Challenge 3:
Extending the Web to the Developing World

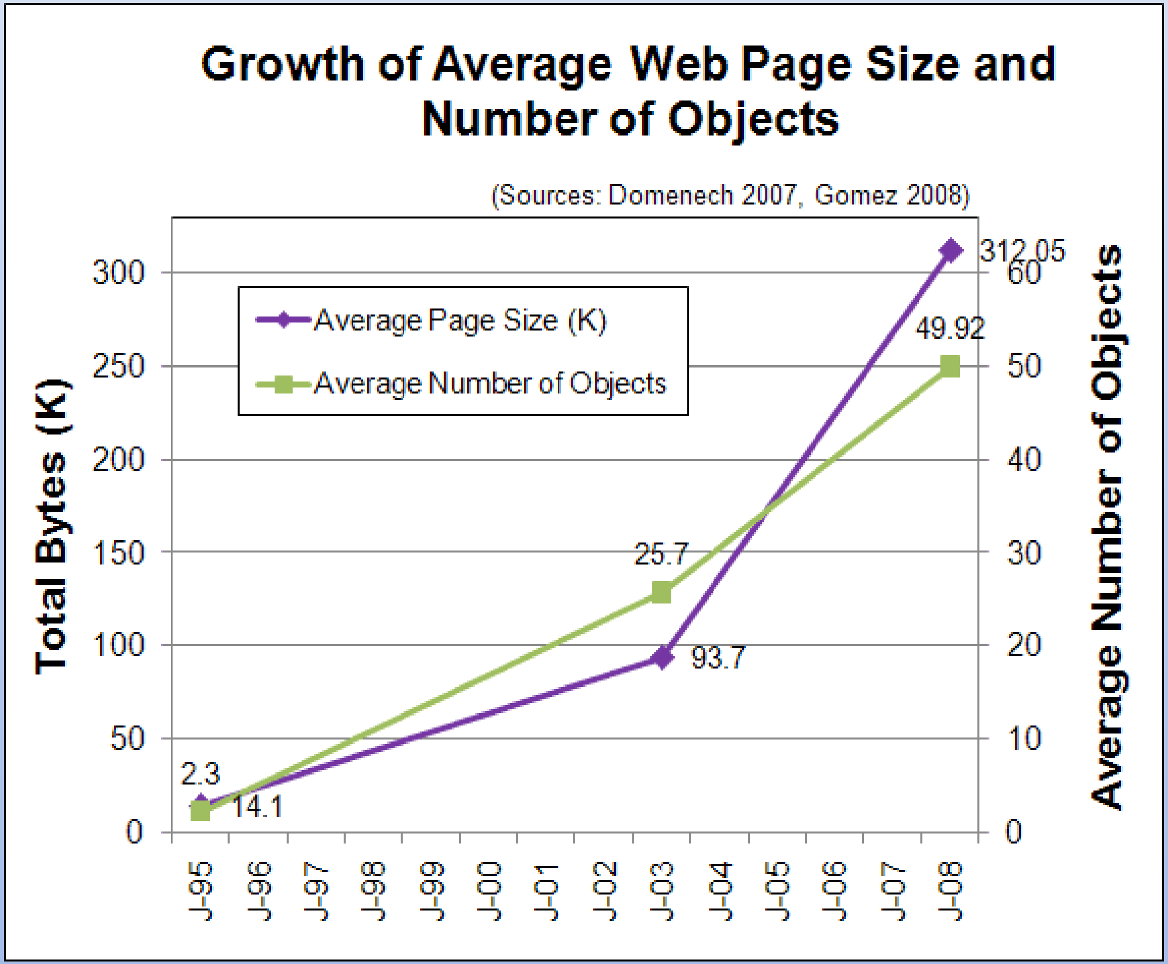
NYU | CATER

High School – 60km Outside Nairobi

Extremely low bandwidth connectivity in many parts of the world

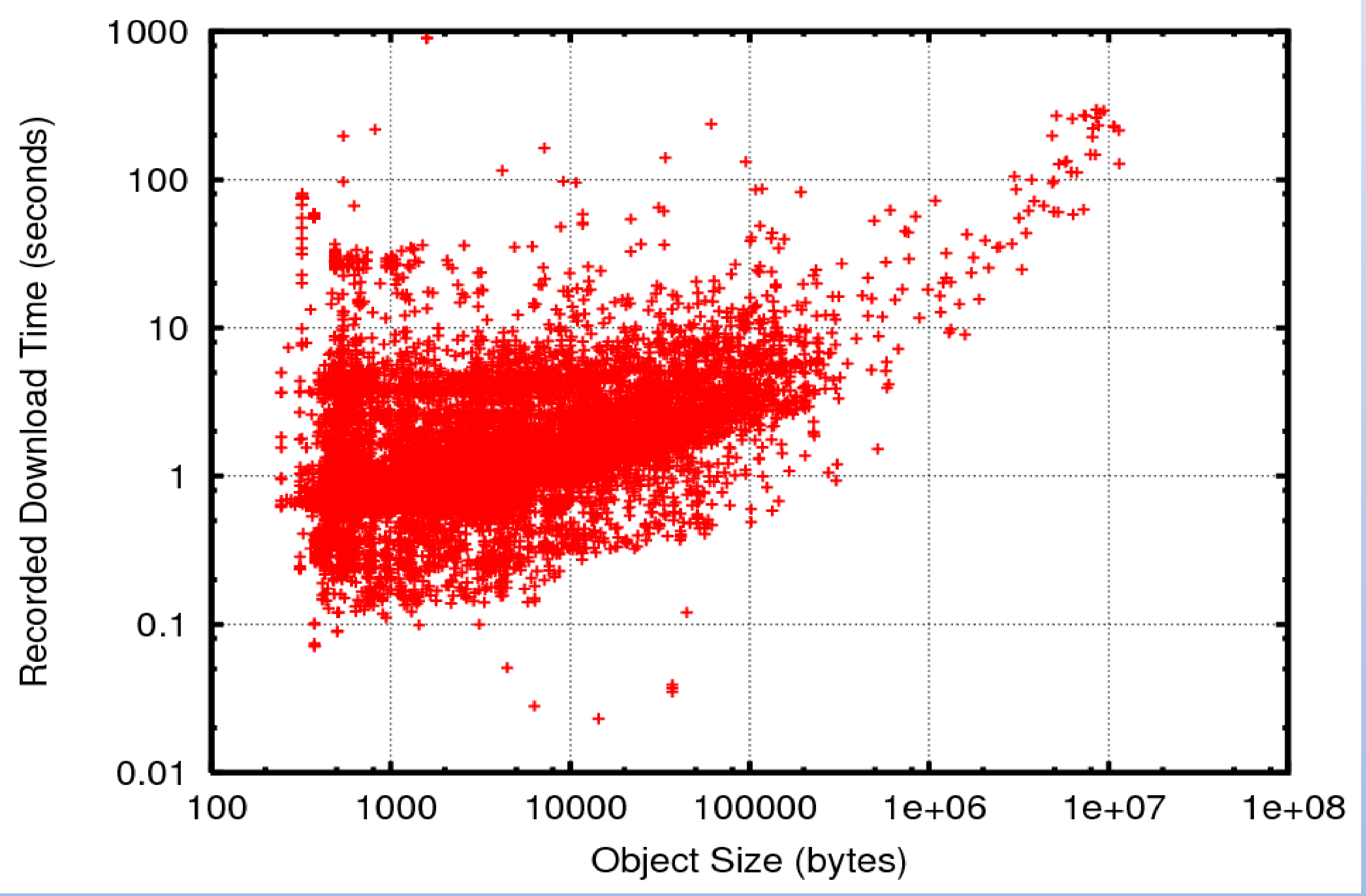
NYU | CATER

Web Page Size



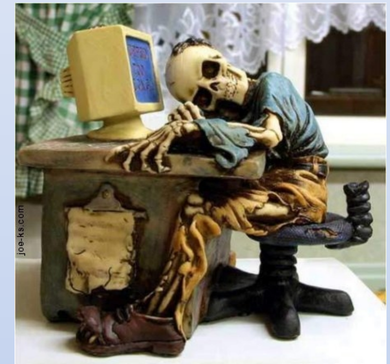
NYU | CATER

2Mbps Connection



The Web under Poor Connectivity

+

=


(acoustic modem)

- **Video + audio + images => web pages are huge**
- **A couple Kbps per user => User waits indefinitely**
- **Browsers make too many connections**
- **Iterative search infeasible!**
- **TCP itself actually starts breaking down**
  - **Not designed for these *“sub-packet regimes”***

How do we fix this?

Towards a Usable Web?

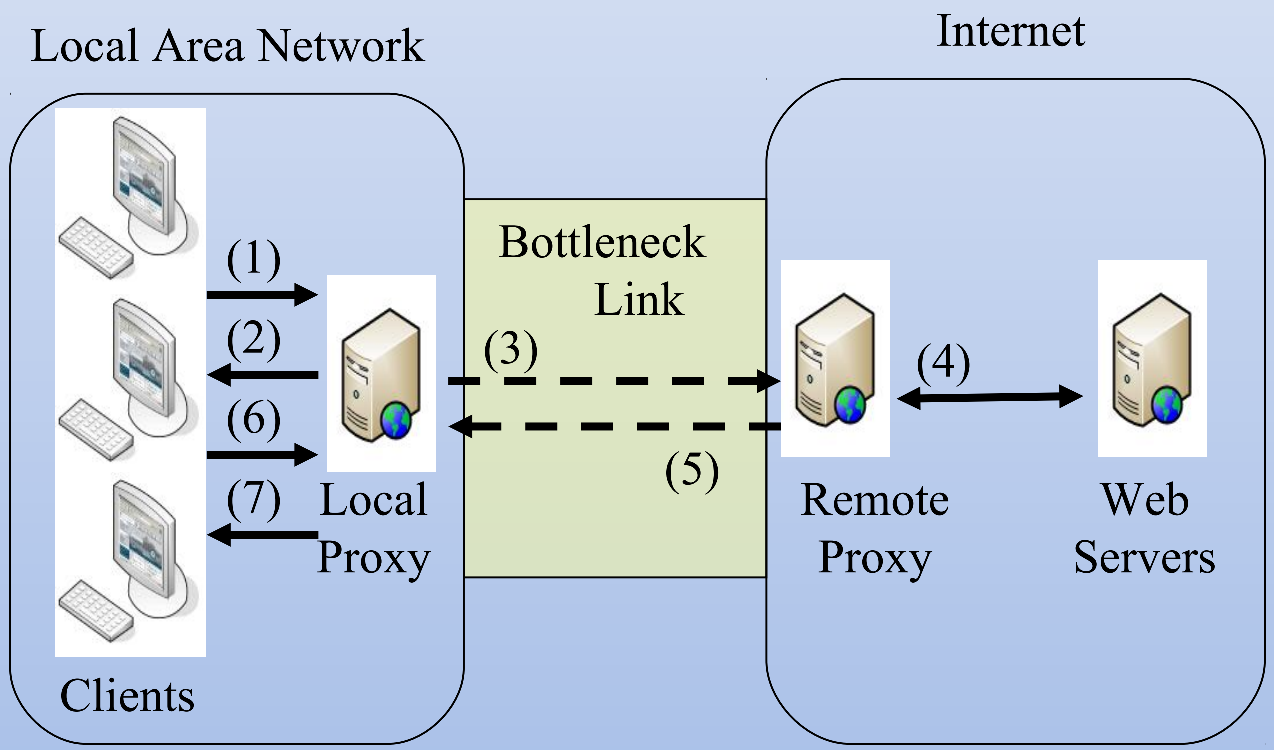
- Interactive
- Works over low bandwidth (without TCP breakdown)
- Intermittent/Delay tolerant
- User feedback (intermittent aware)
- User control

Step 1: Provide connectivity!
(or the illusion thereof)

Application Level Solutions

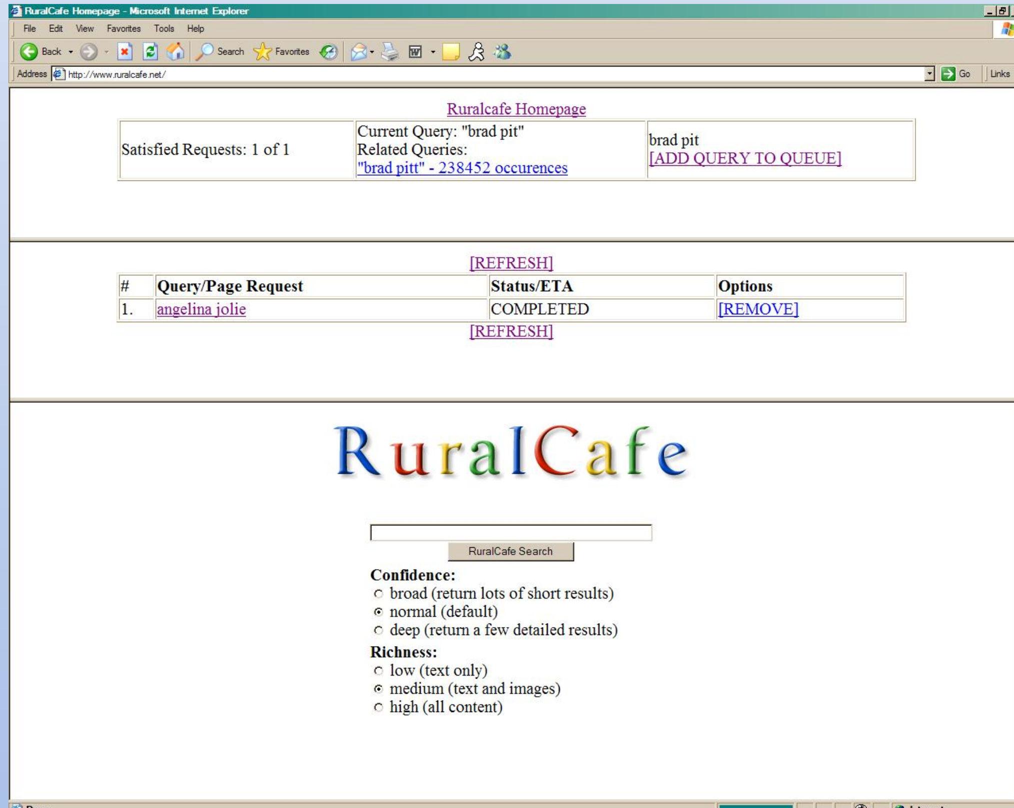
- Use proxies to provide an illusion of Web connectivity
  - One proxy allows users to interact immediately
  - While another proxy works to fetch web pages asynchronously that are queued by the user
- Time Equals Knowledge (TEK) - SMTP as its transport protocol
- RuralCafe – user feedback and control over the content to be downloaded

RuralCafe: Intermittent Web Browsing



NYU | CATER

RuralCafe User Interface



Positive user experiences from a deployment at Amrita University, India

NYU | CATER

Step 2: TCP breakdown problem?

NYU | CATER

The Sub-packet Regime

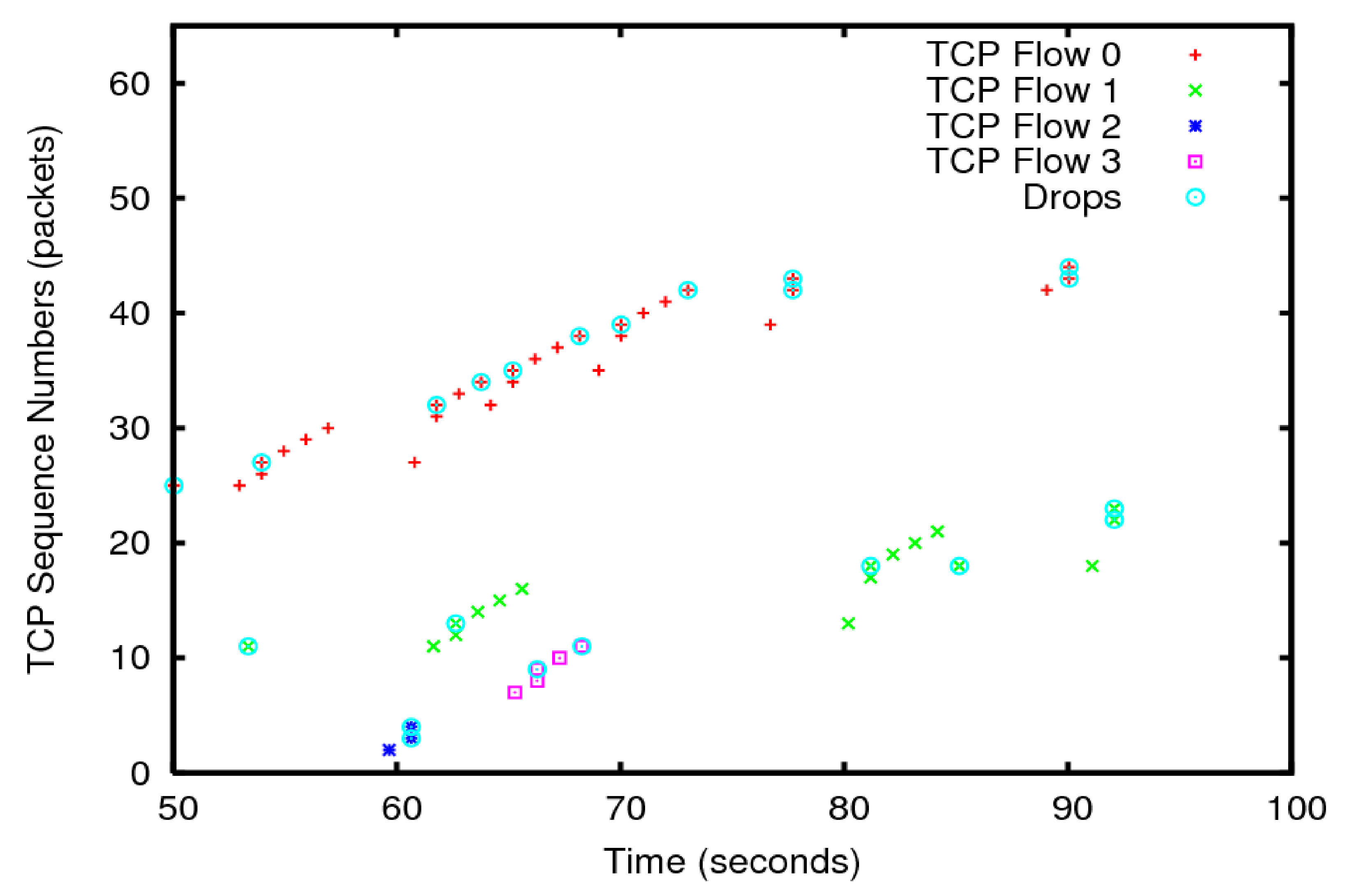
- Number of competing flows, $N >> 1$
- Per-flow fair share, $C/N < kS/RTT$, where
  - C is the link capacity,
  - k is a small integer (e.g. less than 3),
  - S is the packet size, and
  - RTT is the round trip time.

Pathological Sharing: A TCP View

- High packet loss rates
- Elongated and highly variable timeout periods
- Extreme unfairness in the "short" and

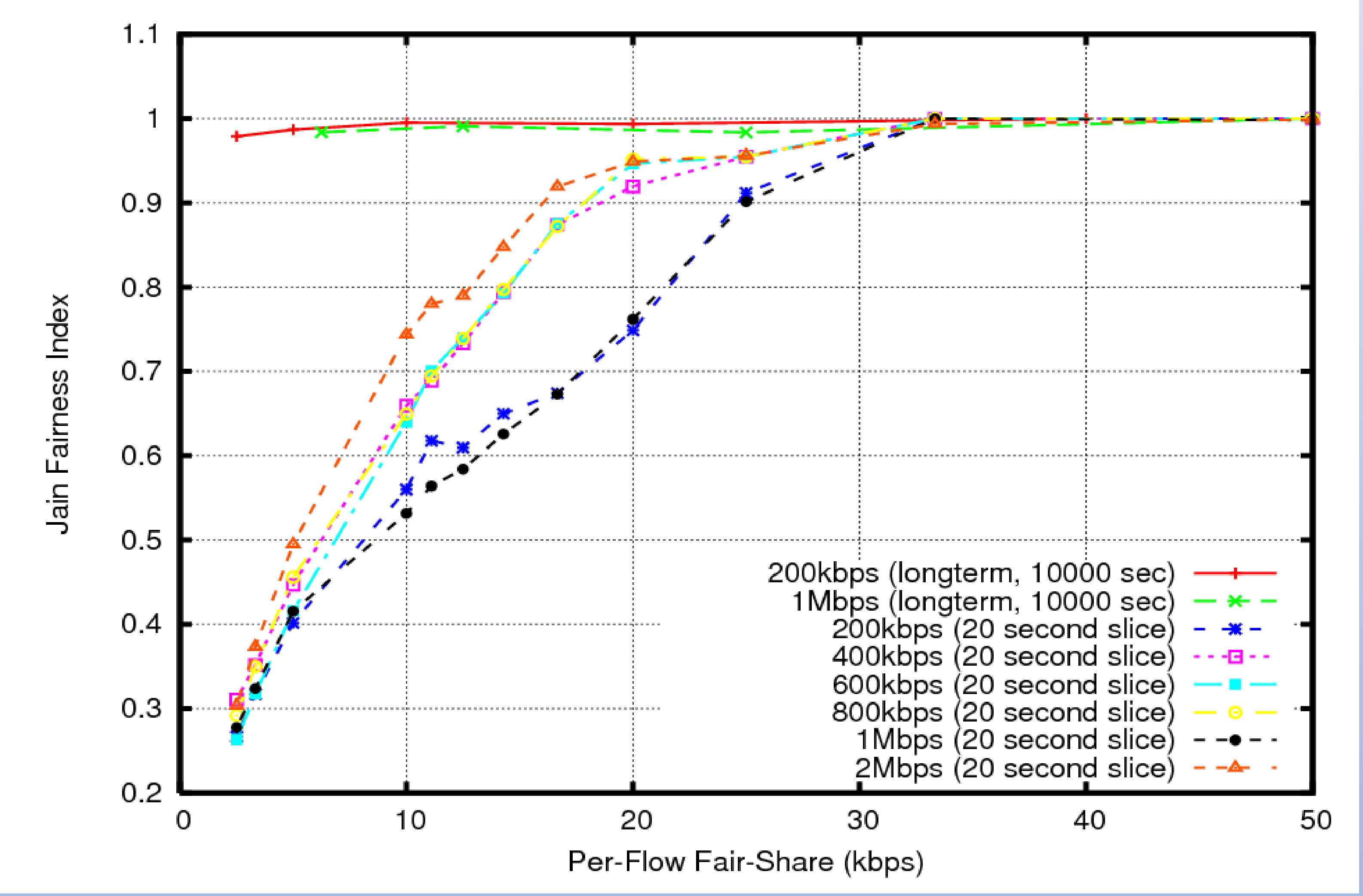
  "long" term
- Resulting in unpredictable flow completion times

Loss Rates and Timeouts



NYU | CATER

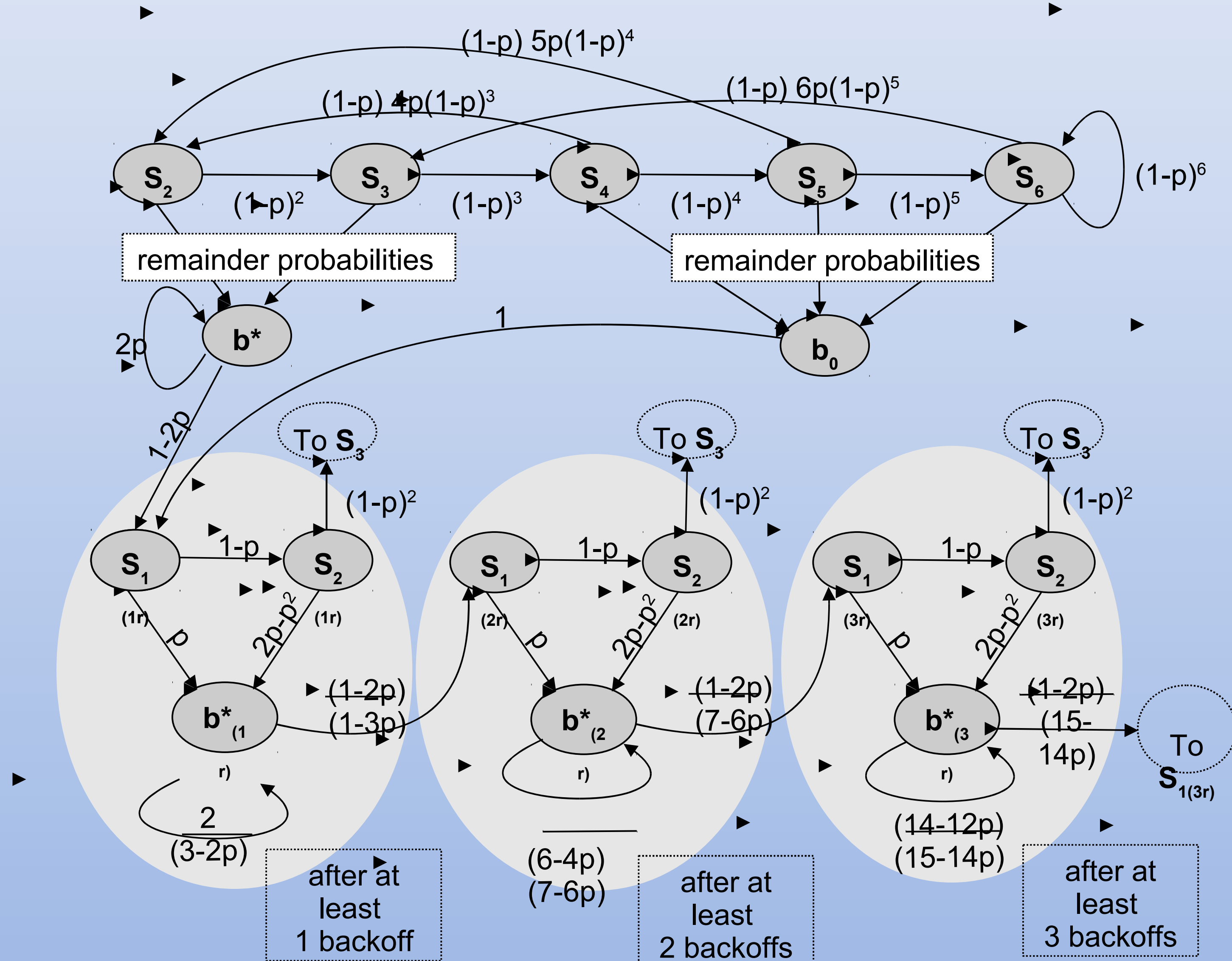
Fairness



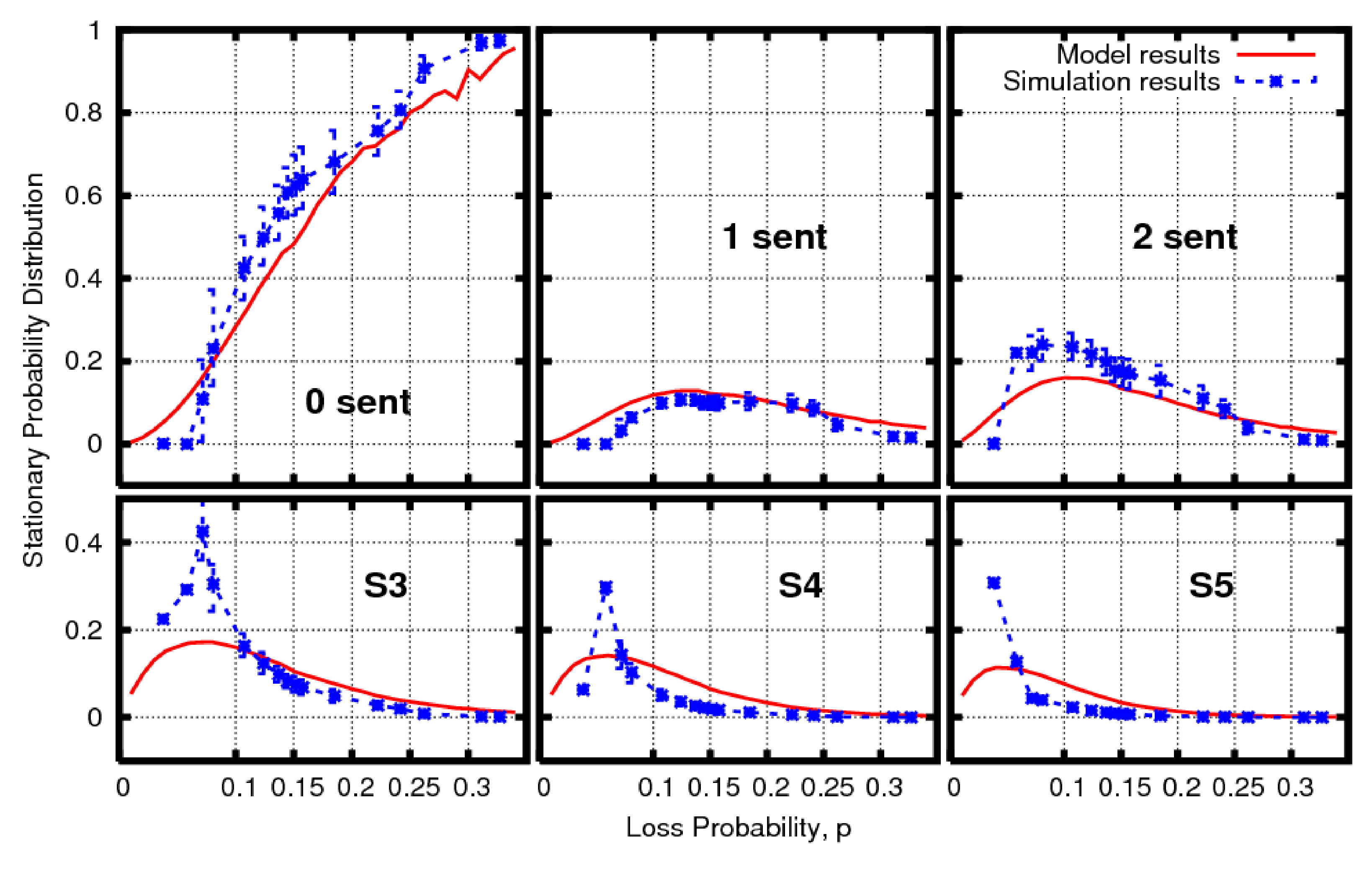
NYU | CATER

Why TCP breaks down?

Model


(1-p) 5p(1-p)4
(1-p) 4p(1-p)3
(1-p) 6p(1-p)5
S2
S3
S4
S5
S6
(1-p)6
(1-p)2
(1-p)3
(1-p)4
(1-p)5
remainder probabilities
remainder probabilities
2p
b*
1
b0
1-2p
To S3
To S3
To S3
(1-p)2
(1-p)2
(1-p)2
S1
1-p
S2
S1
1-p
S2
S1
1-p
S2
(1r)
p
2p-p2
(1r)
(2r)
p
2p-p2
(2r)
(3r)
p
2p-p2
(3r)
(1-2p)
(1-3p)
(1-2p)
(7-6p)
(1-2p)
(15-14p)
b*(1r)
b*(2r)
b*(3r)
2
(3-2p)
(6-4p)
(7-6p)
(14-12p)
(15-14p)
To S1(3r)
after at least 1 backoff
after at least 2 backoffs
after at least 3 backoffs

Validating the Model



Takeaways from the Model

- Beyond a loss rate of 10% the stationary probability of TCP in timeout states rapidly increases
- Loss of retransmissions incur the high cost of increasing timeout periods (flow shut-off)
- At high contention levels 60-90% of flows are shut-off for elongated time periods
- TCP waits for a new data packet before updating the RTT estimate

Fixing the TCP-breakdown problem without Modifying end-hosts:

Key Idea: Avoid the sub-packet regime

Admission Control

- TCP can only handle some number of flows before it breaks down
- Use admission control to keep TCP in the good operating range < 10% loss
- If we preform admission control on a per-flow basis, some applications that require multiple flows to make progress will still fail

Flow Pools

- A collection of inter-related flows from the same source to different destinations that are initiated within a short time-period
- So a single application can make progress with all of its required flows being admitted simultaneously
- A new flow is admitted if:
  - It belongs to a flow pool which has already been admitted
  - It belongs to a new flow pool and the current number of flow pools is below the maximum

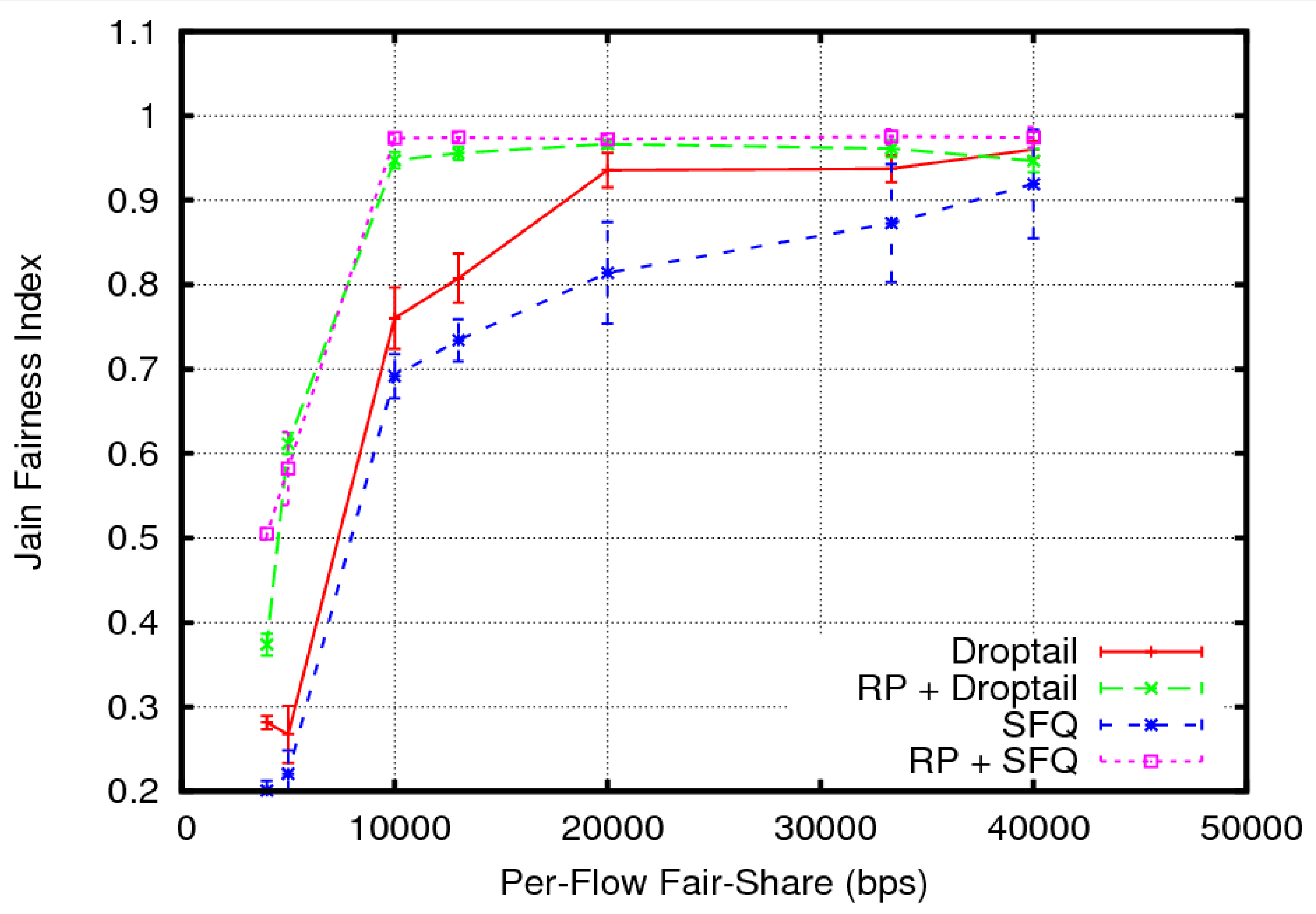
Fair Share

- Each flow pool should be isolated from the other so a single flow pool does not consume all of the resources by simply creating more flows

Fine-grained Packet Drops

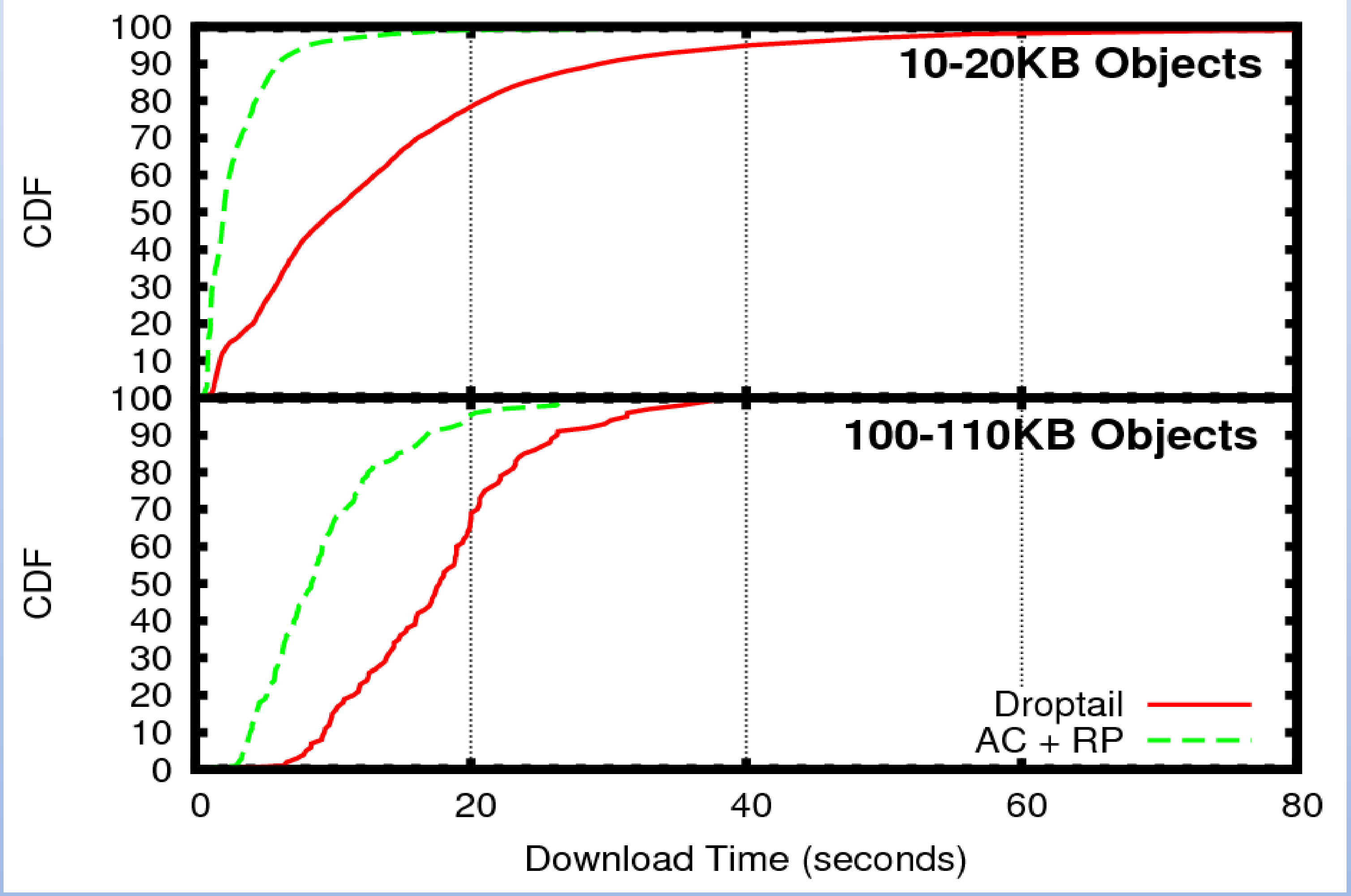
- Retransmissions are important
- Repeated drops cause TCP to collapse
- Prioritize retransmitted packets

Short-term Fairness



NYU | CATER

Overall - Object Download Time



NYU | CATER

Key Takeaways

- We've “fixed” TCP for these sub-packet regimes
  - Provided fairness and isolation
  - Improved predictability
  - Allow progress without hangs
  - Improved overall capacity

Step 3: Contextual Web Caching
Getting appropriate content locally?

Content

- How to get appropriate content?
  - In the local spoken language
  - For specific environmental or social settings
- Small communities with very specific information needs: schools, villages, hospitals, NGO offices, kiosks
- But they also have very broad information “wants”

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

Takeaways

- RuralCafe
- Sub-packet Regime
- Contextual Caches

Challenge 4:
SMS based applications

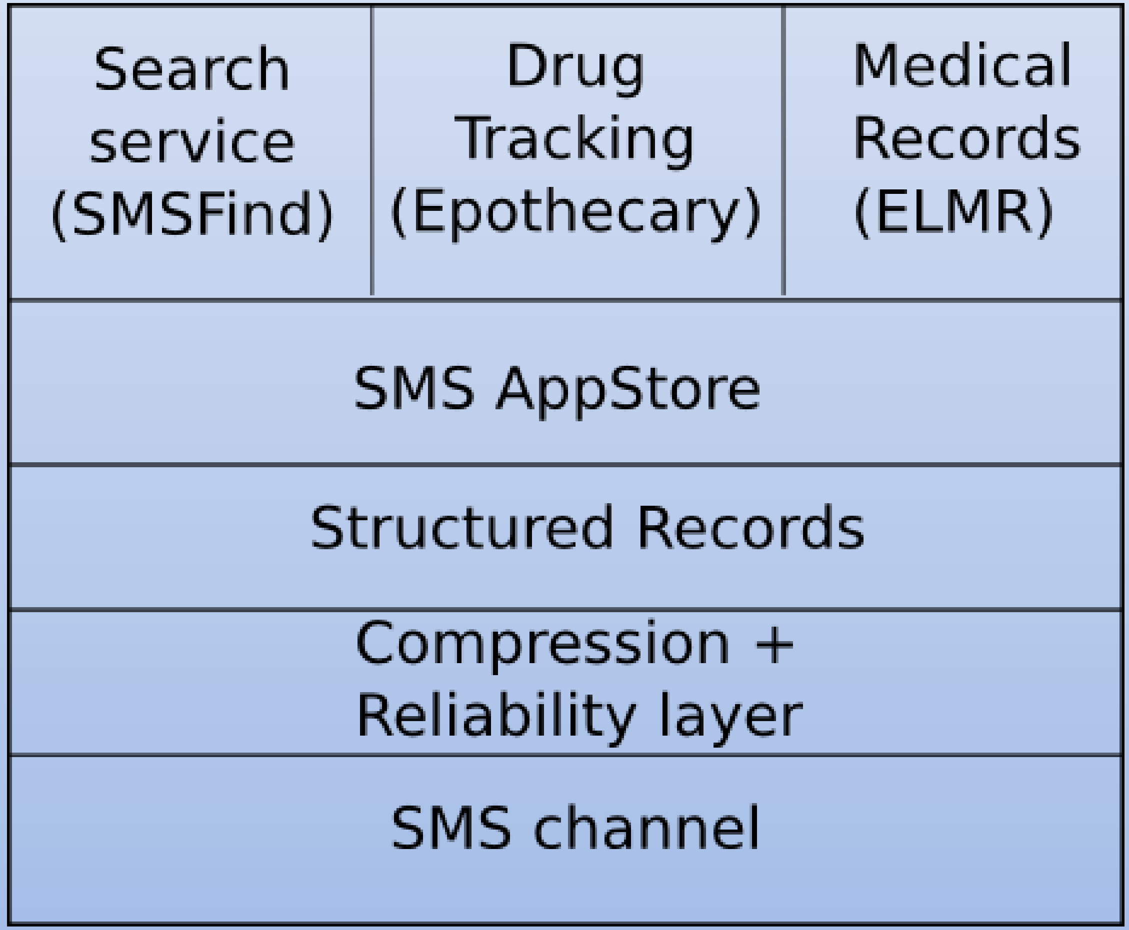
Existing Systems

- Mainly for Smart Phones
- Rely on GPRS network connectivity
- Rural settings have only voice and SMS.
- Examples: OpenRosa, Voxiva, OpenMRS

SMS apps

- Why SMS based apps are hard to create
  - 140 characters (or bytes)
  - Operational dependencies: need to have carrier permission
  - Examples: Frontline SMS, Rapid SMS

SMS stack


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

ELMR Challenges & Solutions

- SMS is just 140 byt — *Restricted Operations Set*
- Each SMS costs — *Semantic Compression*
- Reliability is an issu — *Lightweight Reliability Layer*
- Patients want priva — *Lightweight Privacy Layer*

Symptoms Form

Symptoms Form:

1. Patient ID ….

2. Do you, or have you ever had tuberculosis? ▯-Yes ▯-No

3. Do you sleep with a treated bed net? ▯-Yes ▯-No

At any time while being on ARV therapy have you experienced any of the following?

4. Rashes or skin problems anywhere on your body?
   ▯-Yes ▯-No

5. Sensation of burning, stinging, stiffness, tickling or numbness in the feet, toes or hands
   ▯-Yes ▯-No

6. Diarrhea
   ▯-Yes ▯-No

7. Weight Loss
   ▯-Yes ▯-No

8. Do you have pain while swallowing?
   ▯-Yes ▯-No

9. Weakness
   ▯-Yes ▯-No

10. Shortness of breath
    ▯-Yes ▯-No

11. Coughing up blood
    ▯-Yes ▯-No

12. When was the last time you had malaria?
    ▯-This year ▯-Last year

140 Bytes SMS

Form having **350** Symptoms questions

NYU | CATER

SMSAppstore

- SMS based application-store
  - Separate the application from the mobile platform
- Features
  - Automatic semantic compression
  - Standard operations: Fetch, update, create, search, structural data specification
  - Local operations: User defined
  - Bulk update/fetch
- Apps
  - Health records, Drug tracking, Mobile sensing, Mobile Craigslist, SMS-search, Solar monitoring

SMS-Based Search for Low-End Phones

Problem Statement

- *“Low-Cost handsets to account for over 50% of mobile phones by 2014”* -Juniper Research '09
- These low-end phones will be owned by the world's poor, and only have voice and SMS capabilities, where SMS is the only data channel available.
- For mobile information services, efficient SMS search is critical.
- We seek to: Answer an arbitrary Web search query or question with a single SMS message (140 bytes)

Solution Idea

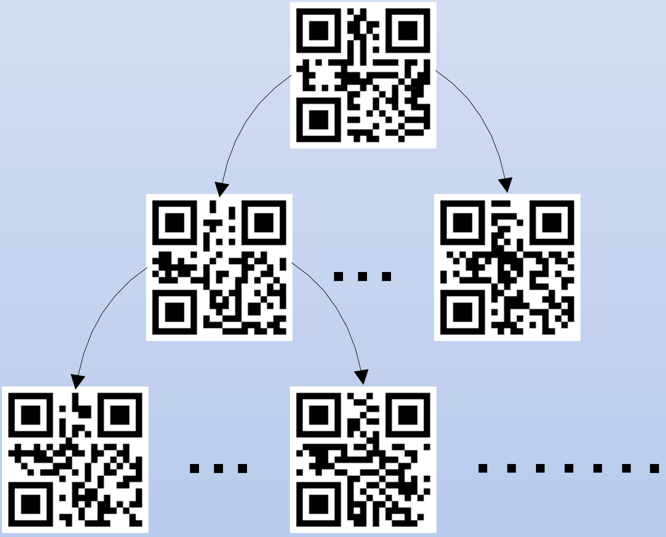
- Allow the user to explicitly disambiguate their query by including a “hint”
- Using the hint, we tailor existing Information Retrieval
- Techniques to find a short snippet answer (140 bytes long) among relevant documents returned by a search engine
- Conceptually: For the query “barack obama wife”

  where “wife” is the hint, we find that “michelle obama” often appears near the word “wife” in the top search result pages

Secure Drug Tracking

Epothecary

- Uses camera phones to scan unique glyphs affixed to units at each level of packaging and tags representing transacting parties



- Uses SMS or GPRS as available to convey scan information back to a central authority

What Does It Get Us?

- Fine-grain track and trace of sold pharmaceuticals
- Strong assurances to the consumer of the authenticity of drugs purchased through a participant in the system
- Greatly speeds tracking of problems in the supply chain

Questions?

Challenge 5:
Security challenges in the developing world

Security: a hard problem?

- Scenario:
  - Non existence of ID cards
  - Trust is always an issue
  - Constrained resources (infrastructure is sparse, low tech devices)
  - Low connectivity or no connectivity
  - Offline authentication
  - People are street smart!

Security: a hard problem

- Traditional security often fails
  - Constrained resources
  - Human in the loop
  - Low tech devices
- Mobile banking transactions are SMS based!
- Outdated GSM standards
  - How unsecure is that!?

Representative projects

- Secure mobile services
  - Epothecary: Secure drug tracking
  - Signet: Low cost auditable transactions
- Trust and Identity management
  - PaperSpeckle: Paper based secure transactions
  - Secure branchless banking
- Outdated GSM standards

Low-cost Auditable Transactions Using SIMs and Mobile Phones

NYU | CATER

Problem

- Paper receipts are ubiquitous: used in microfinance, healthcare
- But, extremely unreliable: repudiation, fabrication, damage

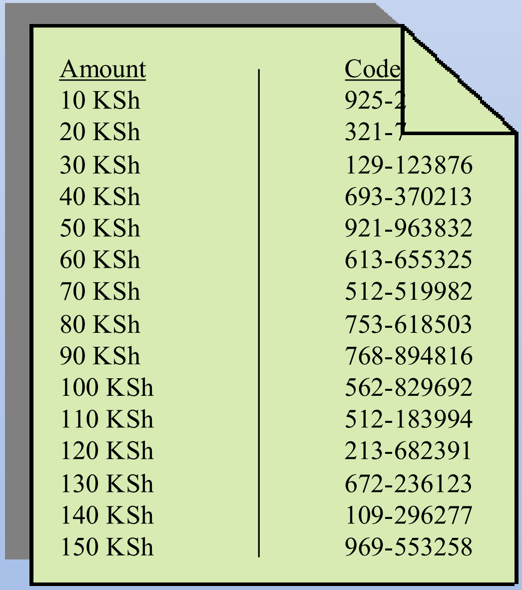
- Need a low cost, secure transaction process

Existing approaches

- POS devices: expensive (~ $400)
- Build the network: expensive
- GPRS/SMS: Coverage not completely ubiquitous, high marginal cost relative to transaction value, particularly with SMS, still requires you to trust other people to hold your data

Signet

- Uses secure computational capacity in SIM cards to perform lightweight signing of transactions
- Confirms transactions OOB to ensure tamper evidence
- Uses SMS or GPRS as available or affordable to ‘lazily’ synchronize central server.



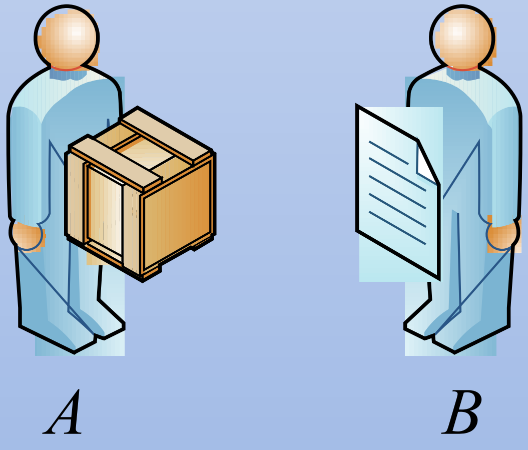
Amount	Code
10 KSh	925-2
20 KSh	321-7
30 KSh	129-123876
40 KSh	693-370213
50 KSh	921-963832
60 KSh	613-655325
70 KSh	512-519982
80 KSh	753-618503
90 KSh	768-894816
100 KSh	562-829692
110 KSh	512-183994
120 KSh	213-682391
130 KSh	672-236123
140 KSh	109-296277
150 KSh	969-553258

Protocol: Prerequisites

- Each user of the system receives
  - A (U)SIM card with a signing application installed
    - If the user has his own programmable phone, he also receives the client application, OTA or otherwise
  - A printed booklet containing transaction amounts and associated signatures

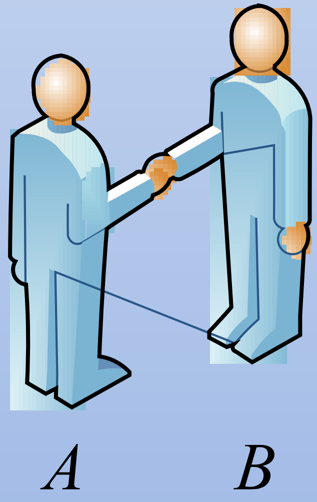
Protocol: Transactions

- Each party brings something to transact
  - In the simplest case, some funds or goods and a receipt
- Each party inserts his SIM into a phone
  - One party may supply both phones



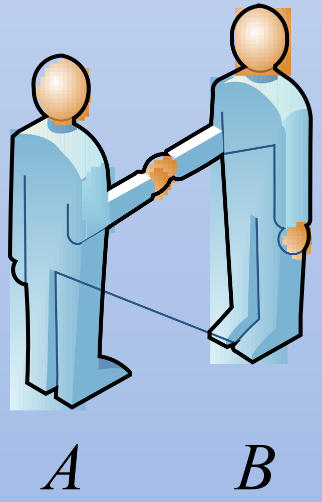
Protocol: Transactions

- The two parties agree on terms
- Party *A* inputs metadata about the transaction into the handset
- The SIM in the handset signs the data it receives and also returns its public key signed by a central authority



Protocol: Transactions

- The handset communicates this to Party *B*’s handset, and *B*’s SIM verifies the relevant signatures, signs the metadata symmetrically, and returns them to *A*’s handset.

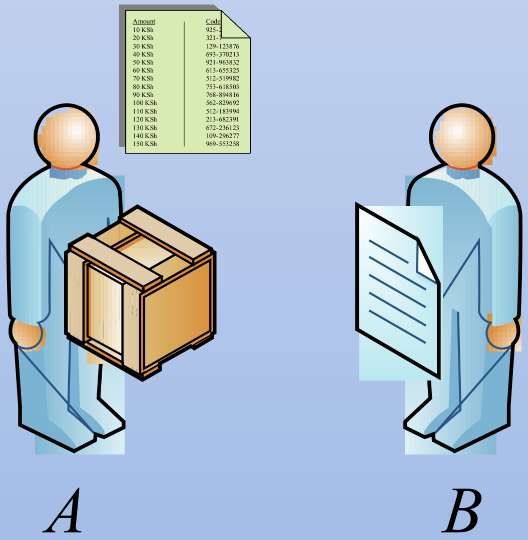


NYU | CATER

Protocol: Transactions

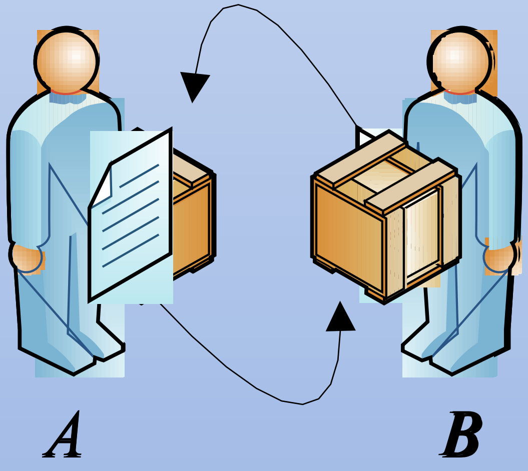
- *A*’s SIM verifies the signatures, then re-signs the metadata with a different key and displays the last *n* bits as a number

- *A* verifies this number with his printed reference book to ensure that the data he got back was accurate



Protocol: Transactions

- Once this check passes
  - *A* surrenders his goods
  - *B* takes possession
  - And the parties part



NYU | CATER

Protocol: Communication

- Communication between handsets takes place over Bluetooth or IR
- Transaction metadata are batched by default in order to amortize transmission cost across several transactions when sending to third party storage

Protocol: Verification

- Assuming a well-known third-party keysigner and associated keys
  - Receipts are nonrepudiable as each party's keys are signed by the keysigner/CA
  - Each party retains an independently verifiable digital copy

Secure Branchless Banking

Rural banking

- ~ 1B people have cell phones but no bank account[1]
- Banking, money transfer a major problem
- Not cost effective both for banks and people

1. CGAP survey June 2009.

NYU | CATER

Branchless Banking

Use existing retail infrastructure and agents

Shopkeeper

Use existing technology

NYU | CATER

Setting

- Rural villages have no banks.
- Traveling to a bank is long and arduous.
- Banking is needed – put away harvest money, get money for seeds.
- Banks deputize shopkeepers to act as agents.
- Shopkeepers literate. Farmers can read numbers. Shopkeepers have cellphones.

Trust Relationships

- Shopkeeper trusts bank.
- Farmer trusts bank.
- Shopkeeper does not trust Farmer. Farmer does not trust Shopkeeper.

Goals

- Farmer and shopkeepers travel to the bank seldom.
- Farmer can do banking (deposit, withdrawal) with shopkeeper and each can prevent cheating from the other or from intermediaries in the insecure phone line.

Protocol: Withdrawals

- F -> S : X_i , ID_f
- S -> B : Keyin(X_i, Am, ID_f, ID_s, N_s)
- F -> B/S : Voicein(Trans details)
- B -> F/S : δ_i | stale(X_i) ;

Compute δ_i = (Am, Y_i); Compare(δ_i , δ_i’)

- F -> B/S : Keyin(Z_i)
- B -> F/S : Accept/Reject
- S -> F : Am
- S -> F : Receipt(N_s)

Embodiment: matrix of numbers

- Suppose that F wants to deposit 534 rupees.
- Bank responds with “Y[i]-matrix”
- 5 3 4
  4 7 3
  2 5 6
  8 2 9
  4 9 3

Y[i] is a set of relationships

- =4 +4 +9
  -3 =5 =6
  =8 +9 =9
  -1 +6 -1

- =x means that the value should be x;
  +x means to take the value of the
  transaction digit and add x modulo 10;
  -x, similarly.

Creating the Y[i] matrix

- Start with 5 3 4
- Apply Y[i]:
  =4 +4 +9
  -3 =5 =6
  =8 +9 =9
  -1 +6 -1
- Result:
- 4 7 3
  2 5 6
  8 2 9
  4 9 3

Existing
(M-Pesa, GCash, WIZZIT)

- Pure cellphones with SMS. All money is electronic.
- But cellphones are not secure.
- Sim card number can easily be hacked. Numbers rerouted.
- Cellphones are often shared.

What Have We Accomplished

- No replay attack, once Z[i] is revealed, the amount of the transaction cannot be changed and X[i] and Z[i] can’t be used later.
- No man in the middle attack – insufficient information.
- No need for crypto.
- No need for secure phone lines.

Questions?