Motivation and Design of a Content Distribution Architecture for Rural Areas

Aaditeshwar Seth
Department of Computer Science, IIT Delhi

Joint work with A. Mahla, D. Martin, I. Ahuja, and Q. Niyaz
Social media in rural areas

- Variety of mechanisms
  - Community radio
  - Community video
  - News-over-phone
  - ...

- Development ==
  - Empowerment
  - Accountability
  - Contextual info
  - Inclusiveness
  - Bonding
Trends: People want to express themselves

- Growing initiatives
  - Video Volunteers: 50+ community correspondents, heading to one per district; 1 video per CC per month
  - Digital Green: 2000+ videos in the last 3 years across 5 states
  - Community radio: 120+ stations, 5GB of new audio per month
  - News over phone: 1000+ listener calls per day from 3 districts

- Dropping costs of technology
  - $35 Aakash tablet
  - Smartphones and feature phones

[Chart showing Jharkhand Mobile News usage]
A voice/video social media platform for the BoP

Scaled up distribution and participation from end-users

- Business model to subsidize cost of access
- Ecosystem to translate social media to development
- Application design to access and participate on content
- **Content distribution under poor Internet connectivity**
Common use-case

- **On-demand** access of published **static content** rather than transactional applications
  - Publish-subscribe, browsing, content-download

- Digital Green comes closest: Participatory video production followed by offline screenings

- Dataset on video production and screenings
  - Jan 2009 to March 2011
  - 1000+ videos across 5 states/20 districts/50 blocks
  - Screened over 20,000 times
Typical social media patterns

- Large variations in production and consumption rates across different sites
- Power-law with cutoff indicating editorial influence in viewing content
Spatial and temporal locality

- 30% of content viewed within same block, 80% within same/neighboring districts
- Seasonal correlation is expected
- 60% of screenings happen within 3 months of production
Design principles

- Assuming other social media content will follow similar patterns

<table>
<thead>
<tr>
<th>Observations</th>
<th>Design principles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application use-cases: Publish-subscribe, browsing, content download</td>
<td>Content-based network</td>
</tr>
<tr>
<td>Locality in content production and consumption. Metadata can potentially reveal access patterns</td>
<td>• Content objects are first class entities</td>
</tr>
<tr>
<td></td>
<td>• Routers can cache content, examine metadata, pre-fetch and replicate</td>
</tr>
<tr>
<td>Content transfer capabilities to/from local rendezvous points in villages</td>
<td></td>
</tr>
<tr>
<td>Applications are tolerant of delays</td>
<td>• Delay tolerant data transfer</td>
</tr>
<tr>
<td>2G coverage is ubiquitously available although not sufficient for large content transfers</td>
<td>• Separate always-on content channel on 2G for route initializations and content download/upload requests</td>
</tr>
</tbody>
</table>
A content distribution network for rural areas

- Broadband Internet access available at district headquarters
- Block, village end-points, mobiles only have 2G connectivity
- Data distribution done via USB keys, SD cards, DVDs
Related work

- Mechanical backhaul
  - Daknet [Pentland et al], KioskNet [Seth et al], PostmanNet [Wang et al]

- Online-offline synchronization systems
  - CoCo [Shah et al], OCMP [Seth et al]

- Content based networks
  - CCN [Jacobsen et al]
  - LIPSIN [Jokela et al]
Network stack

Client: Mobile device/media station/kiosk

Content pub/sub application
- Partial key/value cache (offline)

Data SAR
- polling

Scheduler
- push/pull

Control
- syncing

Gateway
- (Content-ID, Station-ID, time)

Internet

Replication optimization, routing and scheduling

Registration of content replica locations

Content lookup

Service
- Full key/value store

Content pub/sub application
- Download/upload request
Network stack

Registration of content replica locations

Gateway

Replication optimization, routing and scheduling
(Content-ID, Station-ID, time)

Content lookup

Content pub/sub application
Partial key/value cache (offline)

Client: Mobile device/media station/kiosk

Data SAR
Scheduler

polling

syncing

push/pull

Data SAR
Scheduler

DTN
USB

TCP
GPRS

DTN
USB

TCP
GPRS

Cached node

Internet

Content pub/sub application
Full key/value store

Service
Network stack

Client: Mobile device/media station/kiosk

Content pub/sub application
- Partial key/value cache (offline)

Data SAR
- DTN
- USB

Scheduler

Control
- push/pull
- syncing

Gateway
- Replication optimization, routing and scheduling
  - (Content-ID, Station-ID, time)

Internet
- Registration of content replica locations

Service
- Content pub/sub application
  - Full key/value store

Cache node

Data SAR
- DTN
- USB

Scheduler

Control
- TCP
- GPRS

Client: Mobile device/media station/kiosk

Content pub/sub application

Internet

Service

Gateway

Content lookup

Replication optimization, routing and scheduling

Client: Mobile device/media station/kiosk

Content pub/sub application
- Partial key/value cache (offline)
Simulation analysis

- Topology layout
  - Content production/consumption at villages
  - Caches at districts, blocks, some villages

- Movement schedules
  - Village-block by ad hoc means of transport. Once a day
  - Block-block and block-district by bus. Few times a day

- Algorithms
  - Unicast with caching, multicast, multicast with pre-fetching, optimal multicast
  - Cache replacement: LRU, seasonal preference
Download requirements at gateway

<table>
<thead>
<tr>
<th>District</th>
<th>GB/month w/o cache</th>
<th>GB/month with 1GB cache</th>
<th>Improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dewas</td>
<td>8.73</td>
<td>1.13 ± 0.05</td>
<td>87.07</td>
</tr>
<tr>
<td>Rajgarh</td>
<td>7.94</td>
<td>4.70 ± 0.12</td>
<td>40.82</td>
</tr>
<tr>
<td>Guna</td>
<td>8.51</td>
<td>3.79 ± 0.18</td>
<td>55.44</td>
</tr>
<tr>
<td>Keonjhar</td>
<td>20.07</td>
<td>3.71 ± 0.3</td>
<td>81.50</td>
</tr>
<tr>
<td>Dindori</td>
<td>17.97</td>
<td>2.59 ± 0.18</td>
<td>85.59</td>
</tr>
<tr>
<td>Barwani</td>
<td>2.42</td>
<td>0.25 ± 0.01</td>
<td>89.73</td>
</tr>
</tbody>
</table>

- Cache size: Monthly download volumes ~ 1:10 can provide up to 80% improvements in download requirements
Effect of network topology on latencies

- Mesh topologies benefit more from village caches
- Temporal locality of access around blocks helps keep cache size requirements small
Room for optimization does exist

- Pre-fetching to high-degree nodes did not help much
  - Nodes often already included in the multicast trees
  - Well-connected locations are also more active participants

- Seasonal preference did not help much either
  - 60% screenings happen within 3 months of production, only few videos have a longer lifetime

- Distance based cache replacement may help

- More details in the paper
Conclusions

- Significant need for content distribution and participation in rural areas
- Content based architectures and caching are ideally suited in scenarios with spatial and temporal locality
- Control and data separation helps get around several DTN routing knowledge problems

Future work
- Different realizations: IVRs to request, SMSes to notify, etc
- Mobile application design for voice/video-based participation
- Bluetooth and WiFi-direct P2P content swapping on devices
But... content access is not free

- Somebody is paying!
  - Cellular data, toll-free phone calls, public screenings, radio broadcast, mechanical backhaul...
  - Moderation, curation, content quality

- Discovery of appropriate price points with relevant stakeholders to subsidize content access costs

- Need to fight a double battle!

<table>
<thead>
<tr>
<th>Stakeholders</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Communities</td>
<td>Have no local media, want news and information, want to express themselves</td>
</tr>
<tr>
<td>Advertisers, classifieds, platform customers</td>
<td>Access to rural consumers</td>
</tr>
<tr>
<td>Telecom companies</td>
<td>Increase ARPU from rural consumers</td>
</tr>
<tr>
<td>NGOs, governments</td>
<td>Better tools to improve development</td>
</tr>
</tbody>
</table>

Local entrepreneurs and livelihood opportunities
Acknowledgements: Amit Dubey, Sidharath Brahma, Gaurav Luthra, Gram Vaani Community Media, and Digital Green

Supported by: The Department of Information Technology, India and Intel India

Thanks for listening!

aseth@cse.iitd.ernet.in
Voice and video user generated content

- Local news | Opinion | Entertainment | Announcements
- Health | Education | Agriculture | Local governance

Incentives for community reporters

Livelihoods for local entrepreneurs

- Ads | Classifieds | User subscriptions | Content distribution
Community media and development

Role of intermediaries, identity of the media, text-free interfaces, low access cost, in-person community interaction

How can we make this more efficient? more scalable? more accessible?
Community radio

- CR 2.0: Radio + mobile phones
  - Poor telephony integration
- Media players used for scheduling
- Content stored in folders across different machines
- Many moving parts: Headphones, mics, mixer adjustments
Multiple use-cases for radio + telephony

- Community engagement 2x
- Live Panchayat group discussions
- Antakshari competitions across schools
- NREGA helpline, agriculture helpline…
  - Suggestion of new works
  - Project announcements and vacancies
- Stratified database of community members
  - Group updates
  - Personalization
  - Assessment of campaign effectiveness through micro-surveys
Complicated social systems...

[Bailur, 2007]
- “My wife had some ideas for programs. She went to the station. But they told her she had to belong to a self-help-group.”
- “… and it is not every group, but only good groups – meaning, you behave well, keep good books…”

[Krishna, 2002]
- Location of station effects inclusiveness and representativeness
Ideas and awareness for creating relevant programs

Topic of the month
- Employment
- Right to Food
- Water and sanitation
- Maternal and child health

Produce impactful programs
- Civic activism
- Political change
Social networking and content sharing
News-over-phone service: Citizen journalism for the BoP

Citizens call into a toll-free number to leave news

Validated news is published online

Network of social workers assist in grievance redressal

Other citizens can call into the toll-free number to listen to the validated news

Validated news is also published for access on same toll-free number

One stop view to performance of government services in rural areas
Impact can happen if a synergistic ecosystem can be created

- Gram Swaraj Abhiyan, NREGA Sahayata Kendra, Prabhat Khabar, Gram Vaani, etc help follow-up on issues
  - Bribery during UID enrollment: BDO visited the site and the officials were fined
  - Non-payment of wages: Immediate remuneration after the citizen report was relayed to the DDC
  - Malaria deaths in village: Mobile ambulance was dispatched, with fumigation equipment and first aid
- Mass media, government connections, strong NGOs, transparency… somehow synergize in Jharkhand
Community driven complaint management system

- Slum areas of Seemapuri and Sundernagari, East Delhi. In partnership with Action India

Community members call into a toll-free number to leave complaints. NGO listens to complaints and categories/transcribes them on the Internet. Youth wing of NGO files complaint officially, and sends a report to the local councilor. Redressal statistics also displayed on website (and wall newspapers).
Internal dynamics

Key points of leverage: Community empowerment, transparency of grievance statistics, political accountability.
v1: Complaint management system for MCD

Few complaints and statistics posted on MCD’s Facebook page
v2: Citizen based monitoring of garbage collection sites – [http://mericity.in](http://mericity.in)

**Gram Vaani**

**Dhalaos** | **Issues** | **Report Archives**
---|---|---

**General Information**
To dispute Dhalaos status:
Call: 9910153713
Email: modelward@gramvaani.org

<table>
<thead>
<tr>
<th>Dhalaos/WZ</th>
<th>Location</th>
<th>Color</th>
<th>Last Report Time</th>
<th>Comments</th>
<th>Disputed (Yes/No)</th>
<th>Citizen Reports</th>
<th>Cleaning Consistency</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>81/103/WZ</td>
<td>ROAD NO.41, W.P.BAGH</td>
<td>GREEN</td>
<td>Feb. 25, 2012, 2:09 a.m.</td>
<td>Reported as Green</td>
<td>No</td>
<td>2 reports open</td>
<td>92%</td>
<td>dispute</td>
</tr>
<tr>
<td>82/108/WZ</td>
<td>JHEEL PARK, MADIPUR</td>
<td>GREEN</td>
<td>Feb. 25, 2012, 2:10 a.m.</td>
<td>Reported as Green</td>
<td>No</td>
<td>1 reports open</td>
<td>96%</td>
<td>dispute</td>
</tr>
<tr>
<td>83/103/WZ</td>
<td>CENTRAL BANK MADIPUR</td>
<td>GREEN</td>
<td>Feb. 25, 2012, 2:10 a.m.</td>
<td>Reported as Green</td>
<td>No</td>
<td>1 reports open</td>
<td>94%</td>
<td>dispute</td>
</tr>
<tr>
<td>85/103/WZ</td>
<td>A-BLOCK, MADIPUR</td>
<td>GREEN</td>
<td>Feb. 25, 2012, 3:09 p.m.</td>
<td>Reported as Green</td>
<td>Yes</td>
<td>2 reports open</td>
<td>88%</td>
<td>dispute</td>
</tr>
<tr>
<td>NUC-3/103/WZ</td>
<td>ISKCON Temple</td>
<td>GREEN</td>
<td>Feb. 25, 2012, 2:10 a.m.</td>
<td>Reported as Green</td>
<td>No</td>
<td>1 reports open</td>
<td>94%</td>
<td>dispute</td>
</tr>
<tr>
<td>NUC-7/103/WZ</td>
<td>Rajiv Gandhi Camp</td>
<td>GREEN</td>
<td>Feb. 25, 2012, 2:10 a.m.</td>
<td>Reported as Green</td>
<td>No</td>
<td>No issues reported</td>
<td>100%</td>
<td>dispute</td>
</tr>
</tbody>
</table>

- Self reports from MCD and contractors
- Site related complaints
- Citizens can dispute reported stats
Community media and development

- Translating community media to development is a complex process
  - Appropriate technology and feature development
  - Cost
  - Training
  - Technical support
  - Ecosystem
  - Funding
- Simple voice and video technologies can go a long way
- Lots of scope for technological intervention