Reliable data collection in highly disconnected environments using mobile phones

Brian DeRenzi

Department of Computer Science & Engineering
University of Washington

Joint work with Yaw Anokwa, Tapan Parikh, and Gaetano Borriello

Data collection

- Four and a half billion people in the developing world
  - All need access to basic services like healthcare

- Offering and maintaining services requires information
  - Digitizing information has advantages, but how do we transport that data?
Data transmission

- Most rural locations have limited or no network coverage
  - In rural southern Tanzania, only 34% (n=134) of health facilities have mobile coverage¹

- Use data mules to transmit data
  - Mobile devices that ferry data across regions of little or no connectivity

¹ Lindi & Mtwara Health Facility Survey 2004 IPTi in southern Tanzania IHDRC
"Some details"

- Collection of medical data
  - Obtained through examinations and interviews
  - Often collected in isolated rural villages

- Difficulties
  - Expensive to collect (field workers)
  - High error rates (misinterpretations, incomplete data, incomprehensible notes, etc.)
  - High-latency to usefulness (delay in data entry)

Data Collection Days
Village to Village

Rural Tanzania
Data Collectors

Objectives

- Simplify data collection
  - Dynamic forms on mobile device
  - Easy to add metadata (pictures, audio, video)
  - IDs, biometrics, geography for identification
  - Low-latency collection

- Safeguard data
  - Automatic backup of data
  - Get it to servers as soon as possible
Outline

- Motivation
- System components
- New applications
- Current progress
- Next steps
- Acknowledgements

Definitions

- Nodes
  - Mobile devices out in the field
- Neighbors
  - Nodes are neighbors if their devices can talk to one another
  - One-hop neighbors
- Server/Safe store
  - A centralized server where data is safe
- Source
  - A node which is sending, or has sent, data to a server
System components

Data Routing
- Use existing social networks to route data in disconnected environments

Location Profile
- Using GPS traces, build a profile of where the person has gone and where (s)he might go next
- Automatic exchange of location profile to determine good candidates to ferry data

Receipt acknowledgements
- Due to collection cost, do not delete data until it is been marked as received
- Path back to source will not necessarily be the same

Storage reclamation
- Data furthest from the source is deleted first
- Devices with data one hop away from source explicitly notify if there is a need to remove the data
- Source notifies those one hop away when a receipt is received
Routing to the server

- Forward to neighbors
  - Increment hop count
- 1-hop neighbors are special
  - Most likely to be of same organization
  - Backup data for each other
- Multi-hop neighbors are opportunistic best effort
  - Preference for large mules
  - Busses, trucks, government vehicles

From field worker to server

- Not all paths will be successful

![Network diagram with nodes A, B, C, D, E, F, G, H, I, J, and edges connecting them. The diagram includes symbols for disconnected, connected nodes, and internet server, with potential paths and successful paths indicated.]
Routing from server

- New data and acks of delivered data
- Acks
  - Mule that delivers gets ack
  - But also sent along other routes
- Data route determined by location profiles
  - Collected by devices GPS unit
  - Routed to mules likely to be heading where the packet has to go (may not be same place as teams move, multiple possible geographic destinations)
  - Also based on social links

From server to field worker (node)

- Nodes are mobile and may have moved, so the path will not necessarily be the same.
Safeguarding Data

- Share data among members of medical team
  - Groups of field workers travel together
  - Backup each others’ collected data
- Leverage social/work movements to move data
  - Each field worker can “mule” data to where there is connectivity
  - Exploit encounters with other NGO workers, truck drivers, government workers, etc.

Other considerations

- Data Prioritization
  - Send only core data first
  - Send smallest data first
- Usage Incentives
  - Could provide airtime for those transferring data
- User Privacy
  - PKI to protect patient data being transmitted
  - Variable granularity for location profiles
Outline

- Motivation
- System components
- **New applications**
- Current progress
- Next steps
- Acknowledgements

Applications

- Data collection for surveys in rural villages
  - Public health maternal and neonatal survey for the IPTi project\(^1\)
  - Child morbidity monitoring project\(^1\)
  - Stillbirth and miscarriage survey\(^2\)
  - Accompagnateurs in PIH
- Drug monitoring and patient data
  - Dispensaries refer to health clinics that refer to district hospitals
  - Data from medical protocols on mobile devices

---

\(^1\) Ifakara Health Research and Development Centre
\(^2\) Rachel Haws, Johns Hopkins University
**Medical Protocols on mobile devices**

- Protocols exist to aid clinicians
- Currently delivered on paper-based chartbooks.
- Using digital devices has a number of advantages
  - Patient record
  - Updatable protocols
- Need to get data back to policy makers

---

**Outline**

- Motivation
- System components
- New applications
- Current progress
- Next steps
- Acknowledgements
Regions of Operation

- Summer 2007
  - Brian DeRenzi with D-Tree in Mtwara, Tanzania
  - Yaw Anokwa with PIH in Rwinkavu, Rwanda

- Summer 2006
  - Brian was in Kilimanjaro and Tanga regions in Tanzania

Status

- Composed pieces on Nokia phones
  - First version of dynamic forms for data
  - GPS data collection (separate unit)
  - Message passing from device to device

- Prototyped simple email application
  - Computer to phone via MSR’s smstoolkit
  - Phone to phone via Bluetooth
Other Applications

Guatemala
- Remote village with cell phone connection only at top of nearby mountain
- Local entrepreneur shuttles messages for villagers (on paper)
- Could support voice, SMS, and e-mail by muling from villagers’ own phones
- Better privacy, shared use of the only phone in the village

Outline

- Motivation
- System components
- New applications
- Current progress
- Next steps
- Acknowledgements
Next steps

- Development of the system over the next year
  - Add more automation to current prototypes

- Deployment and field testing next summer and the following year
  - Explore the open questions about requirements

Acknowledgments

- Thank you
  - SIGCOMM/NSDR
  - International Service Learning (ISL)
  - Ifakara Health Research & Development Centre (IHRDC)

- Questions/comments?