Computing for Development

A New High-Impact Research Area

Lakshminarayanan Subramanian
NYU

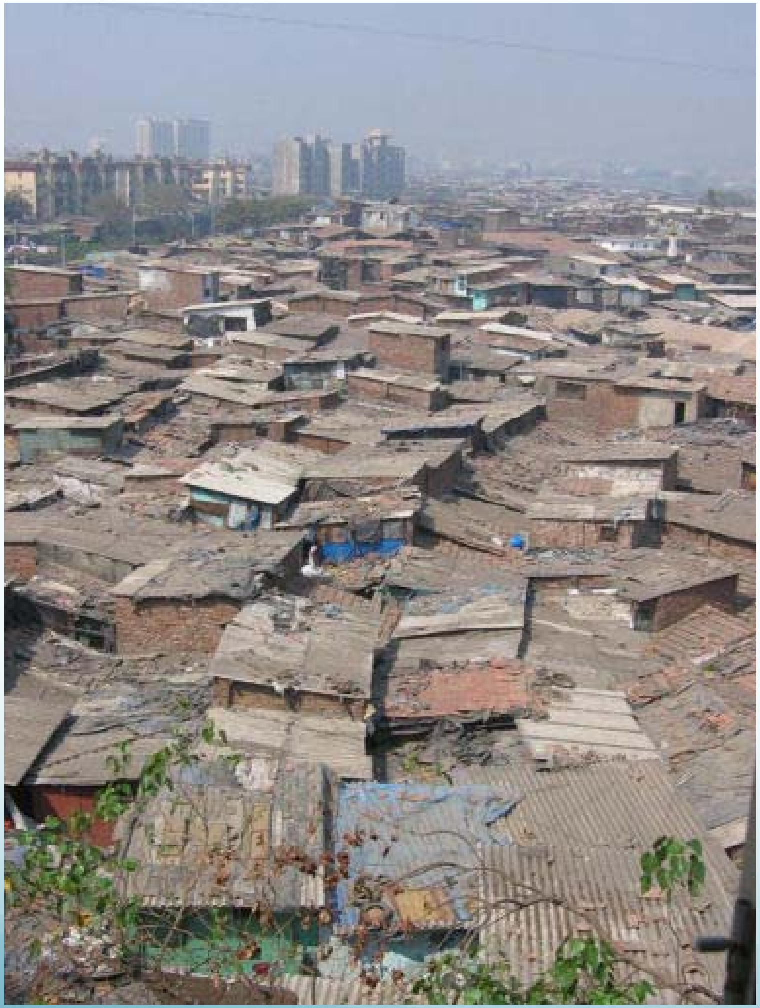
Joint work with many
CATER (NYU), NeWS(NYU), TIER(Berkeley)

Sustainable Development

- Sustainable Development Theories:
  - Jeffrey Sachs: End of Poverty
  - Bill Easterly: Elusive Quest for Growth
  - C.K. Prahlad: Fortune at the Bottom of the Pyramid
  - Amartya Sen: Development as Freedom
  - Paul Collier: The Bottom Billion
- Commonality: “Rural Empowerment critical to sustainable development”

“Appropriate Technology a potential enabling factor to empower rural markets”

The Untapped Rural Market



- Dharavi: Largest Slum in India
  - High cost of being Poor!
  - 85% have a TV
  - 50% have a pressure cooker
  - 21% have a telephone
  - … but can't afford a house
- In Bangladesh:
  - Poorest devote 7 percent income to communications (GrameenPhone)
- These are valid markets…

Enabling Rural Markets

- The Cellular Revolution
  - 70% own a phone/SIM in Africa
- Mpesa, Gcash – Mobile Microfinance
  - 1 million transaction/days in Kenya
- Aravind Telemedicine Network
  - Telemedicine services for 500,000 patients/year
- Digital Green + Digital Study Hall
  - Teaching Farmers and Students using Recorded Video
- eSoko
  - A popular mobile marketplace

Aravind Telemedicine Network

Computing for Development

- Focus: *Design, implementation and evaluation of new computing innovations that enable global social and economic development*
- First world technology - a bad fit!
- Hardest Challenge: Identifying the “right problem”
- Key requirements for technology adoption
  - Locally appropriate
  - Cost-effective
  - Easy to use
  - Extremely robust

The Hard Challenges!

- Need for Cost-effective solutions
  - Minimalistic Computing: Design with minimal resources
- Low-cost high-bandwidth connectivity
- Appropriate Design + Accessible Technologies
- Reliability + Sustainable Power
- The Language Barrier
- And many more….

Challenges encompass several areas of CS

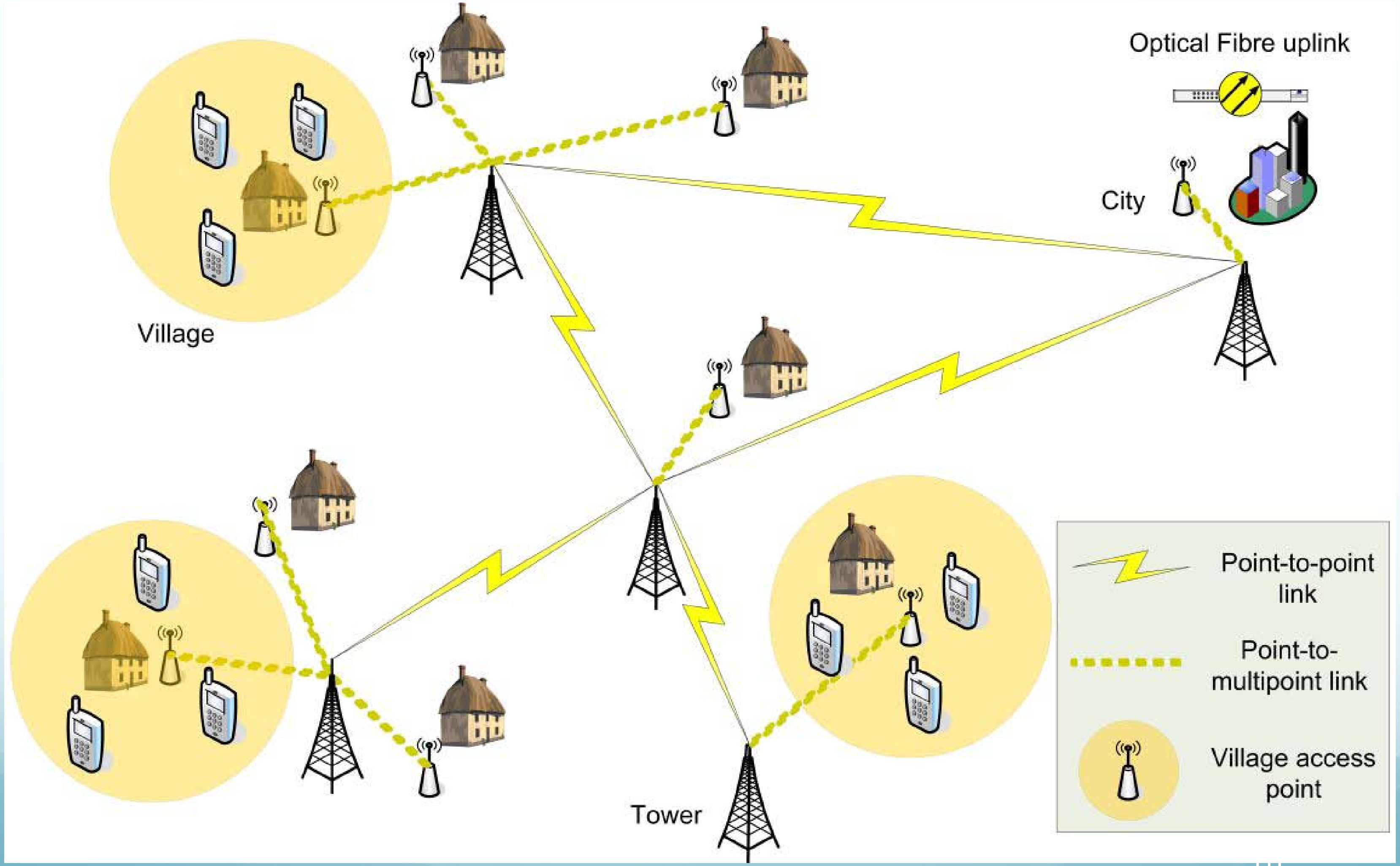
SIGDev

- Proposed new SIG, in “Computation for development”
- Areas:
  - Networks, Systems, Security
  - HCI and Applications
  - AI, NLP, Data mining, Speech, Vision
- Starts this year with DEV 2010
  - http://dev2010.news.cs.nyu.edu
  - December 17-18,2010

Rest of the talk

- Connectivity for the next billion
- Next generation mobile services
- Web architecture for developing regions

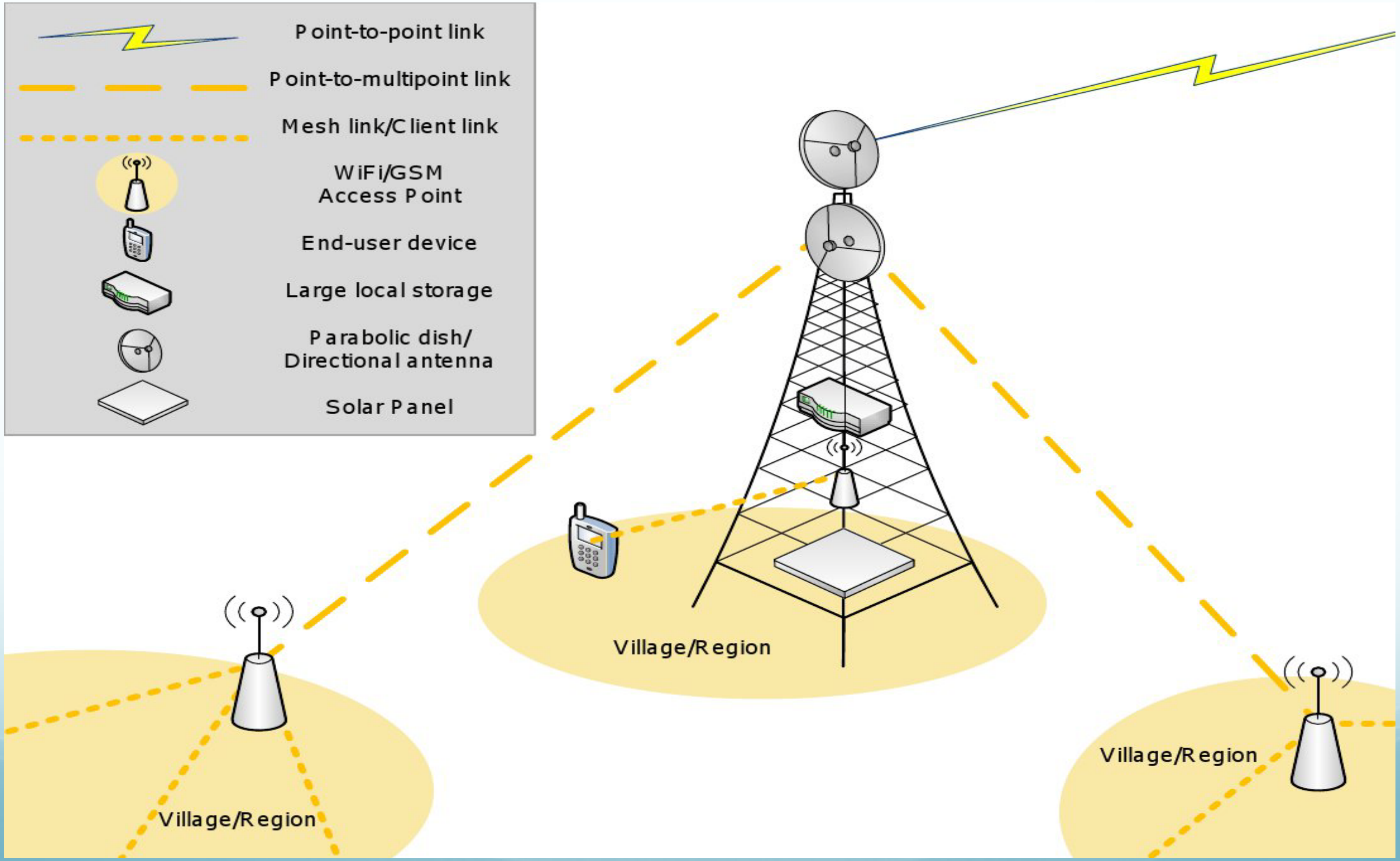
WiRE Architecture



The WiRE vision

- Extremely cheap focused connectivity
  - At least 10 Mbps connectivity
  - Voice calls < 0.1 cents/minute
- Every user owns a cheap mobile device
  - The go-to device for communications, information access and business transactions
- All devices are solar-powered
- Network management should be made easy
- Enable vibrant rural markets with mobile devices, cheap connectivity and next-generation mobile services

WiRE Node Architecture



Challenges

- Physical layer
  - Steerable antennas, better radios, 802.11n?
- MAC layer
  - Combinational wireless network challenges
- Network layer
  - Naming, addressing, routing
- Robustness
  - Power, maintenance
- Application layer
  - Security, End-to-end performance

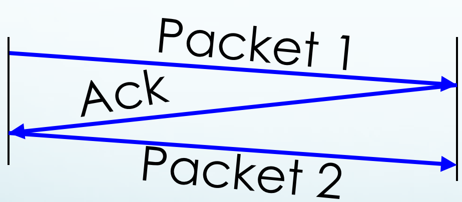
WiFi-based Long Distance Networks

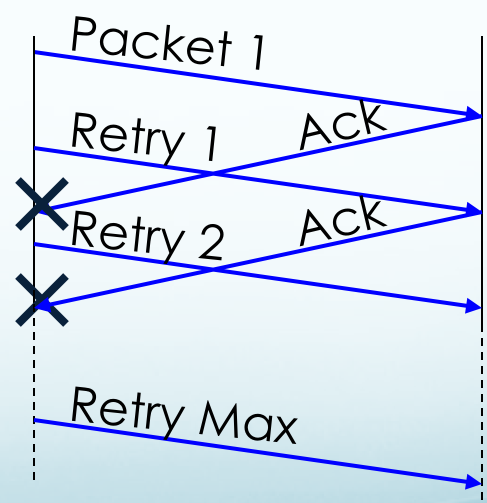
- **WiLD** links use ***standard 802.11*** radios
- Longer range up to **150km**
  - Directional antennas (24dBi)
  - Line of Sight (LOS)
- Why choose **WiFi**:
  - Low cost of $500/node
    - Volume manufacturing
  - No spectrum costs
  - Customizable using open-source drivers
  - Good datarates
    - 11Mbps (11b), 54Mbps (11g)


New World Record – 382 Kms
Pico El Aguila, Venezuela
Elev: 4200 meters

Problem with 802.11: ACKs

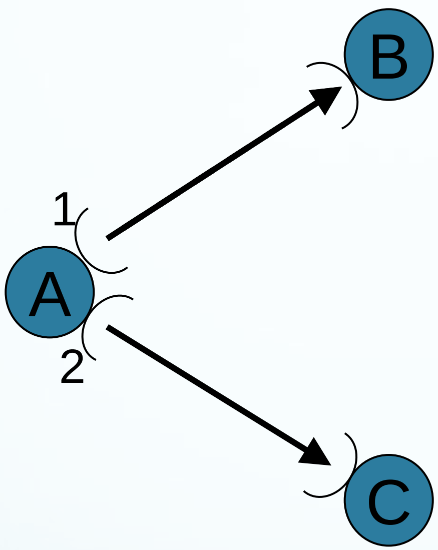
- Low utilization
  - Large propagation delays
  - Stop & wait inefficient
  - RTS/CTS makes it worse
- ACK timeouts
  - ACK doesn't arrive in time
  - Retransmissions until retry limit reached



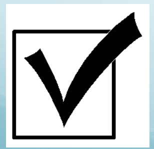


Inter-Link Interference

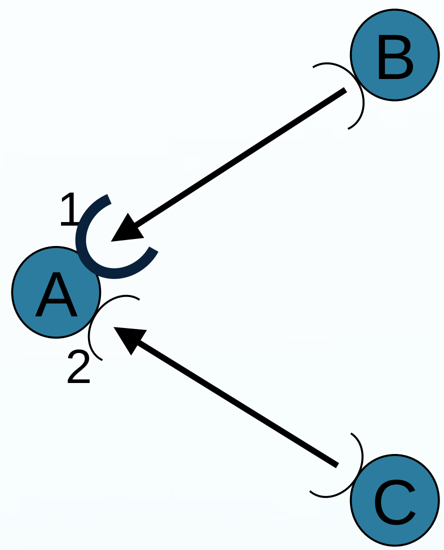
Simultaneous Send



- Disable CCA

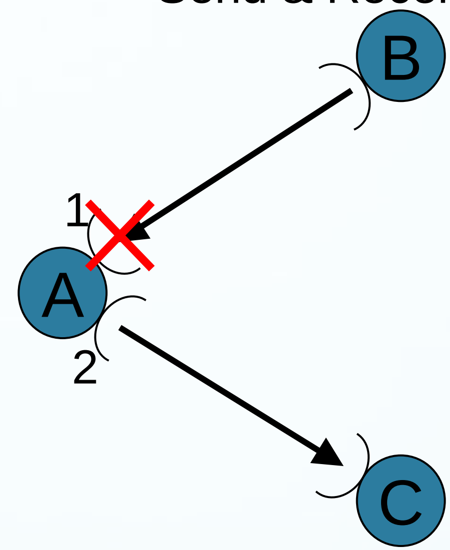


Simultaneous Receive



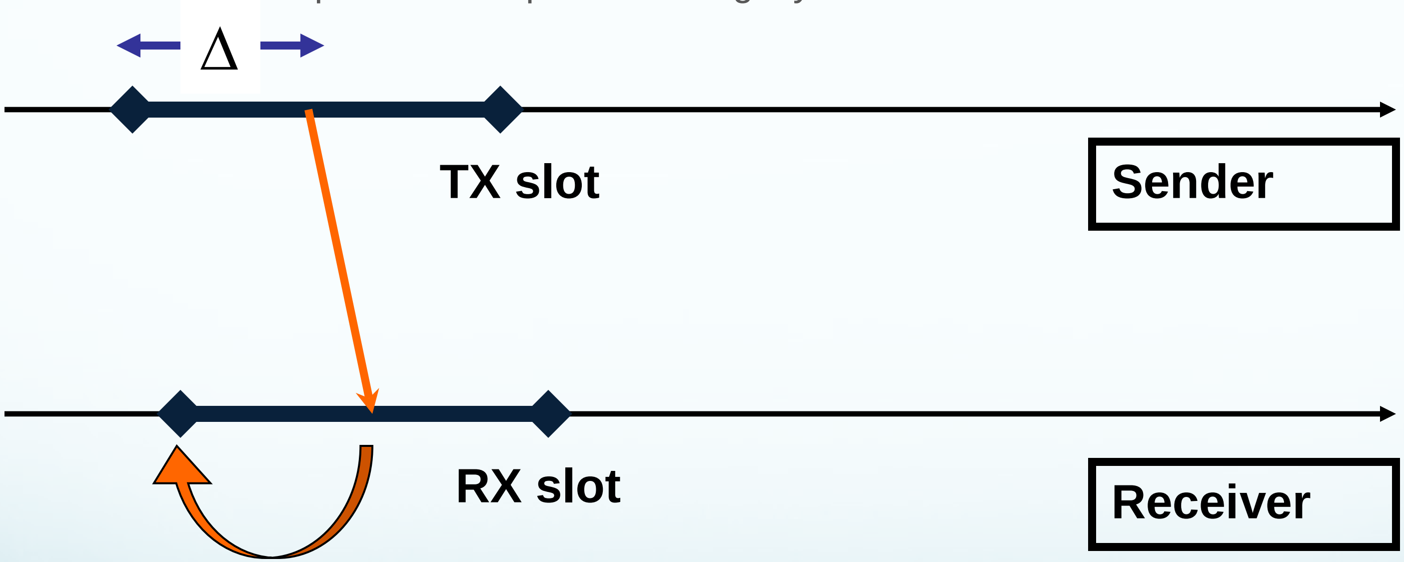
- 12dB isolation

Send & Receive



Implicit Synchronization for TDMA

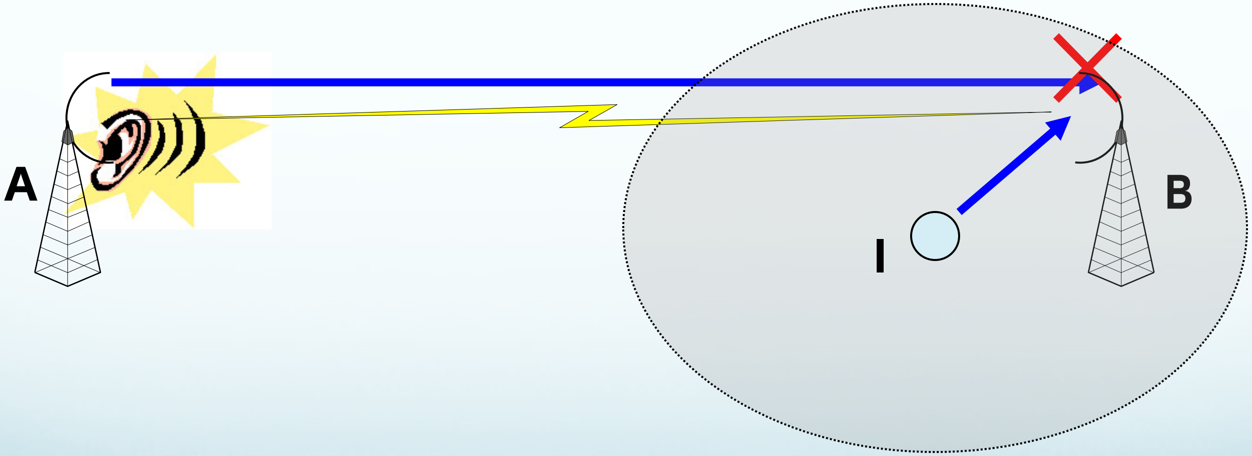
- Every packet is time-stamped in TX slot
- Slots are offset because of propagation delay
- We don't use explicit marker packets to signify end of TX slot*

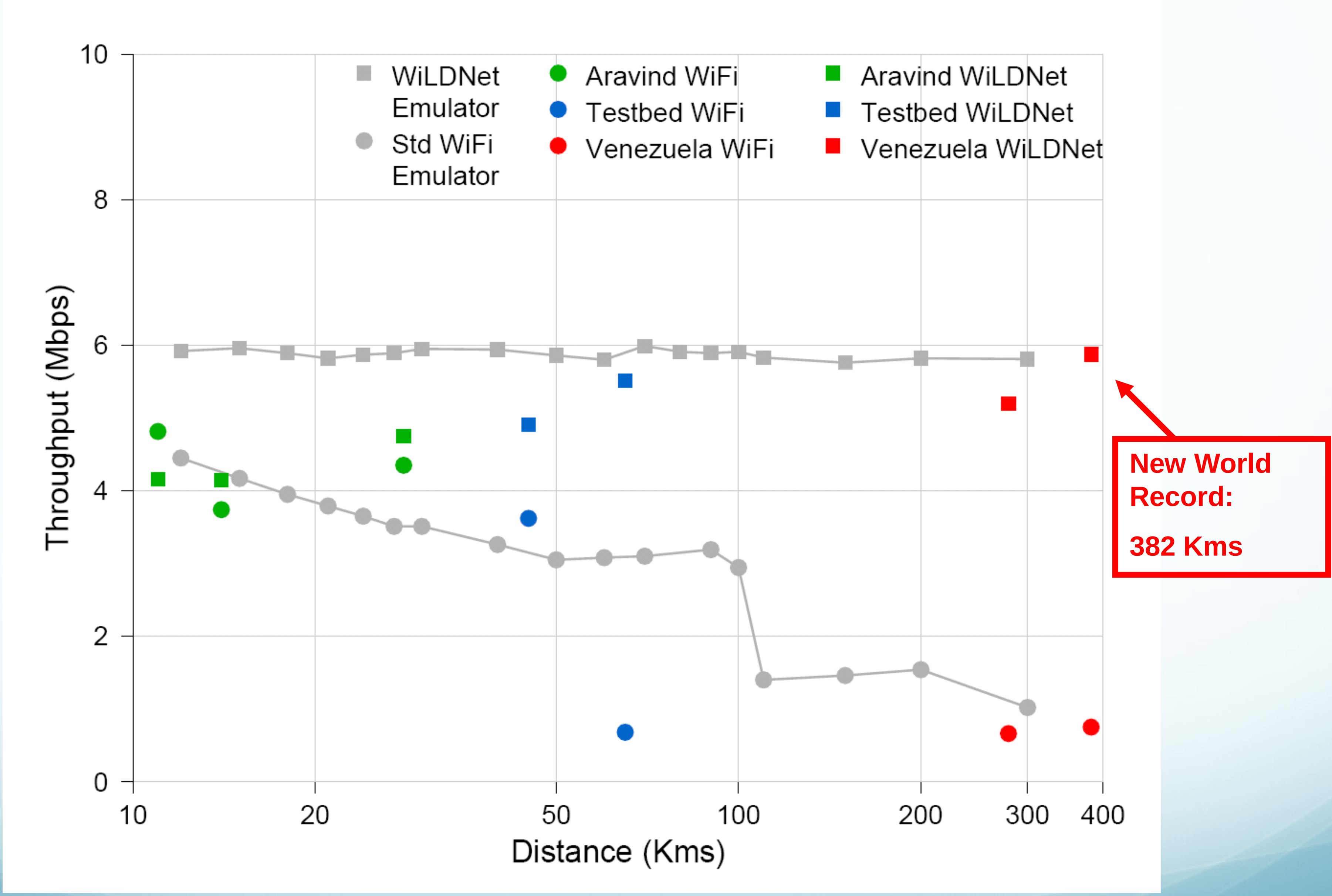


* 2P MAC protocol (Raman et al. Mobicom '05)

Channel Loss: From external traffic

- **Strong correlation between loss and external traffic**
- **Source (A) and interferer (I) do not hear each other**

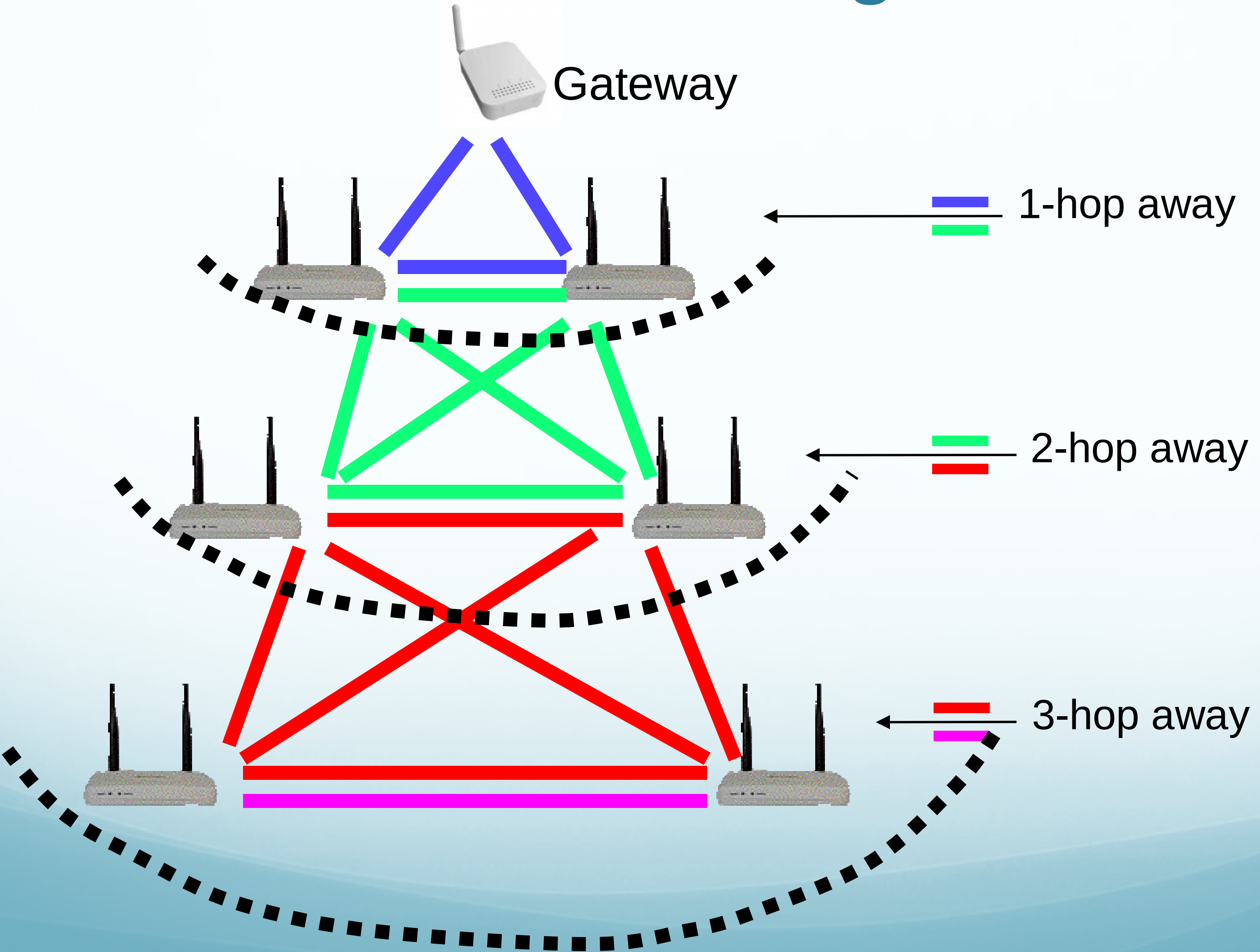




10
8
6
4
2
0
Throughput (Mbps)
10
20
50
100
200
300
400
Distance (Kms)
WiLDNet Emulator
Std WiFi Emulator
Aravind WiFi
Testbed WiFi
Venezuela WiFi
Aravind WiLDNet
Testbed WiLDNet
Venezuela WiLDNet
New World Record:
382 Kms

High performance Multi-radio mesh networks

Routing-driven Interference-free Channel Assignment



A Stable ETT metric

- ETT/ETX over-estimate link performance.
- Besides average loss, other factors affect performance:
  - Loss variations
  - External load
- ROMA's link metric:

$$ETT = \frac{1}{(p_a - p_v) * (p_a' - p_v')} \qquad L$$

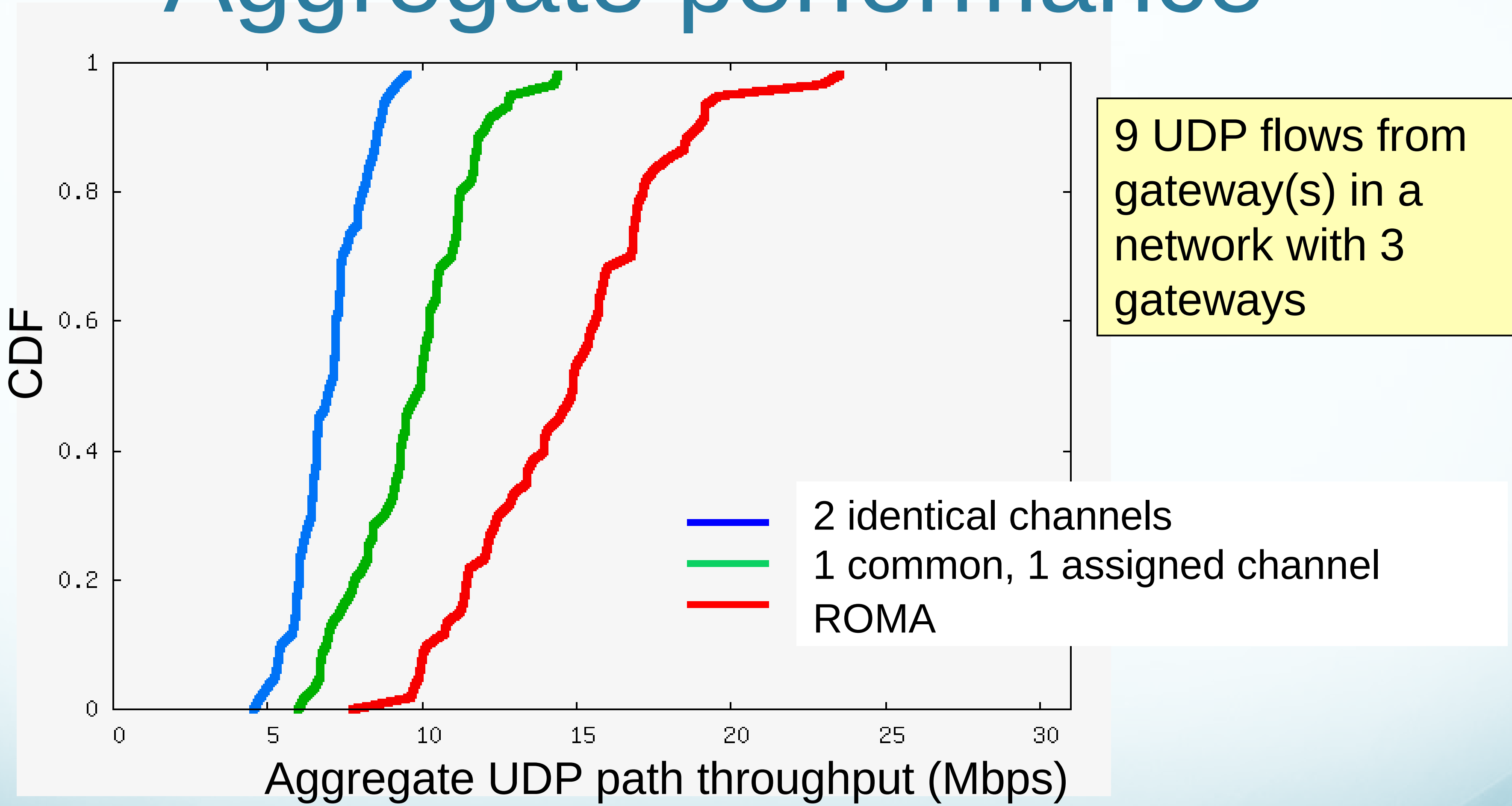
Robust Routing Metric

- SIM route metric [Das et al. NSDI'08] trades off performance and overhead
- Extend SIM to account for external load and variation

$$\left(0.2 * \underbrace{\sum ETT_i}_{\text{Capture tx overhead}} + 0.8 * \underbrace{\max(ETT_i)}_{\text{Capture bottleneck link(s) performance}}\right) *(1+L)$$

- Discover better routes through "investigation"

Aggregate performance

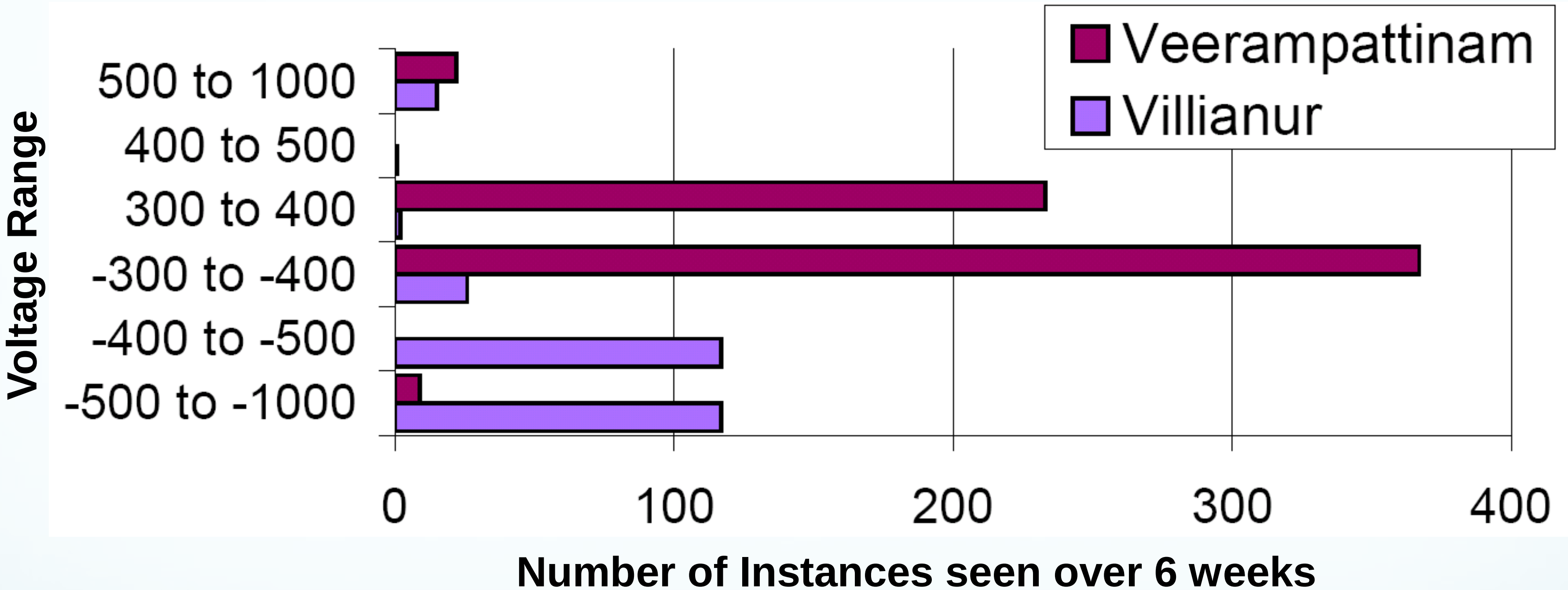


9 UDP flows from gateway(s) in a network with 3 gateways

- ROMA can utilize many available channels to improve aggregate throughput

Reliable Power

Poor Quality Power



Spikes and Swells:

- Lost 50 power adapters
- Burned 30 PoE ports

Low Voltages:

- Incomplete boots
- HW watchdog fails

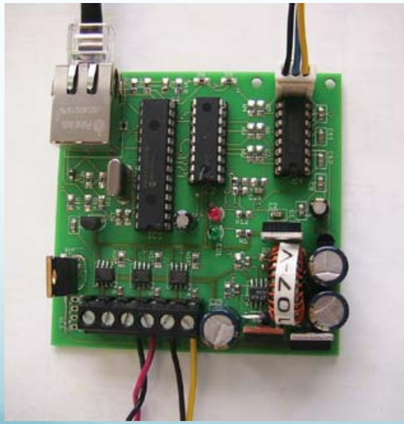
Frequent Fluctuations:

- CF corruptions
- Battery Damage

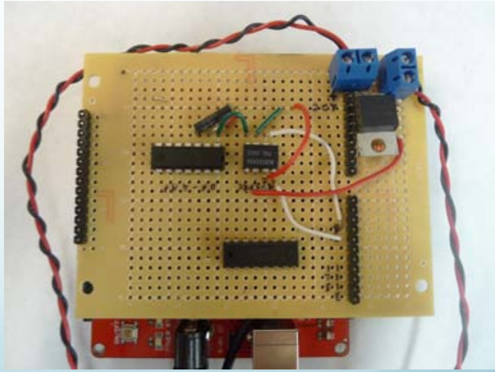
Reliable Solar Power

Installations in Ethiopia

Installations in Ethiopia

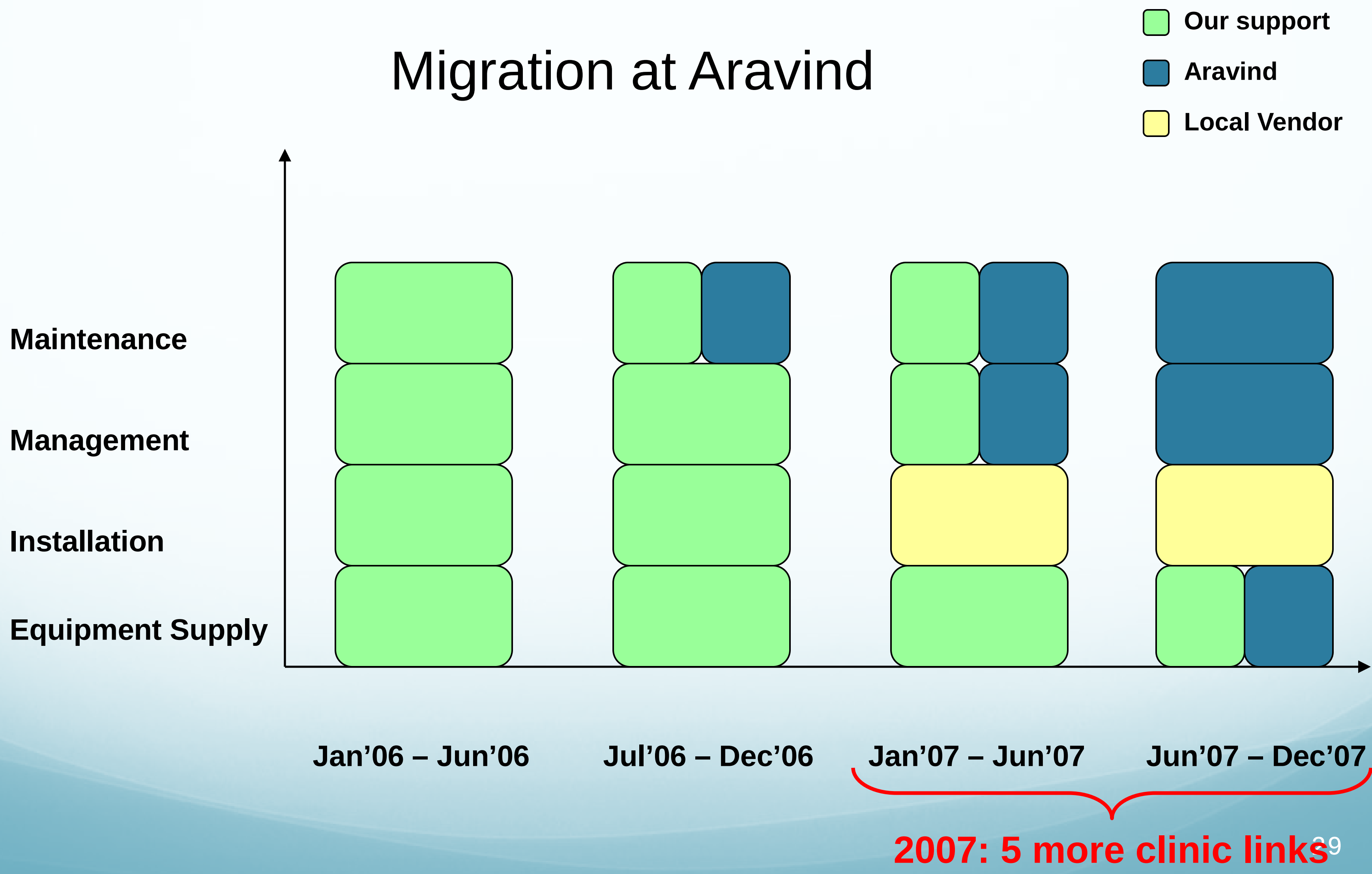


Low-cost Solar Power controller



Solar panel monitoring system

Operational Results



Rest of the talk

- Connectivity for the next billion
- Next generation mobile services
- Web architecture for developing regions

Need for SMS apps?

In many developing regions a data plan is not accessible

- No cellular data network
- Data plans are expensive
- Fancy phones are costly
- Deployable immediately

Ecosystem of SMS/Voice Services

World Health Organization

RICOH
RICOH INNOVATIONS

NEW YORK UNIVERSITY
SCHOOL OF MEDICINE

NIMH
National Institute of Mental Health

NATIONAL CANCER INSTITUTE®

HARVARD SCHOOL OF PUBLIC HEALTH

Elizabeth Glaser Pediatric AIDS Foundation

D-TREE INTERNATIONAL

CLINTON FOUNDATION

AED · SATELLIFE
Center for Health Information and Technology

Abt
Abt Associates Inc.

THE EARTH INSTITUTE
COLUMBIA UNIVERSITY

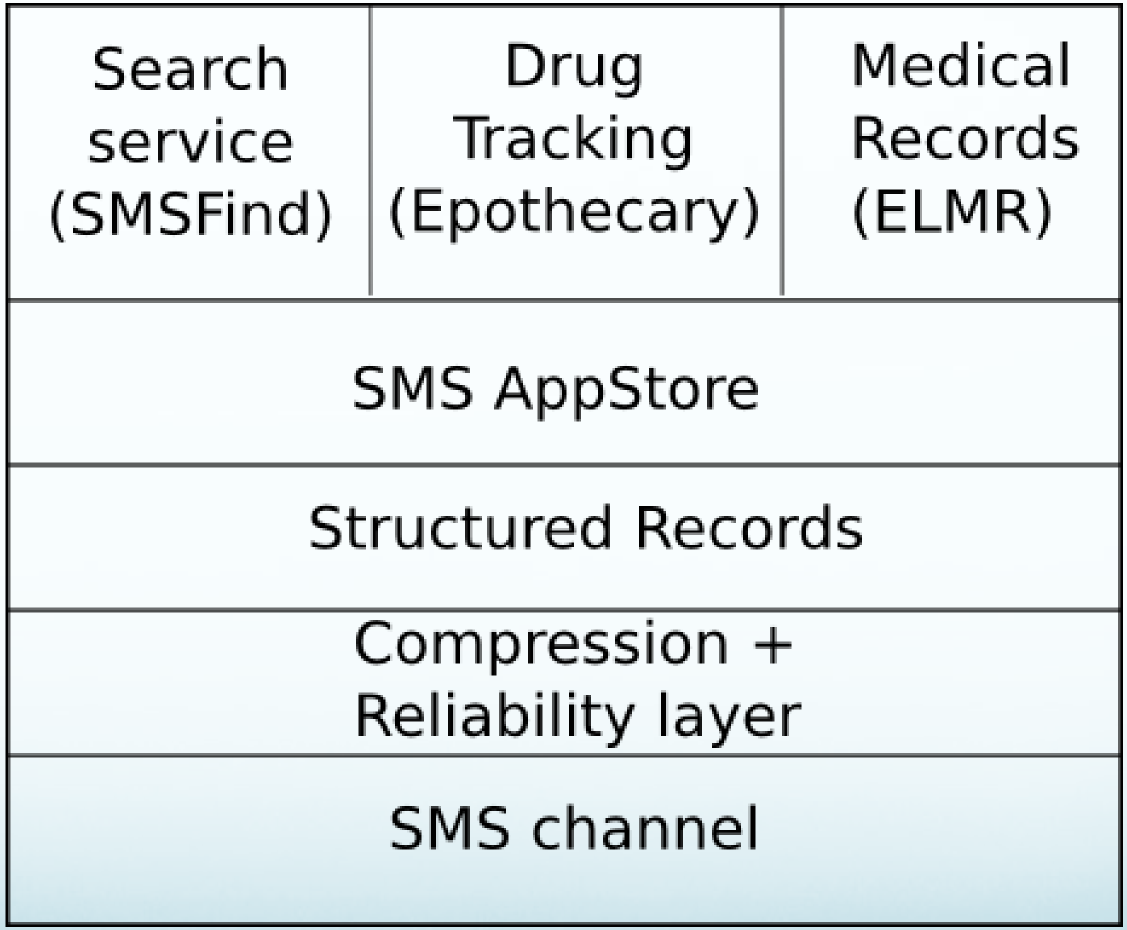
unicef

iavi
International AIDS Vaccine Initiative

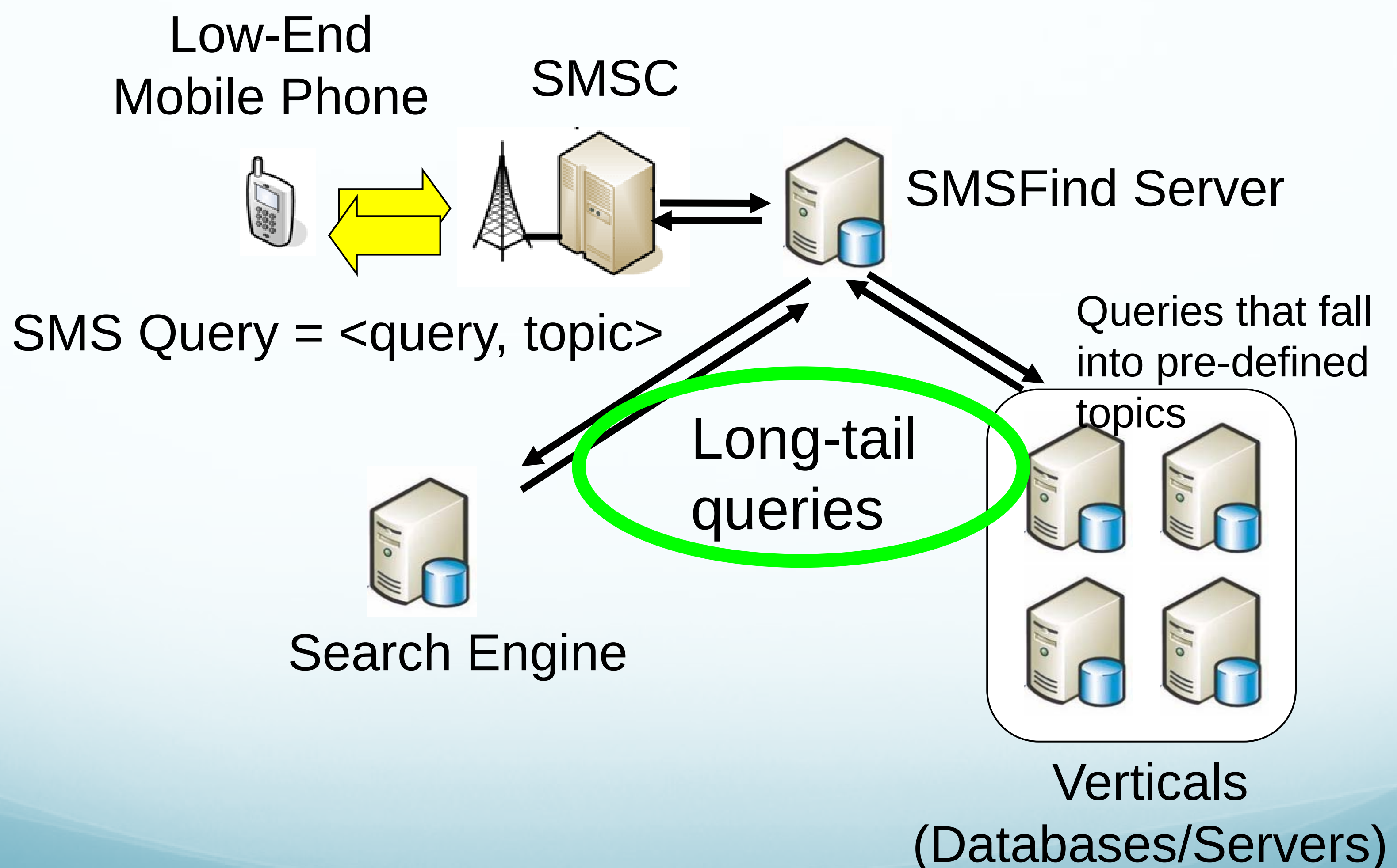
THE WORLD BANK
Working for a World Free of Poverty

UNAMESA

SMS stack


Search service (SMSFind)
Drug Tracking (Epothecary)
Medical Records (ELMR)
SMS AppStore
Structured Records
Compression + Reliability layer
SMS channel

SMSFind – SMS Search



Different from Q/A Systems

SMSFind	Q/A Systems
Unstructured queries	Structured queries
SMS/Mobile queries	Typically manually generated or search engine queries
Document corpus is a function of the query (dynamic, noisy)	Corpus is typically fixed and much smaller
Output is a 140 byte snippet answer	Output is a document, short answer, or summary

SMSFind Algorithm

- Problem: Given a <query, topic>, get all web search result pages, and look for the appropriate 140 byte answer (snippet)

- Intuition: Answer is somewhere in the search result pages, use the topic as a hint

- Algorithm Key Steps:
  - Extract candidate snippets, n-grams
  - Score and rank n-grams
  - Rank snippets using n-gram score

Main Result

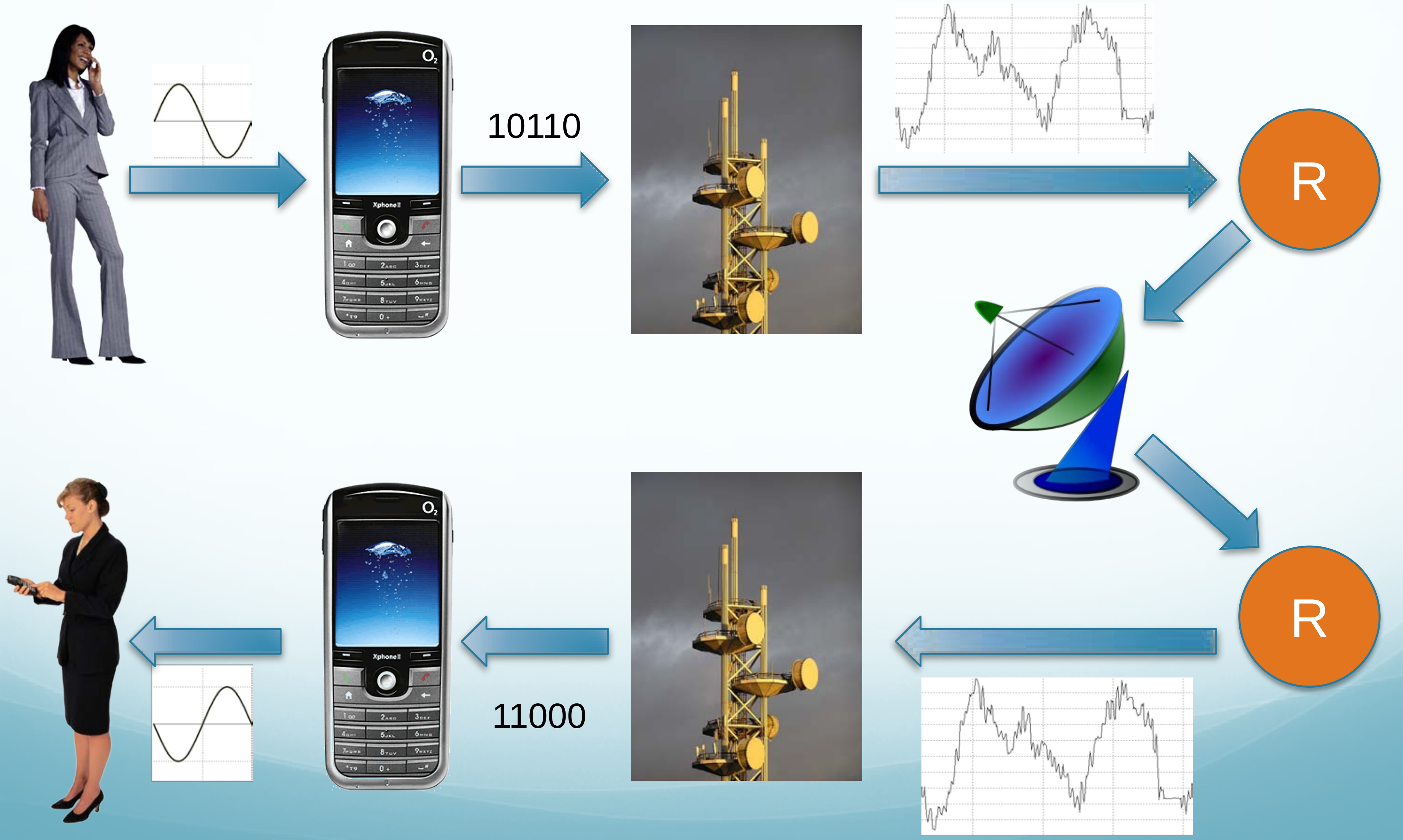
System, Input	% Correct
SMSFind (All queries)	**57.3%**
Google SMS (All Queries)	9.5%

Pilot covering 2000-3000 people in Nairobi, Kenya

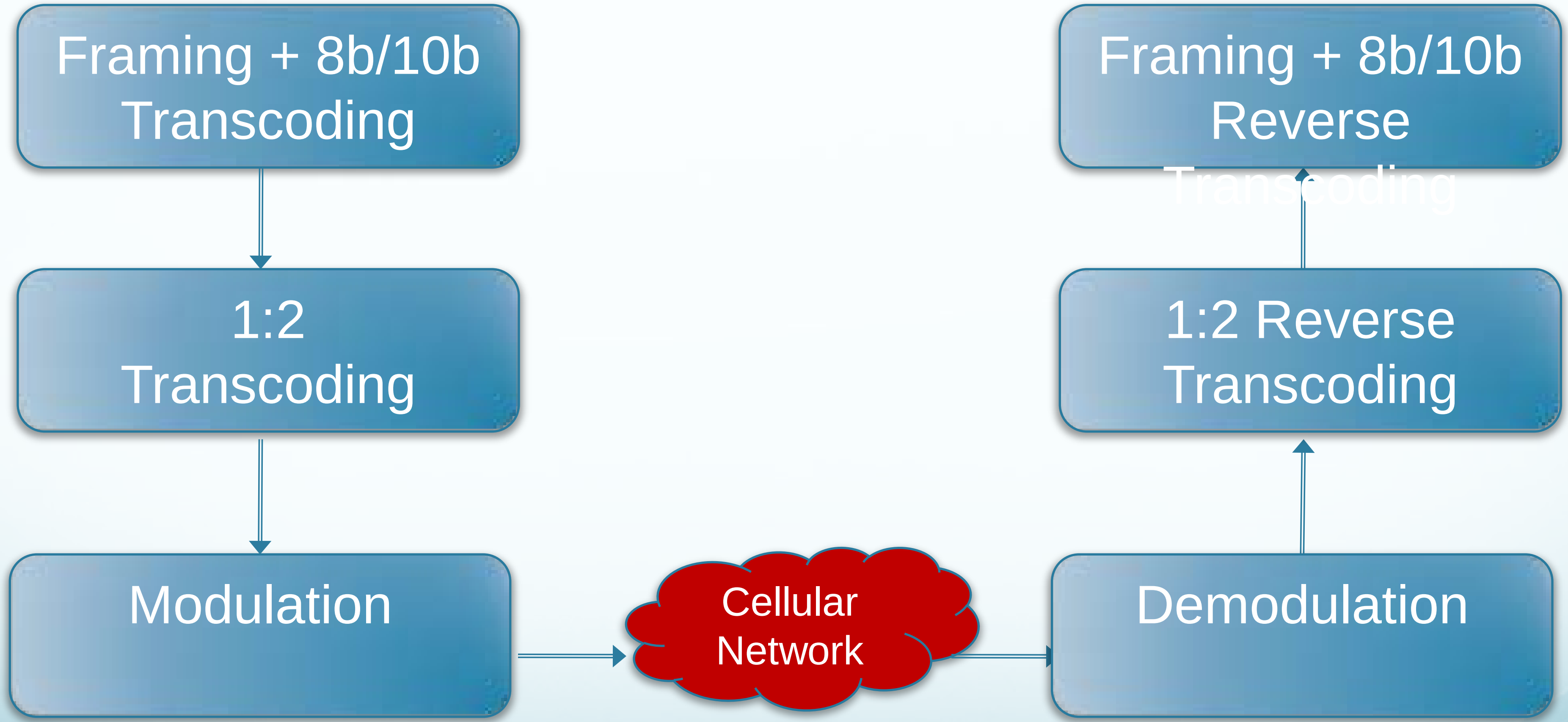
Hermes: Data over Voice Channels

- Scarce / expensive data connectivity
- Ubiquitous cellular connectivity
  - Voice and SMS services.
  - No data connectivity. Why?
  - Cost per bit for SMS is very high.
- Can we modulate data on sounds and send it over a voice call?
  - Functionally like a modem, perhaps?

Cellular Voice Channels



Hermes: Protocol Stack



All algorithms should be simple!

Modulation

Algorithm 1 Convert binary data to sound signals to be sent over a voice call.

Given: base frequency f_{base}, delta frequency δ.
$f = f_{base}$
for each bit b in the input string **do**
  if $b = 0$ **then**
    $f = f - \delta$
  else
    $f = f + \delta$
  end if
  Generate a sinusoid of frequency f
end for

Demodulation

Algorithm 2 Convert received sound signals back into binary data.

Given: Input sound signal.
for each sinusoid in the input sound signal **do**
  Let f_{curr} = frequency of current sinusoid
  Let f_{prev} = frequency of previous sinusoid
  if $f_{curr} \leq f_{prev}$ **then**
    Output 0
  else
    Output 1
  end if
end for

1:2 Transcoding

$$0 \rightarrow 01$$
$$1 \rightarrow 10$$

Input	0	1	1	0	0
Output	01	10	10	01	01

What does this give us?

- Fixed fundamental frequency (Voice-like)
- Operation within very narrow frequency ranges

1:2 Reverse Transcoding

$$01 \rightarrow 0$$
$$10 \rightarrow 1$$

Input	01	10	10	01	01
Output	0	1	1	0	0

- What about error detection?
  - Bit flips?
  - Insertions/deletions?

Performance: Raw Performance

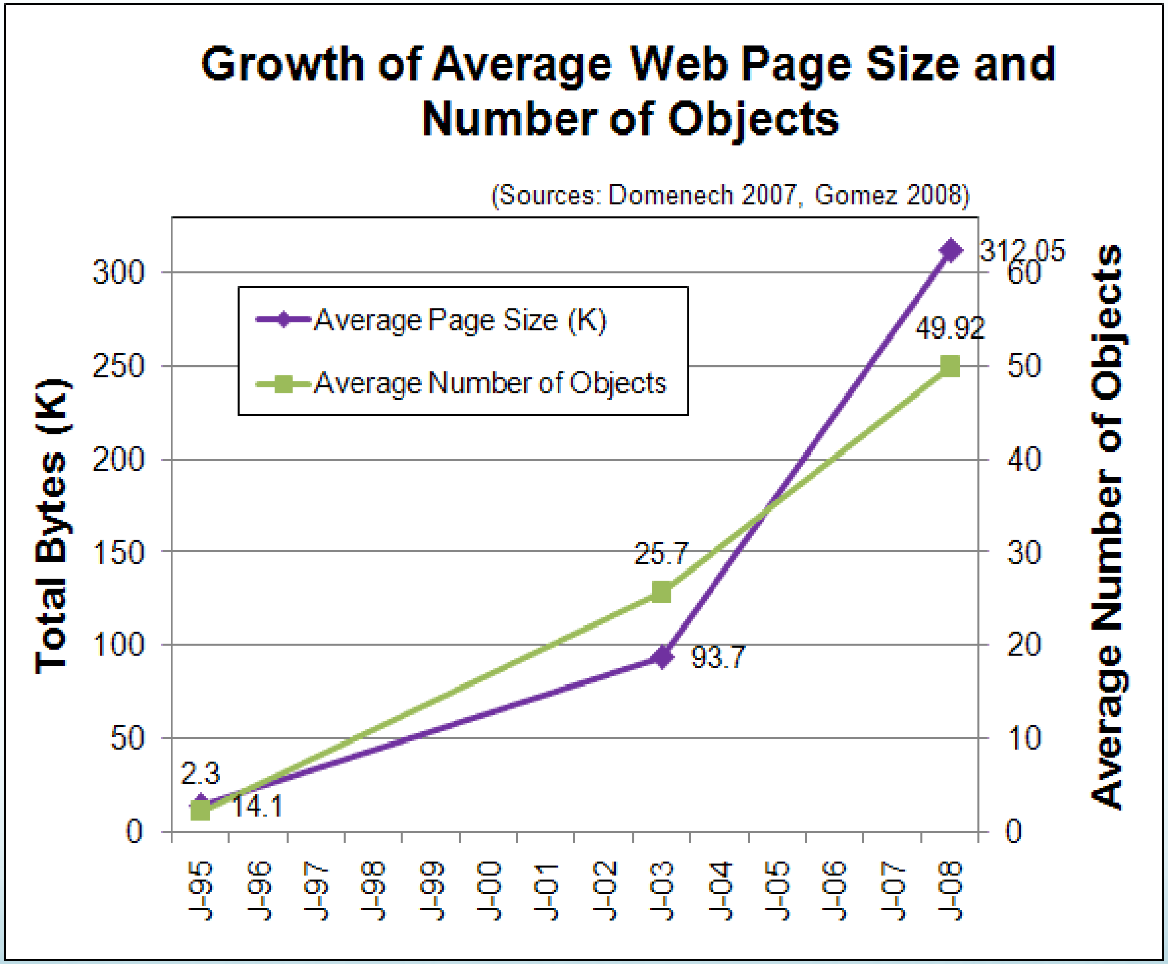
	f_{base}(Hz)	δ(Hz)	BER
AT&T $\rightarrow$ AT&T	2200	480	1×10^{-5}
T-Mobile $\rightarrow$ T-Mobile	2400	640	1×10^{-5}
AT&T $\rightarrow$ T-Mobile	2170	470	1×10^{-5}
T-Mobile $\rightarrow$ AT&T	2130	640	1×10^{-5}

Data Rate = f_{base} * 0.4

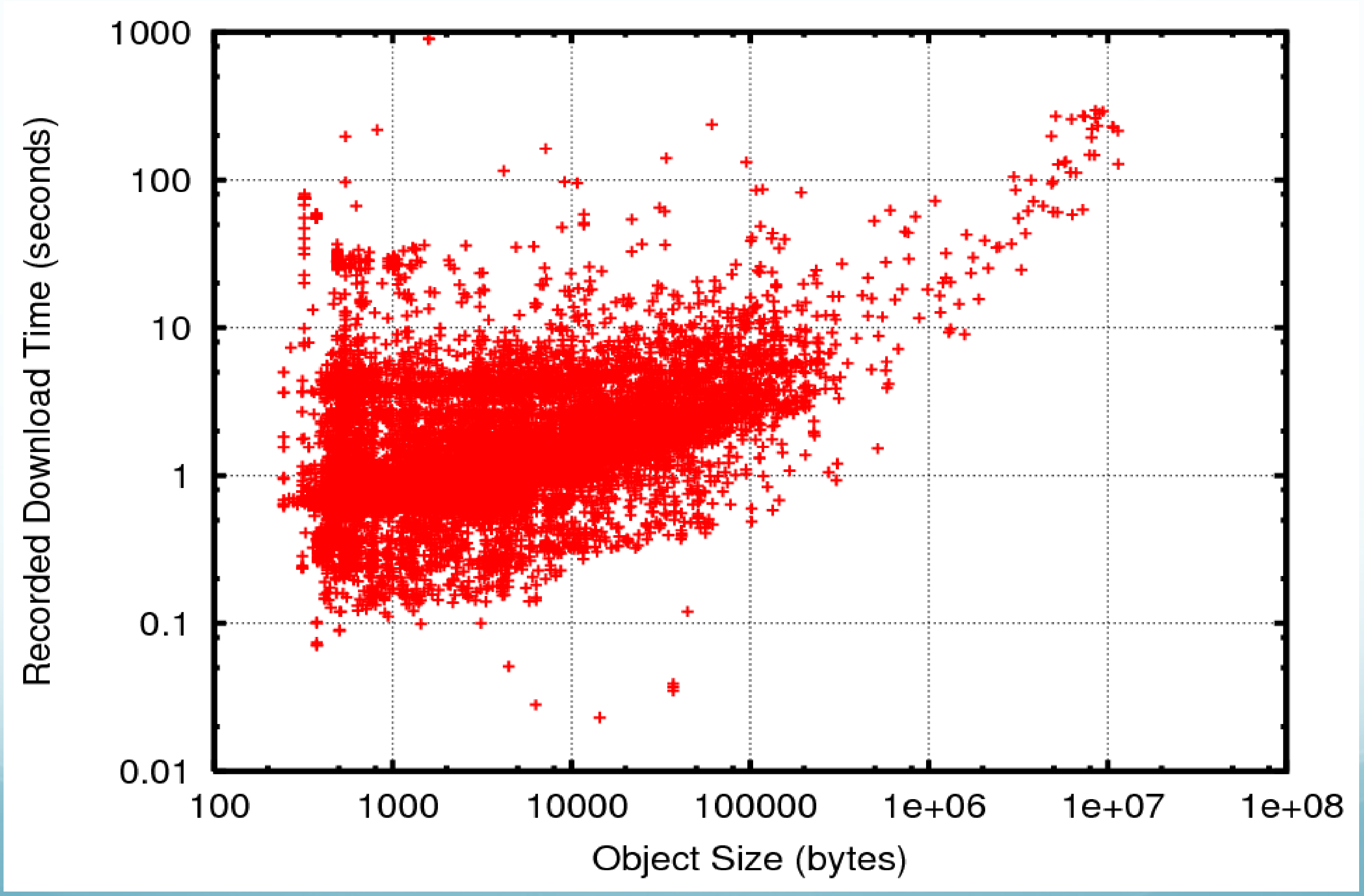
Rest of the talk

- Connectivity for the next billion
- Next generation mobile services
- Web architecture for developing regions

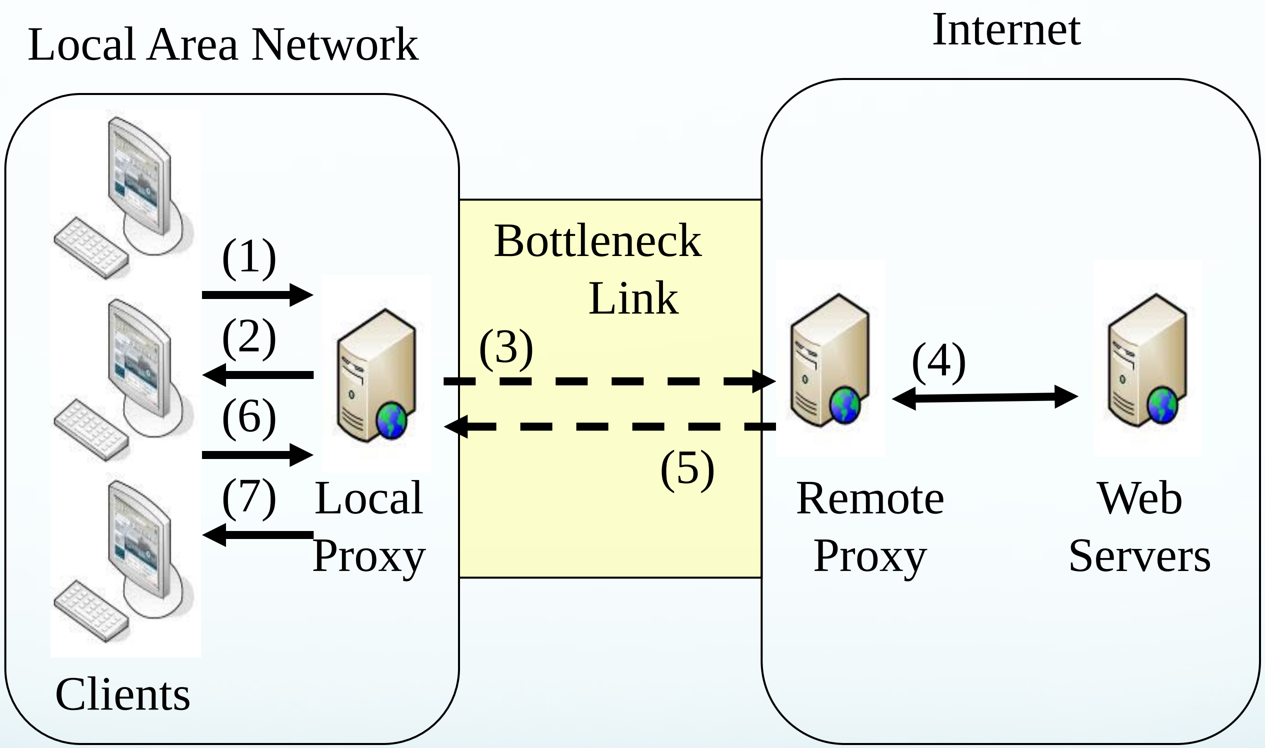
Web Page Size



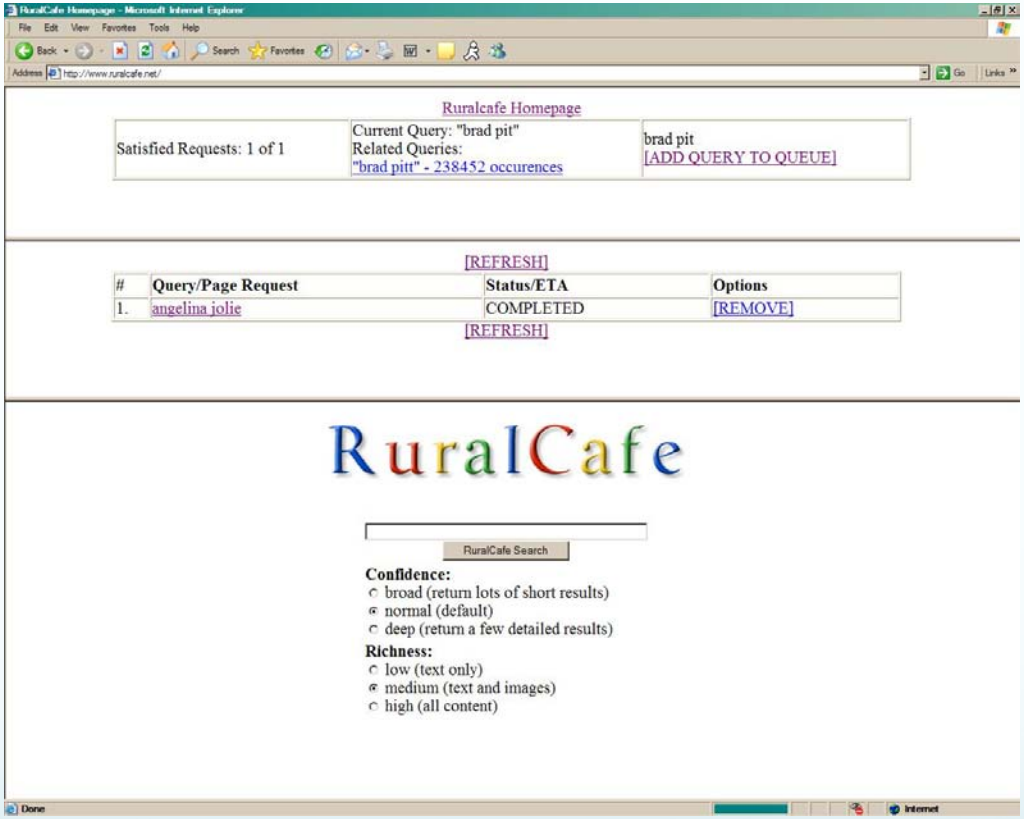
2Mbps Connection



RuralCafe: Intermittent Web Browsing



RuralCafe User Interface

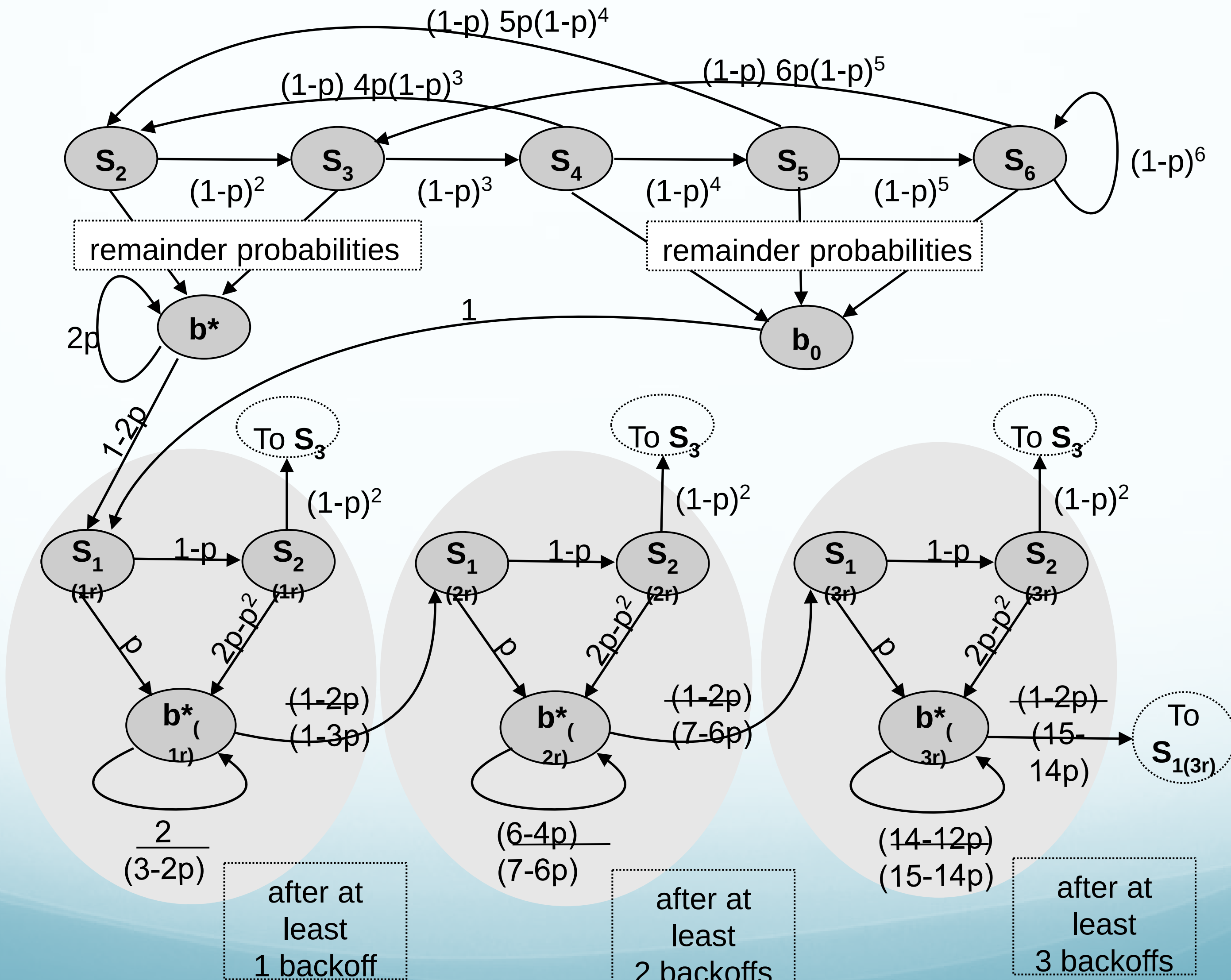


Positive user experiences from a deployment at Amrita University, India

The Sub-packet Regime

- Number of competing flows, $N >> 1$
- Per-flow fair share, $C/N < kS/RTT$, where
  - C is the link capacity,
  - k is a small integer (e.g. less than 3),
  - S is the packet size, and
  - RTT is the round trip time.

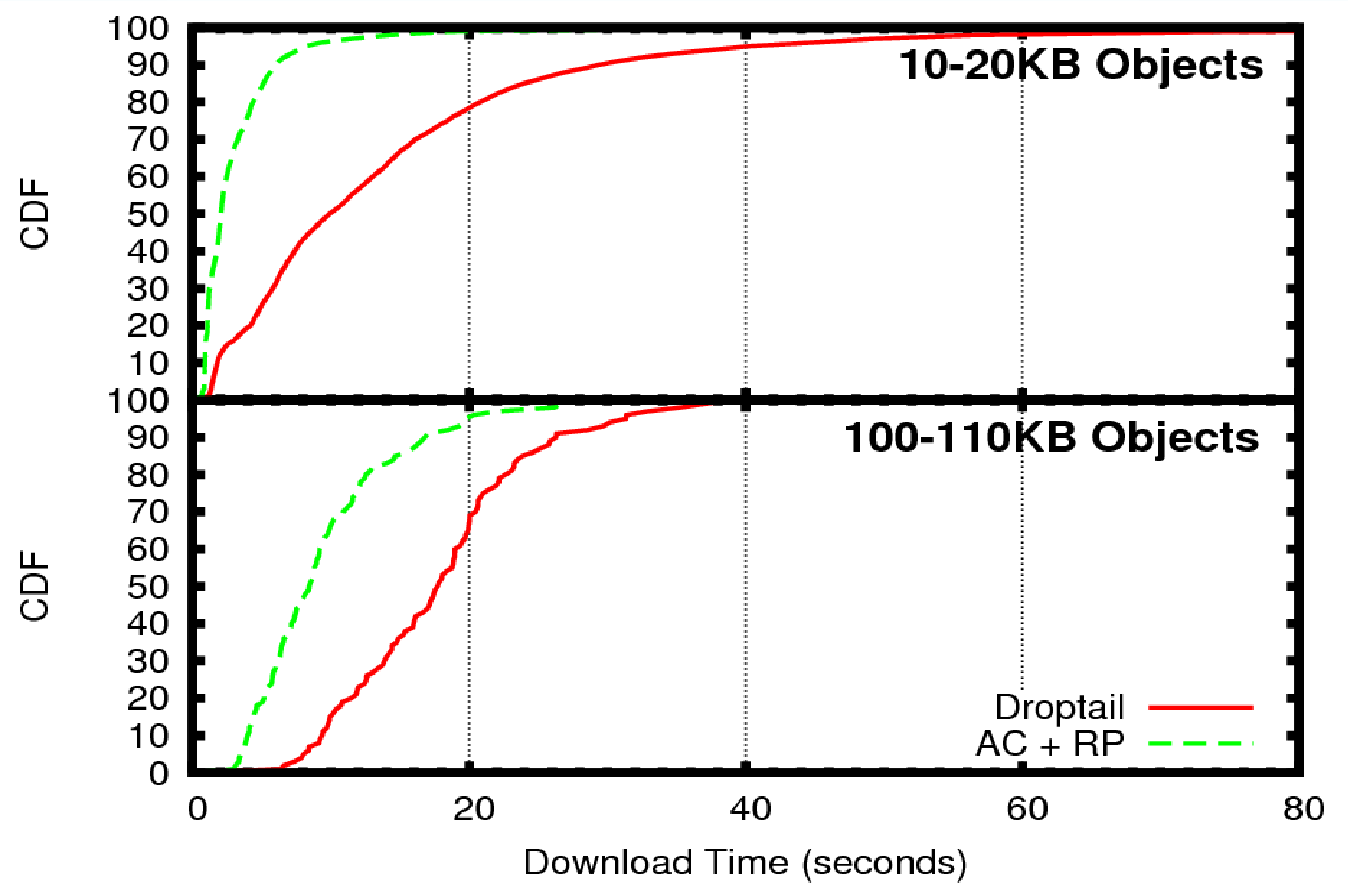
Why TCP breaks down?



Fixing the TCP breakdown

- Key Idea: Avoid the Sub-packet Regime
- Solution Approach
  - Recognize flow pools
  - Use admission control to keep TCP in the good operating range < 10% loss
  - Fine grained packet scheduling
  - Avoid timeouts due to dropping retransmissions

Overall Performance Gains



Seachable Contextual Caches

- Build a cache a smart cache that understands 'topics'
  - Allow users to search the cache for the *information* they need rather than the exact *URLs*
  - Cache by topic hit rate rather than page hit rate
  - Make each “topic-specific” cache searchable
    - A local Google experience

Building Contextual Caches

- Identify topics
  - queries, content, domains
- Identify cached authorities for each topic
- Popularity-driven focused crawling
  - document classifier for topic
  - vertical crawl
- Local indexing per topic
- Updating topic-specific portals

Recap

- Connectivity for the Next Billion
  - WiRE, WiLDNet, Mesh networks, Reliable Power
- Next generation Mobile Services
  - SMSFind, Hermes
- Web Architecture
  - RuralCafe, Sub-packet regime, Contextual Caches

Acknowledgements

- NYU: Jay Chen, Aditya, Ashlesh, Matt, Michael, Jinyang
- Berkeley: Rabin, Sergiu, Sonesh, Eric