Opportunistic Data Services in Least Developed Countries: Benefits, Challenges and Feasibility Issues

Nicolas Anciaux, Luc Bouganim, Thierry Delot, Sergio Ilarri, Leïla Kloul, Nathalie Mitton, Philippe Pucheral

Opportunistic Data Services in Least Developed Countries: Benefits, Challenges and Feasibility Issues

Nicolas Anciaux[1, 2], Luc Bouganim[1, 2], Thierry Delot[3,1], Sergio Ilarri[4], Leila Kloul[2], Nathalie Mitton[1], Philippe Pucheral[1, 2]

(1)INRIA, France fname.lname@inria.fr
(2)PRISM, UVSQ, France fname.lname@prism.uvsq.fr
(3)LAMIH, UVHC, France Thierry.Delot@inria.fr
(4)Univ. of Zaragoza, Spain silarri@unizar.es

ABSTRACT
According to many studies, IT should become a key facilitator in establishing primary education, reducing mortality or supporting commercial initiatives in Least Developed Countries. The main barrier to the development of IT services in these regions is not only the lack of communication facilities, but also the lack of consistent information systems, security procedures, economic and legal support, as well as political commitment. In this paper, we propose the vision of an infrastructure-less data platform well suited for the development of innovative IT services in Least Developed Countries. We propose a participatory approach, called Folk-IS, where each individual implements a small subset of a complete information system thanks to highly secure, portable and low-cost personal devices as well as opportunistic networking, without the need for any form of infrastructure. In this paper, we focus on the exploitation and feasibility analysis of the Folk-IS vision. We also review the technical challenges that are specific to this approach.

1. INTRODUCTION
Least Developed Countries (LDCs) are 50 countries with nearly 1 billion people, which meet UN criteria in terms of poverty, human resource weaknesses and economic vulnerability. LDCs have no global access to the Internet and are more than ever left by the wayside with respect to Information and Communication Technologies (ICT).

Does this mean that e-services are a superfluous luxury for LDCs? On the contrary, several reports [14, 20, 31] make evident that ICT are called to play a catalytic role in these countries: ICT can help to achieve universal primary education, promote gender equality and empower women, reduce child mortality, combat HIV and other diseases, ensure environmental sustainability, and develop a global partnership for development.

However, the challenge of making e-services practical and affordable still remains. While 60% of the population in LDCs is already covered by a mobile cellular signal, only 0.5% has a mobile broadband subscription and a 3G service is offered only in at most 25% of the LDCs, very often at a prohibitive cost [20]. Hence, mobile phones in these areas are primarily feature phones used for voice and SMSs. Several innovative solutions are thus based on the use of mobile phones and text messages to address specific issues like keeping track of vaccine cold chains [13], improving agriculture [33], or easing administrative procedures [29], but cannot be generalized to the broader scope of data-driven applications. Major industrial actors like Google (project Loon for All) and Facebook (Internet.org) are also conducting pioneering projects in this field, showing the potential for new applications and business in these areas. However, providing general data services in LDCs faces many obstacles which go far beyond low connectivity: severe lack of IS infrastructure, high initial investment, difficulty to maintain the system operational, reluctance to use the system due to security concerns, etc.

According to Non-Governmental Organizations (NGOs), four main requirements must be met to build a practical technical solution: (1) **privacy protection**: this is a major prerequisite due to local opaque practices and the lack of any security infrastructure (coercive laws, secured servers, trusted authorities, etc.), leading to a self-enforcement of privacy principles; (2) **immediate personal benefit**: because of the lack of strong economical or political incentives to impose the solution in the field, an immediate benefit must be provided to each user; (3) **self-sufficiency**: the solution must not rely on an hypothetic improvement of the existing software and hardware infrastructure; and (4) **very low deployment cost**: the usual scale being a few dollars per user. Besides this, users' empowerment is known to be crucial to make the solution sustainable in LDCs, and its maintenance should ideally generate a source of revenue for new local jobs.

In [3], we proposed the vision of trusted cells, a data platform for personal data services where the shared infrastructure (typically the Cloud) is untrusted, while personal devices (such as smart phones, tablets or set-top boxes) are trusted execution environments. In this paper, we revisit this vision with the goal of providing data services to the inhabitants of LDCs. We propose a new paradigm, that we call *Folk-enabled Information System* (Folk-IS), based on a fully decentralized and participatory approach, where each individual implements a small subset of a complete information system without the need for a shared networked infrastructure. As trusted cells, Folk-IS builds upon the emergence of highly-secure, portable and low-cost storage and computing devices, called hereafter

Smart Tokens. Here, however, the focus is on the low-cost of ownership, deployment and maintenance, and on the absence of a networked infrastructure. In Folk-IS, smart tokens have the capability to host the digital history of their owners and to exchange information in a privacy-preserving manner. This enables people to transparently and opportunistically perform data management and networking tasks as they physically move, so that IT services are truly delivered by the crowd.

The potential advantages are important for the inhabitants (develop economics activities, hold a digital folder, communicate with friends, with NGOs or administrations, etc.), for the community (have communication means with inhabitants used to feed cultural programs or launch e-admin procedures like census, detect epidemics, etc.), and for the NGOs (conduct monitoring programs, consult/fill inhabitants folders when visiting them, etc.).

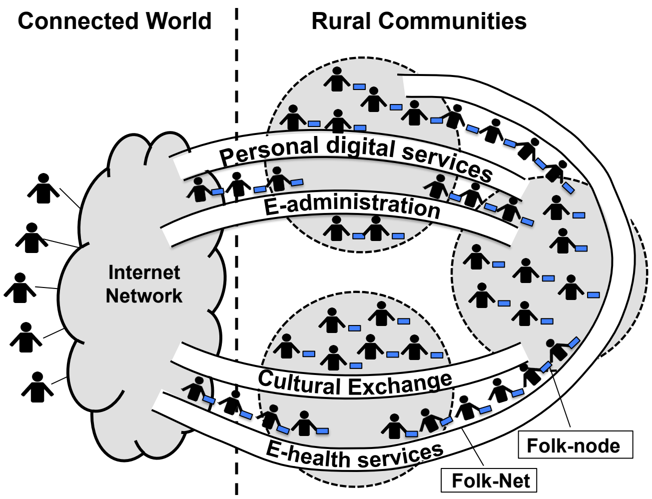

Figure 1. Folk-enabled information system

We do not argue that Folk-IS is the ultimate solution. The future of IT in LDC will probably be multiform, the problem being important and complex enough to leave room for complementary initiatives. Folk-IS has the salient characteristics to enable a smooth and incremental deployment of an information system in a purely infrastructure-less context while taking advantage of existing elements of infrastructure, if any, to improve its own behavior.

In [4], we presented our overall Folk-IS vision. This paper details this vision and shows that this paradigm technically makes sense, conforms to the requirements mentioned above and opens important and exciting research challenges for the database community.

2. REPRESENTATIVE ICT INITIATIVES IN LDCs

Applying ICT in developing countries has recently attracted considerable attention, reflected in recent special issues, such as [14, 31]. Several of the proposed applications aim at making use of mobile ICT services to resolve key challenges such as providing content that is locally relevant and mitigating access barriers linked to literacy and language [40]. However, when developing data-driven solutions for LDCs, important difficulties arise. The coverage of mobile phones in LDCs is increasing, but phones are primarily used for voice and SMSs, not for data-driven applications. Sir Tim Berners-Lee mentioned a recent study that showed that in Mozambique "*using just 1GB of data can cost well over two months wages for the average citizen*"[1]. According to many analysts, this situation cannot evolve rapidly due to a combination of intrinsic barriers, not only technical, but also economical and organizational [9].

Many isolated initiatives have been launched with specific strategies for getting around these barriers, and propose ad-hoc solutions for specific problems. In relation to health issues, for example, we could mention *FoneAstra*, *CVDMagic*, NextDrop, and *mWash*. *FoneAstra* [13] is a low-cost sensing system using mobile phones to keep track of vaccine cold chains. *CVDMagic* [34] proposes the use of mobile-based approaches to detect Cardio-Vascular Diseases (CVD) in India. *NextDrop*[2] provides SMS alerts indicating when a tap will have water, at least 60 minutes in advance. Finally, *mWash*[3] and [10] consider the problem of collecting and disseminating information to ensure a safe access to sanitized water. Other interesting topics tackled in relation to developing countries are related to agriculture [33], and e-government [29], among others; see [28] for more examples of interesting ICT applications for developing countries. These solutions address key issues in developing countries and bring actual improvements. However, despite their undisputed interest, these initiatives usually exploit only text messages for exchanging data, and remain confined to specific applications. Their generalization also faces other barriers like the lack of global information system infrastructures, of security procedures, and economic and political support.

Supporting more powerful form of data-driven applications to provide global e-services undeniably requires improved network connectivity. However, according to many analysts, even if increasing the number of "urban" spots would be an obvious solution, their deployment is very challenging as it would require more base stations, and hence more energy with collateral difficulties, like the need to carry gas-oil and protect it against theft [20]. To overcome the limitations of Internet coverage, several pioneering proposals do not rely on a hypothetic future improvement of the infrastructure, but propose new alternatives. Thus, proposals [30, 35] exploit mobile ad-hoc networks to provide asynchronous connectivity for data-oriented services. Specifically, in *DakNet* [30] mobile access points are mounted on physical means of

---

[1] http://a4ai.org/press-centre/what-our-supporters-say/
[2] http://nextdrop.org/
[3] http://www.pacinst.org/reports/mwash

transportation (e.g., buses, motorcycles) to transport data or information requests from one place to another. The related approach in [35] suggests the use of buses and cars to carry data between villages and Internet gateways. In relation to these proposals, the so-called *Delay-Tolerant Networks (DTNs)* [15] are very interesting, as they suit well with the idea of benefitting from opportunistic links when possible. Moreover, they show that huge benefits (e.g., email services) can be provided to users by means of asynchronous communications, even if this is at the expense of a potentially (unavoidable) high latency.

Google project Loon[4] is another ambitious attempt to connect people in rural and remote areas and bring people back online after disasters thanks to a network of high-altitude balloons. This is promising, although it still remains unclear if such a network infrastructure is durable, since current balloons can only function for a few weeks. In Cape Town (South Africa), Google also participated in *The Cape Town TV White Spaces (TVWS) trial*[5], which evaluated the possibility to benefit from unused parts of the frequency bands used for television to enable Internet access [28]. For its part, *Internet.org* promotes low-cost wireless handsets, compression of Web pages, and data caching on the edge of the network, to reduce data transfer. It is complemented by the Alliance for Affordable Internet[6], a coalition including the public sector and private companies (such as Google and Facebook) as well as civil organizations, which pushes towards progresses in regulation and policy.

These different initiatives demonstrate the growing interest, as well as the high difficulty, to integrate LDCs in global IT networks. Obviously, it is still not possible to predict which of those attempts may succeed, and when. Loon project architects believe that this can achieve a commercial success, but confess that it may take a lot of time. In addition, low connectivity is clearly not the ultimate and unique barrier. As already mentioned, the generalization and deployment of e-services in these regions faces many other obstacles: the lack of consistent information systems, security procedures, economic and legal support as well as political commitment, and the difficulty to ensure the maintenance of the system.

The specificity of our proposal is to consider the aforementioned obstacles as characteristics of LDCs and to propose a holistic and general data management solution that supports secure data-intensive applications, while taking advantage of existing elements of infrastructure, if any, to improve its own behavior. So, we consider initiatives focusing on communication aspects, like DakNet or project Loon, as complementary to Folk-IS, since any additional means to reduce the network latency and increase connectivity would benefit our vision.

[4] http://www.google.es/loon/

[5] http://newweb.tenet.ac.za/tvws/.

[6] http://a4ai.org/

3. FOLK-IS ARCHITECTURE

As stated in the introduction, the Folk-IS paradigm builds upon the emergence of smart tokens, i.e., highly secure, portable and low-cost storage and computing devices, which may have various form factors (i.e., physical layout) and characteristics. In this section, we consider an abstract *Smart Token*, to focus on the foundation of Folk-IS and on its mode of operation with no consideration for specific technical choices. We postpone the discussion on real hardware platforms and concrete architectures to Section 5.

Smart Tokens: We consider Smart Tokens embedding at least: (1) enough stable storage to host the complete digital environment of its holder, (2) enough tamper-resistant computing resources to run a server managing the data and enforcing access control rules, and (3) a biometric sensor to authenticate users (e.g., a fingerprint reader, well adapted to illiterate people). Smart Tokens require also input/output capabilities to interact with users and communication facilities (e.g., short-range communications) to exchange messages. To meet the low-cost requirements, a smart token may inherit part of these functionalities from a terminal it connects to, assuming that it does not introduce security breaches.

Folk-enabled Network: since Folk-IS cannot rely on existing communication infrastructures, an ad hoc network of smart tokens is used for transferring messages (i.e., network packets) among smart tokens and from smart tokens to Internet access points. This can be achieved either by using short-range communications between smart tokens (assuming they are equipped with such communication facilities) or through terminals smart tokens connect to (assuming messages are uploaded on the terminal and retrieved by others smart tokens which will next connect to it). So, communications are opportunistic by nature.

Folk-IS Personal Node (Folk-node): it is associated with each individual and refers to the combination of a Smart Token and of the embedded software components required to manage the holder's data, act as a network node, and enforce the security of the whole system.

Folk-enabled Information System (Folk-IS): based on the existence of Folk-nodes, we can devise a system implementing the three main functions of an information system as follows:

- *Communication management:* by carrying and routing data, each Folk-node acts as a node in the Folk-Net. Assuming that each Folk-node maintains a history of its moves (e.g., by gathering the GPS coordinates of all the devices it connects to), it can forward messages to encountered Folk-nodes (directly, or through terminals) whose moving profile best matches the recipient's location. Such georouting protocol provides a much better resource utilization of the Folk-Net than a basic flooding protocol.
- *Data management:* each individual centralizes in his Folk-node all his personal data (e.g., medical records,

administrative documents, credentials). Each time the holder interacts with a data source (e.g., by physically meeting a data provider, like a doctor, or by receiving a document through the Folk-Net), the application simply inserts these data in a local Folk-node database. From the information system viewpoint, the global database is the "union" of all these local databases. Each Folk-node is assumed to participate in common services, like processing global queries (broadcasted through the Folk-Net) and ensuring data durability (thanks to data replication among Folk-nodes).

- *Access control & authentication:* each Folk-node controls the access to the data it hosts and strongly authenticates the requesters (and its owner). Performing this control on the user's side is the ultimate solution to increase the holder's confidence, enforce his consent, and minimize the benefit/cost ratio of an attack. Indeed, the complexity of attacking the system is increased by the Smart Token tamper-resistance and by the obligation to be physically in contact with it to attack it. In parallel, the benefit of the attack is limited to disclosing/corrupting the data of a single individual. Note that the holder himself must authenticate and does not have all the privileges on his own Folk-node (e.g., he cannot tamper his own medical data to prescribe himself new drugs). In addition, messages carried by a Smart Token or replicas from other individuals stored locally are encrypted with keys never available to that Smart Token.

Hence, the complete system (data storage and network facilities) progressively deploys itself as Folk-nodes and shared devices are distributed to the participants. Folk-IS is by construction highly-redundant and robust, it does not require any central administration, and its global cost is directly proportional to the scale of the targeted population. As discussed next, some Folk-IS functionalities could be delegated to specific workers (e.g., people renting terminals or acting as postmen) in order to improve the quality of service (QoS) while creating a new local economic model (e.g., like people renting their cell phones in LDCs [9]). If elements of infrastructure are available, the Folk-IS QoS also improves accordingly, decreasing the network latency by shortening the route to the destination (or to the nearest Internet access point).

Figure 2 illustrates the Folk-IS mode of operation. It shows two rural communities, residents (black icons with letters) and their possible moves (grey icons), a *school* and a *first aid* room used by residents from both communities, and an Internet access point. Data exchanges through terminals are represented by dashed lines (e.g., the Folk-node of person A is used from the terminal of nurse N in the infirmary), and short-range data exchanges are represented by pink halos (e.g., between A and B in *Community 2*). E serves as a *netman* (i.e., a postman conveying digital messages) and carries data when travelling from place to place (e.g., the school, the infirmary and the Internet access point) thanks to his Folk-node. Different routing paths to transmit a message from A to the Internet are represented, e.g., the path ABTE: person A → person B → T (teacher) →E (netman) → Internet. These paths benefit from the physical mobility of people and their devices.

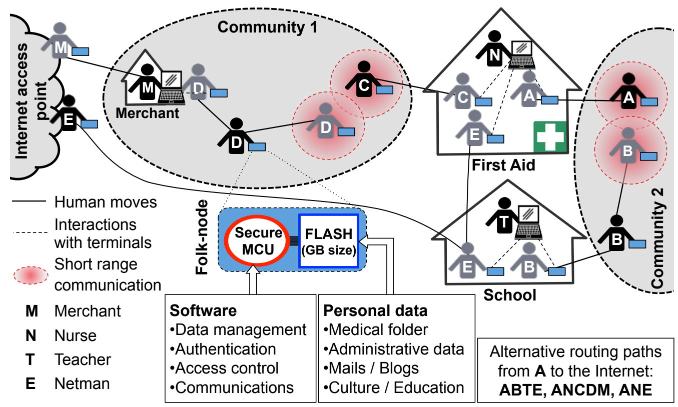

Figure 2. Folk-IS: mode of operation

4. SCENARIOS

Based on this mode of operation, several concrete scenarios can be envisioned. For illustration, we show examples below.

Healthcare scenario: In most LDCs, healthcare is provided by local nurses/doctors working in first aid rooms or intervening in rural communities during drought periods or health programs. They only possess very basic medicines and equipment. Needless to say, they do not benefit from an Electronic Health Records (EHRs) system, which is already highly difficult to organize in developed countries. Even paper-based medical records are too complex to maintain because many people have no identity document to link them to their records and people move according to seasons and drought periods. By using a Folk-node, each patient owns his complete and up-to-date medical folder. Without any Internet connection, a nurse can access the fingerprint-authenticated patient's medical folder, and appends diagnoses and treatment information. She may request medical advice from a distant clinic, by transmitting documents (e.g., a picture of the patient's injury) from the patient's Folk-node through the Folk-Net. She can notify the patient a few days later, still through the Folk-Net, in case a serious problem has been detected. In this case, the same architecture could also be used to detect and anticipate epidemics to better limit and fight them. Global queries can also be broadcasted through the Folk-Net to conduct epidemiological studies or maintain health indicators on populations unreachable so far. Hence, each Folk-node acts as a smart Personally Controlled Electronic Health Record (PCEHR).

Cultural heritage: Benefiting from the distributed and traveling Folk-IS architecture, new practices could be put in place to produce and share pieces of cultural heritage, within communities or towards the outside world through virtual museums. Thus, with the help of cultural organizations, users able to produce pieces of a cultural heritage, related to traditional stories, songs, beliefs or cultural events, could be identified and equipped with Folk-nodes. Other users able to describe or translate such productions could also be equipped, and share/publish the enriched cultural content to virtual museum. The exact contribution of each actor (e.g., content producer, descriptors, translators, etc.) could be recorded, watermarked and certified by his tamper-resistant Folk-node. This may generate a new economic model where actors could be paid according to their contributions.

Generalization: Many usual scenarios can be transposed to rural communities thanks to Folk-IS, provided that they accommodate high-latency data exchanges: scholar folders for children, e-administration procedures (e.g., drought warnings, recording births and deaths), personal data applications (e.g., email, social or networking services), etc. This may also generate a new economic model where every actor could be paid according to his contribution, as exemplified in the cultural heritage scenario. It could also be the case for people playing the role of a *netman*, who could be paid according to the volume of transported data, the amount of population touched and the distance covered.

5. CONCRETE FOLK-IS INSTANCES

Let us now go from abstract smart tokens to the concrete hardware platforms. Given our target, an essential element in the design of the Folk-node's hardware is its cost, since it determines alone the overall cost of the Folk-IS solution. To decrease this cost, we study whether some of the Folk-node's hardware resources can be shared while minimizing the impact on the Folk-node's functionalities. To this end, we distinguish between Folk-node hardware platforms associated with individuals and shared devices which are either made publicly available in specific places or owned by local workers (e.g., nurses, teachers, administration officers, merchants). In the following, we first describe each Folk-node's resource arguing the pros and cons of sharing it. Then we examine the design of shared devices. Finally, we consider the whole Folk-IS architecture and discuss some deployment scenarios.

5.1 Folk-IS Hardware Platforms

5.1.1 *Resources in Folk-nodes*

As mentioned in Section 3, Folk-nodes are expected to embed storage, security, communication, location, and I/O resources.

Storage resources. Flash memory is undoubtedly the most adequate storage medium, thanks to its robustness, compact form factor, low cost, and low energy consumption characteristics. The main storage resources must be located on the Folk-node itself for obvious availability reasons (e.g., the holder's medical folder must be accessible anytime and anywhere when carried by the holder) and then cannot be shared. The storage capacity must afford the storage of all the holder's data as well as the persistent data required by common services (e.g., durability data, network packets, etc.).

Security resources. Secure microcontrollers (SMCUs) [11] can provide tangible security guarantees, and are today available at very low cost, in a small form factor (e.g., a SIM card), generally combined with an external (unsecure) flash chip. Existing SMCUs typically embed a relatively powerful processor (e.g., clocked at about 50 MHz), a cryptographic co-processor, and an internal secure stable storage (around 1 MB) to store the embedded code (e.g., the OS, a data management engine, the authentication mechanism, etc.) and sensitive metadata (e.g., cryptographic material used to protect the holder's data). Since we cannot assume that each user has an electronic identity (this would require certificates and a PKI –Public Key Infrastructure–, which contradicts our infrastructure-less assumption), we consider the use of a biometric sensor to prove the identity of the user. Biometric sensors (e.g., fingerprint readers) seem well adapted to illiterate users since this avoid remembering and typing a password or PIN code. At the personalization step (i.e., before the first use), the user's fingerprint is stored inside the SMCU's secure memory and it is then checked at each access.

SMCUs and fingerprint readers are at the heart of the security system of Folk-IS. They cannot be disassociated from the Folk-nodes (i.e., associated with the terminals the Folk-nodes connect with) without introducing the risk of class-breaking attacks (i.e., breaking a single-shared terminal would enable access to the sensitive data of all the Folk-nodes sharing that terminal). The privacy and security requirement is thus fulfilled, due to a combination of factors: (1) the guarantee of an implicit user's consent (no access to the data without the holder's fingerprint check) and of a strong authentication of local workers; (2) the obligation to be physically in contact with the device to attack it; (3) the tamper-resistance of the Folk-node's processing and storage units, which makes hardware and side-channel attacks highly difficult; (4) the certification of the embedded software platform, which makes software attacks also highly difficult; and (5) the impossibility, even for the Folk-node's holder, to directly access the data stored locally or involved into local computations (he must authenticate and can only get data according to his privileges).

Communication resources. To enable the human network described in Section 3, Folk-nodes need a means to communicate together. This could be achieved either by using short-range communications or by physically plugging Folk-nodes in shared devices: messages can be temporarily stored encrypted on the shared device and delivered to Folk-nodes that will be plugged later. The advantage of enabling Folk-nodes with wireless capabilities is that messages can be exchanged without help from shared devices. However, wireless communications need energy, thus imposing the need of a battery and some mechanism to charge it (e.g., energy harvesting technologies). Solutions with a limited amount of energy can also be exploited to improve the connectivity of the Folk-net, as discussed in Section 6.2.

Location resources. The network routing will be done based on the geographical location of the devices, raising the need to obtain at least an approximate location of Folk-nodes. Different alternatives could avoid embedding a costly positioning mechanism in each Folk-node. For example, the location information can be deduced from interactions with a small subset of localized Folk-nodes, shared devices, or fixed nodes. The user's home address can also be stored on each Folk-node at the personalization step, giving approximate static information about its location.

Input/Output (I/O) resources. Classical I/O resources must be available to enable user interactions with a Folk-node. A screen, a keyboard (or a touch screen), a microphone and a speaker are obvious I/O elements. They have however different form-factor and cost/energy requirements. Fortunately, most user interactions will happen with local workers and thus can make use of their own device's I/O resources. In addition, illiterate users will probably underuse the keyboard or the screen. Such I/O resources can thus be located on shared devices without a significant impact on the platform usability, while keeping a speaker and microphone on the Folk-node would allow basic users' interactions autonomously.

5.1.2 Energy considerations and concrete Folk-nodes

In an infrastructure-less context, the electrical power network is limited and leads to consider energy as a particularly scarce resource. To this respect, let us first consider two extreme designs for the Folk-nodes: passive and active ones (see the top of Figure 3).

The ***passive*** Folk-node embeds only mandatory resources (i.e., related to storage and security) and has no internal battery, and thus cannot be used without a shared device. It achieves the lowest cost at the expense of needing a shared device to be used. The intrinsic characteristics of a passive Folk-node are its low cost and robustness, inherited from its simplicity. It is very similar to a smart token product, provided by Gemalto, that we used in a field experiment [2, 3, 5]. This kind of smart token is available for a few dollars, in a SIM card form factor (plugged in a USB key casing), and simply needs to be extended with a fingerprint reader, in a way similar to traditional secure USB keys.

The ***active*** Folk-node is more complex, since it embeds a battery and the required means to charge it and power a wireless communication element (Bluetooth, led-light communications, ...). It is also equipped with I/O resources to allow basic users' interactions autonomously. Typically, we keep a speaker and microphone, given their small form factor, robustness and low-cost/low-energy profiles. Thus, active Folk-nodes provide certain functionalities to end-users even without being connected to a shared terminal, and they can communicate (typically, they can exchange messages wirelessly with surrounding active folk-nodes).

Capitalizing on various energy harvesting technologies (Piezo-electric, solar, etc.), other intermediate designs would make sense, leading to Folk-nodes being active intermittently, then called ***partially active*** Folk-nodes. For example, a Folk-node could be endowed with piezoelectric energy harvesting technologies and could become active at the will of its owner, a mechanical interaction with the equipment being needed to power the device for short durations.

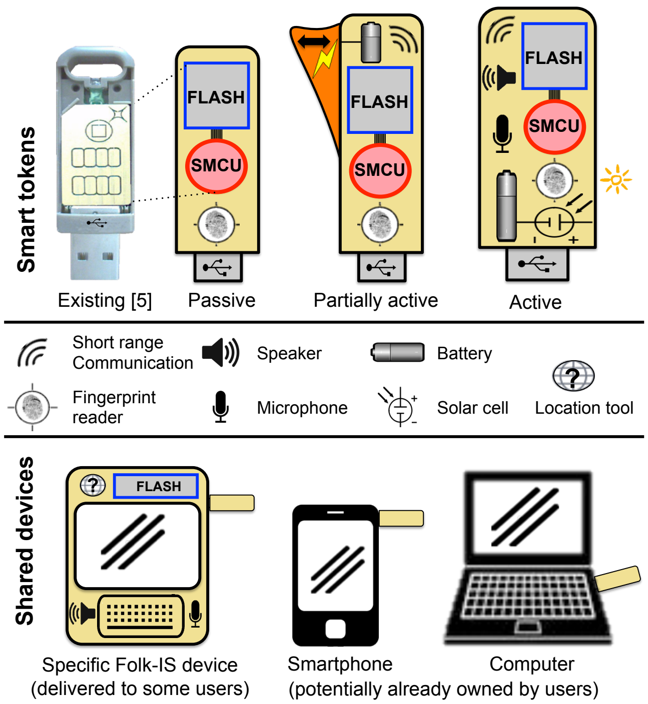

Figure 3. Folk-IS architecture elements

5.2 Shared Devices

Any hardware resource mentioned in the previous section and not present in Folk-nodes should be available in shared devices, that is: (1) for the passive Folk-nodes, all the I/O resources (screen, keyboard, speaker and microphone), some flash memory (for durability and communication purposes), and location mechanisms (for helping routing messages); (2) for the active Folk-nodes, a keyboard, a screen, and a location mechanism.

The bottom of Figure 3 shows three possible instances of shared devices. The specific Folk-IS device (on the left) is designed especially for the Folk-IS architecture. The idea is to increase the robustness of the deployed solution by providing a closed architecture, i.e., a device with no possibility to install any software or hardware components different from those developed and validated by Folk-IS (note that this does not mean proprietary hardware or software). An alternative is to make use of the devices that local workers potentially already own (e.g., a smartphone or a computer, as shown on the right of Figure 3) in order to further reduce the overall cost of the Folk-IS architecture. Indeed, the required shared resources are generally present in these devices. If a device misses some resource (e.g., it has no location facilities, or has a small available storage), then the service will be locally degraded, but this does not hinder the global architecture (typically, the estimated location will be more imprecise, the durability will be provided elsewhere, etc.).

5.3 Concrete Architectures and Deployment Scenarios

Based on the discussion above, different combinations of Folk-nodes / shared devices and different deployment scenarios can be envisioned. The initial deployment of Folk-IS should be based on the most simple and robust combination, in order to minimize the risks of a too-open or too-complex architecture. Thus, we first consider the combination passive Folk-nodes / specific Folk-IS shared device, whose main advantages are robustness and low cost (a few dollars for the Folk-node, less than US$100 for the shared device, with the cost of shared devices having a small impact on the cost of the overall solution).

Then, more expensive partially active or active Folk-nodes (of a few tens of dollars) could be progressively acquired and distributed to unequipped individuals requiring a lower latency of communications or wishing to access their Folk-nodes without the need of a shared device (e.g., by adding a microphone or speaker).

An orthogonal direction for enhancing the initial deployment is to open the Folk-IS solution to existing devices potentially owned by users or local workers. This means allowing Folk-nodes to connect to a larger set of devices (hardware adaptors may be needed to enable physical connections) and to upload applications (retrieved through the Folk-Net) to these devices. Such open architecture is more complex due to device heterogeneity but it brings many benefits: (1) it lowers the overall cost of the Folk-IS architecture; (2) it increases the number of potential terminals, thereby increasing data availability and reducing network latency; and (3) it allows third parties to develop interesting user-centric applications, thus increasing the benefit to the users.

These two directions (i.e., more active Folk-nodes and open Folk-IS) are not exclusive. We believe that a final Folk-IS architecture will indeed include passive Folk-nodes, active ones, specific Folk-IS shared devices, and existing devices, their proportion being highly dependent on the context.

6. DATABASES CHALLENGES

Several challenges of Folk-IS are at the crossroad of Computer Science, Economics, and Social Science. A first one is to formalize and quantify the impact of different Folk-IS configurations on e-inclusion, and more generally on social and economic progresses in LDCs. A second important challenge is linked to the study of specific human-computer interactions, made complex by the fact that some users might be illiterate and lack training. This section however focuses on database related challenges.

6.1 Co-design of the Embedded Data Management Engine

A primary role of a Folk-node is to ensure secure storage and sharing of all data forming the holder's *digital environment*: user authentication, secure data storage, query evaluation and access control enforcement. The fingerprint reader enables authenticating the requester (e.g., a nurse) at each access, so that privileges can be properly associated with him/her in order to protect data confidentiality and integrity. In a data management system, fine-grained privileges granted to subjects on objects are usually identified by expressions (i.e., queries) over the data. These queries determine the information that can be accessed by the subject (other data remain hidden). Hence, answering accurately the Folk-IS privacy and security requirement leads to embed a data management engine within the Folk-node, in charge of evaluating queries over the data, checking the integrity of any data involved in the computation, and delivering the result to the grantee subject. When considering a relational database, the queries are formulated using the relational algebra (involving selection, projection, joins and aggregation operators). Whenever documents are managed, queries are usually defined using expressions over tags/labels associated to the documents.

Evaluating queries within a Folk-node, even expressed using simple expressions over tags, is very challenging due to the strong and conflicting inherent hardware constraints of smart tokens. Whatever their exact characteristics, low cost and low energy, embedded devices are always endowed with: (1) a small amount of RAM (because of its low density and its high relative impact on the overall cost of the device); and (2) NAND flash memory (the most adopted persistent memory for embedded devices, due to its high density and high robustness), which exhibits bad performance for random writes. In addition, in the context of the Folk-node, the raw data should be protected using cryptographic features to ensure the confidentiality and integrity of the stored data. These constraints lead to contradictory objectives: executing queries with acceptable performance with a tiny RAM entails massively indexing the embedded database in NAND Flash, while index updates generate fine-grained random writes, and then unacceptable performance due to NAND flash write costs and cryptographic overhead.

These statements make any state-of-the-art solution impractical, and call for the design of brand-new data management techniques. Existing embedded DBMSs (e.g., SQLite, BerkeleyDB) and light versions of popular DBMSs (e.g., DB2 Everyplace, Oracle Database Mobile Server) target devices far more powerful than smart tokens (like smart phones or tablets), and do not tackle the abovementioned constraints. Most of the recent proposals addressing the problem of storage and indexing in Flash, like [1, 24], adapt the traditional B+-Tree to Flash memory by relying on a Flash-resident log to delay the index updates. When the log is large enough, the updates are committed into the B+-Tree in batch mode, to amortize the Flash write cost. The log must be indexed in RAM to ensure performance. To amortize the write cost by large factors, the log is seldom committed, leading to a higher RAM consumption. Conversely, limiting the RAM size means increasing the commit frequency, thus generating more random writes. Under strong RAM limitations, the commit frequency becomes de facto very high and the gain on random writes vanishes. Flash-aware implementations of key-value stores [39, 25] also use a log structure in Flash to store key-value pairs and exploit sequential writes. But again, an index must be maintained in RAM, with a relatively large size (~1B per key-value pair) to obtain acceptable performances.

Recent proposals more specifically handle the smart token constraints. Microsearch [38] is a search engine designed for sensor nodes, answering full text search queries in a pure pipeline fashion, thus fulfilling the tiny RAM constraint. MiloDB [6] is a database machine for future generations of SIM cards with large storage capacities (a NAND Flash chip is superposed to the secure microcontroller of the SIM card). The database is organized into sequential *Log Containers* in NAND Flash, supports a massively indexed database while avoiding random writes by definition, and is able to compute SQL star join queries within the microcontroller. These proposals demonstrate the growing interest of storing and querying large amounts of data in smart tokens, but remain restricted to particular data models and specific types of smart tokens.

In Folk-IS, a much broader scope in terms of data models and hardware components have to be investigated to co-design and implement the best storage, indexing and query engine considering a triple objective: reducing the overall cost of the device, reducing its energy consumption, and offering a high level of flexibility. To this end, other hardware choices will have to be envisioned. For example, a combination of a very low cost secure (tamper-resistant) microcontroller coupled with a regular microcontroller could be preferred. It would provide the expected level of security, would be more flexible, and would consume less energy (regular microcontrollers have a larger RAM size with respect to tamper-resistant ones, which increases performance and thus reduces energy consumption). Regarding persistent memory (NAND Flash), a micro-SD card would undoubtedly offer better flexibility and lower cost than a raw Flash die: the card could be replaced at will by a cheaper, larger or more efficient one, and it could also be kept when replacing a deficient device. Supporting micro-SD cards has a strong impact on the software design since it leads to access the NAND Flash through a Flash Translation Layer (FTL). An FTL is a software layer, proprietary, and potentially different for each micro-SD card, leading to different behaviors in terms of access time, read/write performance ratios, number of sequential data files that can be written sequentially, etc. Data management techniques designed for such a variable environment must be adaptive: the frequency of index refreshment would have to adapt to the performance of the micro-SD card, as well as many parameters regarding the storage and indexing structures.

6.2 Towards Folk-Net Routing and Mobility Prediction

We present here the global concept the communication protocols should enable in the Folk-IS paradigm. We also describe the directions we will investigate to achieve this based on existing approaches. The communication should be enabled either directly between active Folk-nodes (i.e., both equipped with communication facilities) or indirectly through a shared device (e.g., encrypted messages are exchanged as passive Folk-nodes successively connect to the same device, acting as a relay at a clinic, a school, etc.). Moreover, some humans may play the role of "netmen", gathering data to relay ("letters") and physically carrying them to other shared devices (or "intermediate delivery points") for further distribution. Folk-Net can be homogeneous (i.e., exclusively composed of active or passive Folk-nodes) or hybrid, in which case network latency and throughput will depend on the proportion of each type of node and of netmen. Folk-Net will thus rely on *opportunistic communications* [19] by nature, which poses a set of challenges in terms of routing, mobility prediction and energy consumption.

Routing protocols

The choice of the best candidate(s) to carry the encrypted data towards their destination is a primary concern to avoid flooding the human network. Many routing protocols have been proposed for wireless networks. *OLSR* and *AODV*, designed for *MANETs (Mobile Ad hoc Networks)* [23], are disqualified in the Folk-IS context due to the great amount of messages they require. *ZigBee* cannot be considered either. Indeed, it relies on a routing tree and thus is not reliable against node mobility. Geographic routing protocols are more interesting candidates as they do not need to maintain routing tables and work nearly stateless [32]. However, they rely on the following assumptions: (1) the location of the packets' recipient is known, and (2) every node capable of routing information knows its own position and trajectory. Both assumptions cannot be assumed in the Folk-Net context, motivating new research efforts.

Folk-IS also shares some similarities with *VANETs (Vehicular Ad Hoc Networks)* [18]. Whereas some features differ significantly (e.g., the underlying communication standard, the communication range, the speed and power of nodes, etc.), both environments share the same lack of infrastructure and dynamic topology. Hence, we believe that protocols designed in the VANET context, like *carry-and-forward* [16] or *content-based dissemination* [12], could be adapted to Folk-IS. Carry-and-forward protocols are well suited to sparse networks. They exploit some nodes as "data mules" to carry the information and deliver it to other nodes encountered along their trip. On the contrary, content-based dissemination protocols limit the number of nodes relaying the message in dense networks.

So, in the Folk-IS context, carry-and-forward protocols could be investigated to deliver messages (e.g., a query result), whereas content-based dissemination could be considered to efficiently disseminate a query to the nodes holding potential results.

Mobility prediction and profiling

As commented in Section 5.1.1, no exact GPS location can be assumed for the Folk-IS nodes. Instead, their location must be approximated from interactions with a small subset of localized Folk-nodes, either shared devices or fixed nodes. Besides, active Folk-nodes can also compute an approximate location based on the position of their neighbors using techniques based on *RSSI (Received Signal Strength Indicator)*.

Approximate locations can be combined with mobility prediction techniques to predict the future movements of Folk-nodes. Many mobility prediction techniques only rely on the history of the user's mobility patterns and thus are insensitive to the changes of the user behavior. Other techniques are based on both the history of the mobility patterns and formal models (e.g., [26]). In [7, 8], a new approach is investigated combining the notions of local and global profiles. While local profiles capture the recurrent behaviors of each user (like going to the office every day at the same hour and using, more likely, the same paths), global profiles capture a set of frequently-followed paths by a certain percentage of users (main roads, trains, etc.). This approach achieves a much better prediction success rate and could lead to great delivery and energy performance in Folk-IS georouting. Local profiles could be maintained in each Folk-IS node for privacy concern while global profiles could be anonymized and hosted in shared devices during a certain period of time.

Routing with mobility prediction

New routing protocols mixing geolocation approximation, mobility prediction strategies and social interactions (e.g., [17, 41]) definitely deserve to be studied to select the best "data mules" for each message. In relation to this topic, it is worth mentioning that the problem of detecting and tracking "familiar strangers" (unknown people that one may frequently encounter during his/her everyday activities) has attracted the attention of research, due to the interest in understanding human behavior and diffusion processes [37]. This emphasizes the interest of considering the social behavior in the design of opportunistic routing protocols. As another example, [22] distinguishes between *stable nodes* that have frequent contacts with others within the same community (chosen as *community coordinators* for searching information within the community) and *highly mobile nodes* that frequently visit other communities (chosen as *community ambassadors* for searching information in other communities). These concepts could be applied similarly in the context of Folk-IS.

Energy saving

Active Folk-nodes having a limited autonomy in terms of energy must detect neighbors at low cost and form an ad hoc network whose topology and density may change over time (e.g., the network may be quite stable and dense in a village but quite sparse elsewhere, a dynamic neighborhood may become stable when owners of Folk-nodes enter the same bus, etc.). Every task should thus integrate energy consumption considerations. This could be done through transmission range adjustment when communicating (a node can increase/decrease its transmission power sending data more or less far and reaching more or less nodes) and set up of duty cycling (deciding when to turn off/on the radio to save energy). Range adjustment allows the reduction of the energy spent for sending (the message is sent to a closer node at a lower power) and for overhearing (it will reach less nodes), and generates less interference. It has to be included in the georouting protocol (like in [27]). The duty cycle is based on the node activity and its neighborhood. If a node does not need to communicate and there are enough active nodes in its neighborhood to ensure the relay of other nodes' messages, it can simply sleep and save energy. This operation can also be enhanced through the use of prediction models.

6.3 Application, Data Model and Access Control Deployment

As discussed in Section 4, Folk-IS enables many important applications, like email services, healthcare, and e-administration, some of them novel by their purpose (e.g., sharing pieces of a cultural heritage) or by their target (e.g., epidemiological studies on populations unreachable so far). The infrastructure-less context introduces new challenges with respect to application deployment, unified data modeling, identity verification, access control, and user's consent. Typically, data standardization, central application stores or central authorities identifying people, delivering certificates and enforcing access control rules on central servers, cannot be assumed. Conversely, a salient feature of Folk-IS is the ability to push the control at the edge of the network, that is to say (1) within each tamper-resistant Smart Token, and (2) through face-to-face interactions between users, implementing a de-facto user's consent. This may enable the reestablishment of local spheres of trust. For example, NGOs could produce applications, data models, home-made identification information, access control policies, and push them at the Folk-node level through Folk-Net. Specific workers could also be involved in the application installation and maintenance phases. This paves the way to a semi-decentralized way of managing and deploying applications, each Folk-node guaranteeing that (i) data produced by a given organization will not leak outside that organization, and (ii) the data owners' privacy is always respected.

6.4 Evaluation of Global Queries on a Population of Folk-nodes

Traditional techniques used to process queries in distributed databases or in peer-to-peer networks are not suitable, as they usually assume a good knowledge about the location of the data items in the network and do not consider the mobility of nodes. Here the data items will be dynamically distributed over a network of nodes that may be accessible or not at a certain moment, and the carrier of a replica of a data item may change at any time. Moreover, we cannot assume that all the data items relevant to a given query will always be available and retrievable in a reasonable time period. So, we should adopt the open-world assumption, admit the possibility of approximate answers, decide how to identify the relevant data sources without overloading the network (e.g., based on spatial conditions), how to route queries and results to their recipients (while preserving autonomy) using geographic routing and the physical mobility of nodes, and how to determine when a query and its associated routing tasks have to be finished. Moreover, new types of queries could be of interest in this context, requiring new query processing techniques; for example, reachability queries [36] could be used to study the possibility of propagation of a disease by analyzing past trajectories.

6.5 Structure and Calibration of the Folk-IS Architecture

Folk-IS is built on a large number of highly-secure but seldom-available Folk-nodes (active or passive) on the one side, and on less secure but more accessible and powerful shared devices on the other side. This unusual asymmetric architecture requires deeply rethinking the overall organization of an information system. This means distributing software resources on the hardware elements of the architecture, such that local and global applications can run and provide results with acceptable performance, security and resiliency. This also means associating different responsibilities to different devices (e.g., *super-nodes*), associating specific roles to humans (e.g., *netmen* to reduce latency), and designing new distributed protocols to implement global functionalities like query evaluation or data durability. For the latter, data replication is required because Folk-nodes may be lost, stolen, broken, etc. Data could be replicated based on the mobility profiles of the users. Confidentiality of replicas can be achieved through encryption, leading to the problem of choosing the most adequate set of Folk-nodes to hold a copy of the keys or of key shares (e.g., based on trust or similar mobility profiles). Finally, new simulation models are needed to calibrate IT resources according to the target applications and the local habits of residents. The objective is to determine suitable network topologies (given a certain density of Folk-nodes, communication frequency, etc.) with viable incremental deployments. Those simulation models will also help quantifying the expected gain in terms of social and economic sustainability, which is key for their adoption in the long term.

7. CONCLUSION

As mentioned in [9], time has come for research works, not only commercial initiatives, addressing the expectations of 80% of the population living outside developed countries. The *Folk-enabled Information System (Folk-IS)* paradigm combines low-cost secure devices, embedded software components and opportunistic communications, to meet fundamental requirements of LDCs. The promise of the solution is to fulfill the requirements stated in the introduction:

- *Privacy protection*: Folk-IS inherits its security and privacy protection from the tamper-resistance of each token and from the delegation of the access control and authentication tasks to each Folk-node, at the user's side. Moreover, any user must authenticate to the Folk-node, which securely records any performed action in an audit log, making the complete system accountable. The holder himself may be authenticated by the Folk-node, and may not have all the privileges on his own records.
- *Immediate personal benefit*: Folk-IS makes the Folk-node of each individual play the role of a passport for a wide range of critical applications: healthcare, e-administration, education, market rates for farmers, e-mails, etc. In parallel, governments and NGOs benefit from a new channel to pull/push information from/towards communities previously unreachable.
- *Self-sufficiency*: Folk-IS is infrastructure-less by construction. The complete system (data storage and network facilities) progressively and automatically deploys itself as Folk-nodes are distributed to the participants, with the ability to benefit from future infrastructure improvements.
- *Very low deployment cost*: Folk-nodes are hosted in low-cost smart tokens and the global cost of the system is directly proportional to the scale of the targeted population. In addition Folk-IS is by construction a highly-redundant and robust system that does not require any central administration. The maintenance and performance improvement of the system can be a source of empowerment with new local jobs (e.g., people renting terminals or acting as postmen), crucial to make the solution sustainable.

Hence, we expect that this paper could pave the way to exciting future works for our research community as well as to fruitful experiments in the field as soon as the major challenges are tackled.

8. ACKNOWLEDGMENTS

We warmly thank Léo Dayan, scientific director of APREIS NGO and of the Nomadic World University for Sustainable Development, as well as Pascale Pollack, director of E-NEXUS NGO, for precious discussions and feedback on the requirements of IT solutions for LDCs.

9. REFERENCES

[1] Agrawal, D., Ganesan, D., Sitaraman, R., Diao, Y., and Singh, S. 2009. Lazy-adaptive tree: An optimized index structure for flash devices. In *PVLDB*, 2(1).

[2] Allard, T., Anciaux, N., Bouganim, L., Pucheral, P., and Thion, R. 2009. Trustworthiness of Pervasive Healthcare Folders. *Pervasive and Smart Technologies for Healthcare*, Information Science Reference.

[3] Anciaux, N., Bonnet, P., Bouganim, L., Nguyen, B., Popa, I. S., and Pucheral, P. 2013. Trusted Cells: A Sea Change for Personal Data Services. In *CIDR*.

[4] Anciaux, N., Bouganim, L., Delot, T., Ilarri, S., Kloul, L., Mitton, N., and Pucheral, P. 2014. Folk-IS: Opportunistic Data Services in Least Developed Countries. In *PVLDB*, 7. http://www.vldb.org/pvldb/vol7/p425-anciaux.pdf

[5] Anciaux, N., Bouganim, L., Guo, Y., Pucheral, P., Vandewalle, J.-J., and Yin, S. 2010. Pluggable Personal Data Servers. In *ACM SIGMOD*.

[6] Anciaux, N., Bouganim, L., Pucheral, P., Guo, Y., Le Folgoc, L., and Yin, S. 2013. MILo-DB: A personal, secure and portable database machine. *Distributed and Parallel Databases*.

[7] Barth, D., Bellahsene, S., and Kloul, L. 2011. Mobility Prediction Using Mobile User Profiles. In *IEEE MASCOTS*.

[8] Barth, D., Bellahsene, S., and Kloul, L. 2012. Combining Local and Global Profiles for Mobility Prediction in LTE. In *ACM MSWiM*.

[9] Brewer, E., Demmer, M., Du, B., Ho, M., Kam, M., Nedevschi, S., Pal, J., Patra, R., Surana, S., and Fall, K. 2005. The Case for Technology in Developing Regions. IEEE *Computer* 38(6).

[10] Brown, D., Marsden, G., and Rivett, U. 2012. WATER Alert!: Using Mobile Phones to Improve Community Perspective on Drinking Water Quality in South Africa. In *ACM ICTD*.

[11] Bursky, D. 2012. Secure Microcontrollers Keep Data Safe. PRN Engineering Services, http://tinyurl.com/secureMCU

[12] Cenerario, N., Delot, T., and Ilarri, S. 2011. A Content-Based Dissemination Protocol for VANETs: Exploiting the Encounter Probability. *IEEE Trans. on Intelligent Transportation Systems*, 12(3).

[13] Chaudhri, R., Borriello, G., and Anderson, R. J. 2012. Monitoring Vaccine Cold Chains in Developing Countries. *IEEE Pervasive Computing*, 11(3).

[14] Coceres, R., Belding, E. M., Parikh, T. S., and Subramanian, L. 2012. Information and Communication Technologies for Development - Guest Editors' Introduction. IEEE *Pervasive Computing*, 11(3).

[15] Fall, K. R. 2003. A Delay-Tolerant Network Architecture for Challenged Internets. In *ACM SIGCOMM*.

[16] Fenu, G. and Nitti, M. 2011. Strategies to Carry and Forward Packets in VANET. In *Digital Information and Communication Technology and Its Applications, Communications in Computer and Information Science,* 166.

[17] Gao, L., Li, M., Bonti, A., Zhou, W., and Yu, S. 2013. Multidimensional Routing Protocol in Human-Associated Delay-Tolerant Networks. *IEEE Transactions on Mobile Computing*, 12(11).

[18] Hartenstein H. and Laberteaux, K. 2010. *Vehicular Applications and Inter-Networking Technologies*. John Wiley & Sons.

[19] Huang, C.M., Lan, K.C., and Tsai, C.Z. 2008. A Survey of Opportunistic Networks. In *Advanced Information Networking and Applications-Workshops*.

[20] ITU. 2011. The Role of ICT in Advancing Growth in Least Developed Countries – Trends, Challenges and Opportunities. http://www.itu.int/pub/D-LDC-ICTLDC.2011

[21] Jumira, O., Wolhuter, R., and Mitton, N. 2012. Prediction Model For Solar Energy Harvesting Wireless Sensors. In *Africomm*.

[22] Kang, C., Haiying, S., and Haibo, Z. 2014. Leveraging Social Networks for P2P Content-Based File Sharing in Disconnected MANETs. *IEEE Transactions on Mobile Computing*, 13(2).

[23] Kiess, W. and Mauve, M. 2007. A survey on real-world implementations of mobile ad-hoc networks. *Ad Hoc Networks*, 5(3).

[24] Li, Y., He, B., Yang, R. J., Luo, Q., and Yi, K. 2010. Tree indexing on solid state drives. In *PVLDB*, 3(1-2).

[25] Lim, H., Fan, B., Andersen, D. G., and Kaminsky, M. 2011. SILT: A memory-efficient, high-performance key-value store. In *ACM SOSP*.

[26] Liu, G. and Maguire, G. 1996. A Class of Mobile Motion Prediction Algorithms for Wireless Mobile Computing and Communication. *Mobile Networks and Applications*, 1(2).

[27] Mitton, N., Simplot, D., and Zheng, J. 2013. Guaranteed Delivery in k-Anycast Routing in Multi-Sink Wireless Networks ICST Mobile Communications and Applications. In *ICST*.

[28] Murugesan, S. 2013. Mobile Apps in Africa. *IT Professional*, 15(5).

[29] Ndou, V. 2004. E-Government for Developing Countries: Opportunities and Challenges. *The Electronic Journal of Information Systems in Developing Countries*, 18.

[30] Pentland, A., Fletcher, R., and Hasson, A. 2004. DakNet: Rethinking Connectivity in Developing Nations. IEEE *Computer*, 37(1).

[31] Rossi, G., Murugesan, S., and Godbole, N. 2012. IT in Emerging Markets. *IT Professional*, 14(4).

[32] Rührup, S. 2009. Theory and Practice of Geographic Routing. *Ad Hoc and Sensor Wireless Networks: Architectures, Algorithms and Protocols*.

[33] Sahilu, H., Villafiorita, A., Weldemariam, K., Belachew, M., and Zewge, A. 2012. Designing Distributed Agricultural Information Services for Developing Countries. In *ACM DEV*.

[34] Singh, P., Singh, A., Naik, V., and Lal, S. 2012. CVDMagic: A Mobile Based Study for CVD Risk Detection in Rural India. In *ACM ICTD*.

[35] Seth, A., Kroeker, D., Zaharia, M. A., Guo, S., and Keshav, S. 2006. Low-Cost Communication for Rural Internet Kiosks Using Mechanical Backhaul. In *ACM MOBICOM*.

[36] Shirani-Mehr, H., Kashani, F. B., and Shahabi, C. 2012. Efficient Reachability Query Evaluation in Large Spatiotemporal Contact Datasets. In *PVLDB*, 5(9).

[37] Suna, L., Axhausena, K. W., Lee, D., and Huanga, X. 2013. Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences,* 110(34).

[38] Tan, C.C., Sheng, B. , Wang, H., and Li, Q. 2010. Microsearch: A search engine for embedded devices used in pervasive computing. ACM *Transactions on Embedded Computing Systems*, 9(4).

[39] Vo, H. T., Wang, S., Agrawal, D., Chen, G., and Ooi, B. C. 2012. LogBase: A scalable log-structured database system in the cloud. In *PVLDB*, 5(10).

[40] VOICE project. VOIce-based Community-cEntric mobile Services for social development (VOICE). ICT-2009.9.1. http://www.mvoices.eu/

[41] Wang, Y., Yang, W.S., and Wu, J. 2013. Analysis of a Hypercube-Based Social Feature Multipath Routing in Delay Tolerant Networks. IEEE *TPDS*, 24(9).