and $5,000 honorarium are awarded.

PRESENTATION
The Richard E. Merwin Award is presented at the Computer Society’s Annual Awards Ceremony.

REQUIREMENTS
This award requires 3 endorsements.

NOMINATION SUBMISSION
Nominations are being accepted electronically
http://www.computer.org/portal/web/awards/merwin

AWARDS HOMEPAGE
www.computer.org/awards

CONTACT US
awards@computer.org

STEPHEN L. DIAMOND
2014 Richard E. Merwin Distinguished Service Award

For distinguished leadership and service to the IEEE, IEEE Computer Society and the Computing profession.
WordPress & Powerful App Platforms!
- Supports over 140 popular apps including WordPress, Drupal™, Joomla™, TYPO3, and more
- Security & version upgrade notifications
- Trial version available for all applications
- App Expert Support
- 1&1 CDN powered by CloudFlare™ and 2 GB guaranteed RAM for peak performance

Powerful Tools
- PHP 5.5, Perl, Python, Ruby
- 1&1 Mobile Website Builder
- NetObjects Fusion® 2013 included

Successful Marketing
- Facebook® advertising credits
- Listing in business directories
- 1&1 Search Engine Optimization
- 1&1 E-Mail Marketing Manager

State-of-the-Art Technology
- Maximum availability (Geo-Redundancy)
- 300 Gbit/s network connection

All Inclusive
- 1 FREE domain: .com, .net, .org, .biz, .info
- Unlimited Power: webspace, traffic, mail accounts, MySQL databases
- Secure e-mail addresses with virus and spam protection
- Linux or Windows operating system

Starting at $0.99 per month*

1 (877) 461-2631
1and1.com