Defining Hard Technical Challenges in the ICTD space

Lakshminarayanan Subramanian
Courant Institute of Mathematical Sciences
New York University
lakshmi@cs.nyu.edu

ABSTRACT

To make ICTD into a viable, sustainable and attractive research space within computer science, it is imperative that we address two fundamental issues: (a) identify hard technical challenges that define the space; (b) demonstrate the high impact factor for solutions that solve these hard challenges. I will articulate some technical challenges in this space.

1. INTRODUCTION

We, as computer science researchers working on ICTD are at crossroads simultaneously trying to serve two masters. On one hand, we are trying to convince core computer scientists on the value of ICTD research as a discipline within computer science. On the other end, we are trying to convince global development specialists on the value of new technological solutions and their potential to address pressing developmental problems. While the computer scientist wants to know "what are the challenging research problems", the development specialist wants to see "real-world impact and utility for addressing the problem". Not every developmental problem warrants a technically challenging solution and not every technical challenging problem is a real-world problem. One needs to carefully walk this fine line; especially, given our habit of trying to over-generalize problems which often quickly makes the operating assumptions unreal.

I believe there is a simple way out of this. ICTD as a field needs a unified vision around a set of well-defined "grand" challenges that satisfy the test of both the computer scientist and the developmental specialist. Any new emerging field is centered around a vision that specifies a set of hard challenging problems. For example, researchers working in Bioinformatics have a fairly well-defined goal that is coherent across the entire field. Same is the case with other new areas like sensor networks, quantum computing etc. In fact, most fields within the applied sciences have a well-defined vision that enables the community as a whole to present a unified pitch.

The problem with the current state of affairs in ICTD is that our vision is better articulated for the development specialist than for the computer scientist. While I definitely think that there are several significant research challenges within the ICTD space, we as a community have not clearly articulated these challenges to ourselves and to the CS and development communities at large. Worse still, many of us within the ICTD space are not sure about these grand challenges. This confused outlook is problematic and detrimental in the long run.

As one first step, I will articulate some of the interesting technical challenges in this space that I have worked on or come across (this list is by no means complete).

2. CHALLENGES

2.1 Low-cost connectivity solutions

Goal: Develop network connectivity solutions that can provide basic communications in developing regions at extremely low costs. Any such solution has to be economically viable and sustainable. Existing connectivity solutions are extremely expensive and not economically viable especially in rural areas.

Existing works: WiLD networks, Mesh networks, Mechanical backhauls, Delay Tolerant Networks (DTN).

Impact factor: Wireless solutions deployed in 15+ countries and used by millions. DTNs are also deployed in many countries.

Research contributions:

1. New wireless protocols for tailoring WiFi to achieve high performance over extremely long distances.
2. New protocols for routing, transport, naming and addressing in delay tolerant networks.
3. New network management solutions to ease the deployment and maintenance of rural wireless networks.
4. Building solutions for handling power fluctuations and lack of sustained power.

Possible research directions: Within the wireless space, we are currently investigating these directions:

1. Designing high performance multi-radio wireless networks.
2. Significantly enhancing performance of point-point links by leveraging better physical layer solutions such as MIMO, steerable antennas and Analog Network Coding.
3. Design a unified MAC layer for combinational wireless distribution networks comprising of point-point, point-to-multipoint and omni-directional links.
4. Building a new WiFi-based cellular architecture that consumes much lower power and incurs significantly lower cost.
5. Data over GSM: Achieve a data-connectivity layer on the cellular voice channel.

2.2 Extending the Web to Developing Regions

Goal: Improve the penetration of the World Wide Web in developing countries especially in areas with poor, limited or no affordable network connectivity. Even in places where good connectivity

exists, the Internet connection is often shared across several users.

Existing works: TEK, RuralCafe, HashCache, EdgeXL, Contextual Search Portals, Riverbed Networks.

Impact: A few deployments in a few targeted locations. Riverbed networks is a large commercial entity.

Research Challenges and Contributions:

1. Network protocols require a complete redesign. HTTP and TCP work very poorly in low bandwidth networks and Web browsing is an extremely slow process. RuralCafe and TEK support an intermittent web browser interface to expose the intermittency to the user. We are currently building a new transport layer that can alleviate many of the transport problems in shared low-bandwidth networks.
2. How to build large Terabyte caches to support offline web access in low-bandwidth networks? We are currently building a contextual search engine that can fetch the vertical slice of the Web on any given topic and store this slice on a hard-disk. The idea is to perform offline content distribution and support offline web access by distributing hard disks and DVDs on specific contexts.
3. Wide-area network acceleration (EdgeXL, Riverbed) perform duplicate traffic elimination using block Rabin fingerprints to identify arbitrary length duplicates on the fly.
4. Existing web proxies do not scale for TB caches. HashCache presents a modified memory indexing system to make web proxies scale to support TB caches. We are also working on new algorithms to support fast local search on TB caches.
5. There are several works which have performed on-the-fly content adaptation to reduce the size of web pages to enhance web browsing experience in low bandwidth networks. Given the widespread prevalence of Web 2.0, existing web pages are no more plain HTML and there are many unaddressed challenges in content adaptation.

2.3 Authentication and Identity Management

Goal: Authentication is a fundamental problem that arises in various contexts in developing countries. We need low-cost usable authentication mechanisms appropriate for developing regions. Paper is often the common medium for transactions in these settings and paper forgery is among the leading causes of corruption in many developing regions. Our goal is to develop a low-cost authentication mechanism to verify the authenticity of any piece of paper.

Existing works: Paper authentication, Speech based authentication, biometrics, smart cards.

Research Contributions:

1. We have developed, PaperSpeckle, a low-cost, robust, tamper-resistant paper watermarking technique that extracts a unique watermark for any piece of paper based on the natural randomness present in the structure of the paper. We also show that this verification can be done using a simple mobile device. We show a theoretical result that establishes the hardness in forging the authentication mechanism.
2. We are working a combinational protocol of using speech-based authentication in conjunction with challenge-response based cryptographic techniques to develop a usable authentication mechanism for branchless banking.

Applications: Stamp paper verification, Drug counterfeiting, currency notes, supply chain management.

Identity Management: Closely related to authentication, the notion of identity in most application scenarios in developing countries is very weak and it is significantly hard to overhaul the entire system with a new identity management system. The goal is to develop a scalable and sustainable identity management system suited for developing regions. Cryptographic solutions are not the easiest to deploy since such solutions are not easily usable. Biometrics are possible options but they require a large manual setup process. We are investigating one solution based on leveraging SIM cards as a trusted computing base to bootstrap an identity management system for developing regions.

2.4 Mobile Apps for Developing Regions

Goal: Develop new viable mobile applications that can enable a new class of services in health care, financial services and other areas. While mobile devices are touted to solve a large number of problems in developing regions, there are some important limitations to understand. First, not every one has a smart phone or a phone with a data connection plan. If restricted to only voice and SMS, the functionality is extremely constrained. Second, cellphone usage rates are extremely high and the application needs to be extremely curtailed in bandwidth usage. Third, we are the mercy of the operator for mass adoption which is not always desirable.

Challenges and Contributions:

1. Given that most users have low-end phones, one of our thrust areas is SMS-based mobile applications. We have currently built solutions to four problems: (a) SMS-based health records; (b) Secure drug tracking; (c) Mobile Craigslist; (d) Rural ATMs. Key ideas to enable these applications include semantic compression of updates, a lightweight reliability layer for SMS, supporting aggregation of operations and lightweight privacy and security.
2. TCP/IP is unsuitable for intermittent environments. While the DTN research community, tries to provide a TCP/IP semantics over delay tolerant links, often hiding the intermittency from the application is not a good idea. Given that most applications operate in synchronous environments over TCP/IP, tailoring these applications to intermittent environments is often a difficult challenge. Some traditional synchronous applications like ATM machines need to be fundamentally redesigned when tailored for these asynchronous intermittent environments. To achieve this our rural ATM system supports offline authentication along with an approximate consensus protocol to support cash availability.
3. Mobile platforms need to be extensible and should support easy installation of new applications without operator need or involvement. One possibility we are investigating is to build a class of applications on a common platform, where the application themselves are instantiations of the platform specified as configuration parameters to the underlying platform (like creating a web page). We are currently working towards developing an extensible platform to support a wide range of SMS-based apps.
4. As an alternate to GPRS, we are developing a transport layer for Data over GSM to achieve 2-2.5 Kbps on the voice channel. We can currently achieve close to 2.5 Kbps with a 10% loss rate on the voice channel. This represents a low-bandwidth and low-cost alternative to GPRS.

2.5 Other Important Research Challenges

To enumerate a few other important challenges:

1. **Appropriate user interface design:** A large fraction of the development-targeted population is either semi-literate or illiterate. The user interface is the single deciding factor on whether a system is deemed usable and is widely used by local populations. There are several HCI and UI design challenges in the ICTD space that warrant careful consideration

and there is a sub-community within ICTD studying this space.

2. **Local language challenges:** Each developing region has their own local language with its own spoken and written dialects. This opens up many research challenges to both the speech and NLP research communities. Two of the biggest challenges in local language are speech recognition and content generation, translation and presentation. Addressing either or both of these issues is critical to make systems usable in the local language.
3. **Low, unreliable and intermittent power:** Systems should be reliable and robust in the face of unreliable and intermittent power. Designing low-power computing systems is currently receiving a lot of attention within the CS systems community (HotPower workshop) but much of their focus is on power issues in the developed world where they assume the presence of a sustained reliable power source. Power challenges in the developing world are technically harder to address with lack of reliable power sources and significant power fluctuations. We need to draw this low-power research community to address power design challenges in the developing world.
4. **Computer vision challenges:** We recently developed an automated diabetic retinopathy system that uses machine learning techniques to detect diabetic retinopathy disorders in retinal images. This turns out to be fairly hard image processing and vision problem and existing solutions do a poor job of solving this problem. Our system has high specificity with low false positives and false negatives and improves over prior detection systems. We are deploying our system in Aravind Eye Hospitals where we believe this solution can significantly alleviate their diagnosis burden on 2 million patients tested for diabetic retinopathy every year.

3. SUMMARY

To end on a positive note, there is a wealth of really hard technical challenges in the ICTD space that the CS community should jump on. The beauty is the breadth of challenges in this space covering different disciplines within computer science with significant depth in each area. We, as a community, should be able to better articulate these challenges to both the CS and the development communities at large. I have tried my best to begin articulating a few of these technical challenges in greater detail.