Information Access and Communication Networks in the Developing-world

Umar Saif

LUMS, Pakistan

umar@lums.edu.pk | umar@mit.edu

Overview

- **Poor Man's Broadband**
  - Poor Man's Cache
  - Packet Containment
- **TEK Internet Search**
- **Inverse Multiplexing of Cellular Connections**
- **Teleputer (Time permitting)**

Motivation

Developed World	Digital Divide	Developing World
2 MB Internet Connection < $40		2 MB Internet Connection > $4000
Bulk Data Transfer on the Internet > 70%		Bulk Data Transfer on the Internet < 15%
Average End-user Bandwidth via ISP > 100 kb/sec		Average End-user Bandwidth via ISP < 10 kb/sec

---

Internet in Pakistan

- **Facts of life in the developing world**
  - **Expensive International Bandwidth**
  - **No real peering points**
  - **Internet used over dialup**
    - Poor "Scratch card" provisioning


Internet in Pakistan

- **Average Dialup Bandwidth**
  - Less than 10 kb/sec
- **Almost Never Used for**
  - Exchanging
  - Disseminating
  - Accessing
  - …. Content larger than a couple of hundred kilobyes

How I Stumbled Upon this?

- **"Good research solves *real* problems in a practical way"**
  - Started last year when I wanted to exchange a 3.5 MB PDF file with my dad
  - Two laptops sitting next to each other
  - No way to exchange data if you don't have portable storage!
    - We actually went our and bought a CDR to exchange data….

Problem

Internet

<10kb/sec

~ 56kb/sec

Not a Last Mile Problem
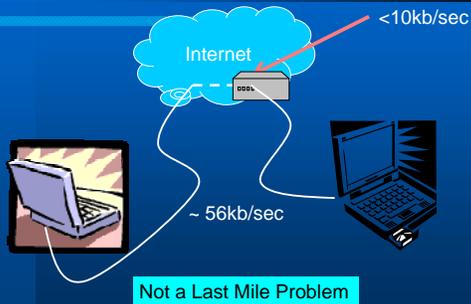
Solution

Bypass the Internet when exchanging large

Internet

~ 56kb/sec

Email Attachments

- **Time to exchange a 3.5 MB file on the Internet ~ 1 hours (16 Kb/sec)**
  - – 30 mins upload and download
  - – Assuming no disconnections
- **Time now (40 kb/sec)**
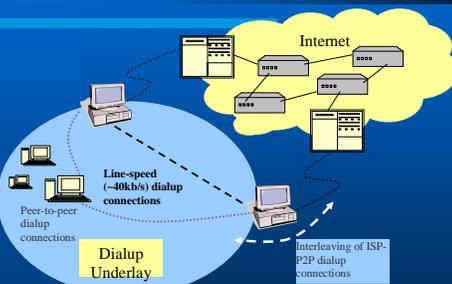 - – 12 mins!!

Disruptive Technology

- **Of course Internet also started as an overlay over the phone lines**
- **A new kind of Internet**
- **Reminiscent of Pre-Internet days**
  - – FidoNet
  - – UUCP

Why is this Practical?

- **Phone bills are becoming "Flat"**
  - – Rs 199/month -- free nation-wide calls
- **But cannot always connect to the server**
- **As long as one can identify a "close-by" host, "broadband access" is free**
- **P2P systems already follow a similar model**
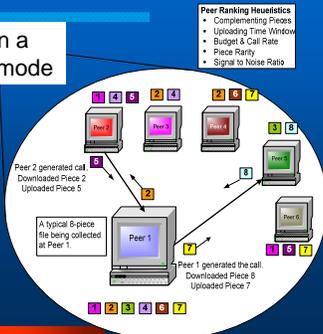 - – Incentive-based BitTorrent

Dialup P2P-ISP Interleaving

Key Idea: Use Internet as a directory service, not as digital pipe

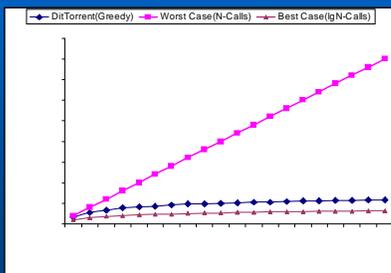Internet

Line-speed
(~40kb/s) dialup
connections

Peer-to-peer
dialup
connections

Dialup
Underlay

Interleaving of ISP-
P2P dialup
connections

2

Dialup BitTorent

BitTorrent in a sequential mode

Peer Ranking Heueristics
- Complementing Pieces
- Uploading Time Window
- Budget & Call Rate
- Piece Rarity
- Signal to Noise Ratio

Peer 2 generated call
Downloaded Piece 2
Uploaded Piece 5

A typical 6-piece file being collected at Peer 1.

Peer 1 generated the call.
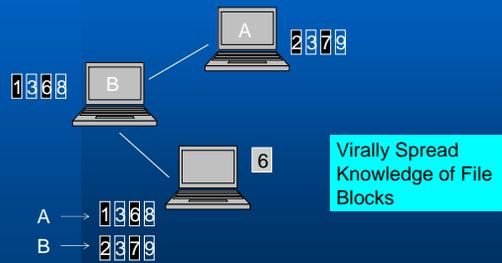Downloaded Piece 8
Uploaded Piece 7

Other Challenges

- **Overhead of Peer connections: ~30 sec**
- **Offline Block Discovery**
- **Last Block Problem**
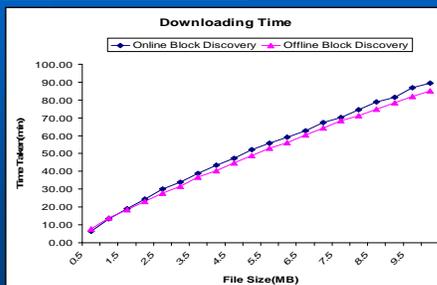- **Flash Crowds (Backoff for congestion control)**
- **Budget-based Download**
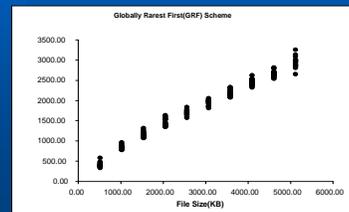

Peer Connection Overhead

DitTorrent(Greedy) — Worst Case(N-Calls) — Best Case(lgN-Calls)

Offline Block Discovery

A

2 3 7 9

1 3 6 8

B

6

Virally Spread Knowledge of File Blocks

A → 1 3 6 8
B → 2 3 7 9

Offline Block Discovery

Downloading Time

Online Block Discovery — Offline Block Discovery

Time Taken(min)

100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00

0.5 1.5 2.5 3.5 4.5 5.5 6.5 7.5 8.5 9.5

File Size(MB)

Last Block Problem

• Grab rare blocks first
• Favor those who will finish at the end of the connection

Globally Rarest First(GRF) Scheme

3500.00
3000.00
2500.00
2000.00
1500.00
1000.00
500.00
0.00

0.00 1000.00 2000.00 3000.00 4000.00 5000.00 6000.00

File Size(KB)

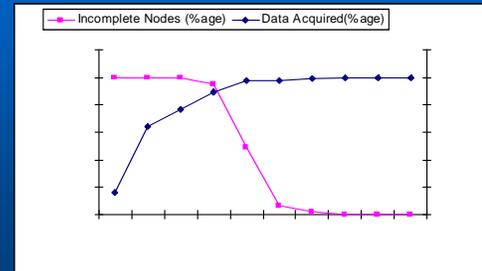## Flash crowds

- Each busy-tone costs ~10 secs
- Wait between calls (nominal)
- Back-off in times of congestion (busy tones)

Optimum Choking Interval

WBC=1 WBC=25

(Time Taken(min) vs Choking Interval)

Budget-based Download

Incomplete Nodes (%age)   Data Acquired(%age)
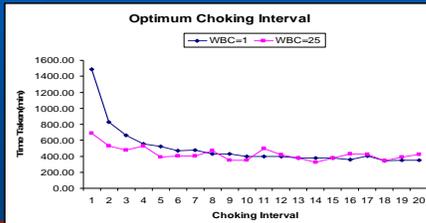
Three Evolving Applications

- P2P file-sharing
- Web-browsing
- Large Email attachments

http://dittorrent.sourceforge.net


Can we do the same on the "Internet"

- Contain packets within the developing-world
  - Routing paths are screwed up
    - "All roads lead to Rome" i.e. transit out of the country
- Low cache hit-rates
  - Too many small ISPs, no sharing of cached content
  - Misconfigured "Proxies"
  - Hit rates < 30% (instead of >60%)
- Poor DNS support

Work-in-progress

- Indirect Routing
  - Force routing paths by vectoring messages through intermediate nodes
  - Initial results show improvement across all traffic metrics
- An ISP-independent distributed cache
  - Similar to CoralCDN
- An ISP-independent DNS

ChoupalLink

**Inverse Multiplexing over cellular connections**
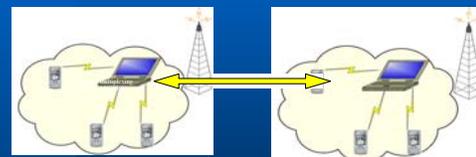
TEK: Time Equals Knowledge

- **Web Search for Low-bandwidth, intermittently connected users**
- **One of the first examples of mainstream ICTD research ~circa 2000**
- **Renewed interest with DTNs coming in vogue**

TEK

- **Internet in the developing-world**
  - **Expensive**
  - **Intermittent**

- **An Email-based Internet Search Facility**
  - **Asynchronous Dialup model**
  - **Search optimized for bandwidth and latency rather than speed**
  - **Heavy client-side caching**

TEK
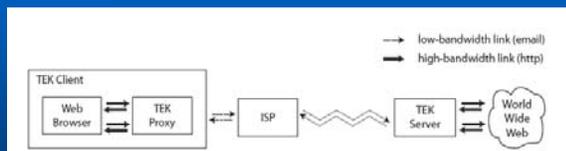
TEK Server

- **Remove Duplicate Content**
- **Cluster results**
- **Distinguish Content from links**
- **Remove images**
- **Remove background code**
- **Compress results**

Rationale

- **Lower operational costs**
  - **Caching vs internet download**
  - **ISP-host connection**
    - **Reliable**
    - **Higher Bandwidth**
    - **Cheaper: email-only account**
  - **Better utilization of Internet Connection**
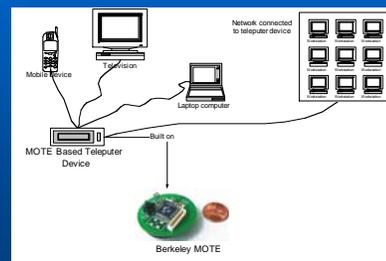
Our Teleputer

Multi-user devices

Teleputer

- **Zero-configuration**
- **Text-free Interface**
- **Sensor-actuator**
- **Cell-phone integrated**
- **Shared Computing**
- **Server-style processing**

Teleputer Sensors

Teleputer Operation

Thank you!

umar@lums.edu.pk

http://www.dritte.org