Internet over Broadband Hybrid Networks

Faculty:	J. Baras, S. Corson, N. Roussopoulos, L. Tassiulas
Research Staff:	S. Papademetriou, M.Y. Liu
Graduate Students:	F. Anjum, V. Bharadwaj, B. Barrett, O. Ercetin, K. Holleman, M. Impett, T. Jen, Y. Jiang, M. Karir, A. Misra, R. Poovedran, P. Ramakrishnan, R. Srinivasan, M. Raissi-Dehkordi, K. Stathatos, M. Stagarescu, A. Tunpan, R. Vaidyanathan
Industry Support:	Hughes Network Systems, Lockheed Martin Global Telecommunications, Bellcore
Industry Interest:	Boeing, Loral, GTE, Lucent Technologies, DirecTV, TCI, MCI, Microsoft, Teledesic-Motorola, Orion
Other Sponsors:	ARL through the ATIRP Consortium, NSA

Industry Advisory Board Meeting
February 17, 1999

Commercial Objectives and Significance

- **Objectives :**
  - Develop and test efficient interconnections of high-data-rate satellites (DBS and future K_a) and broadband terrestrial (wireless and wireline) networks for inexpensive interactive multimedia communications with universal access
  - Develop efficient implementation of TCP/IP over the integrated satellite and terrestrial broadband system including various “last mile” options, especially wireless
  - Develop local improvements to Internet protocols and combine with ATM
  - Develop efficient and dynamic, caching, multicasting, mirroring, multicasting of caches, prefetching algorithms and schemes for satellite supported systems

CSHCN

Commercial Objectives and Significance (cont.)

- Develop efficient algorithms and strategies for the allocation of resources (bandwidth, memory, etc.) and for the dynamic servicing of “push” and “pull” information distribution by the Network Operations Center
- Test and validate these NOC strategies at the University of Maryland OPNET-based simulation testbed of such services, and also using real-life traffic data from industry NOCs

- **Significance :**
  - Satellites are cost effective for broadcast-based information dissemination: Disaster relief efforts, integrated broadband video and interactive data services, distance learning, telemedicine, etc.
  - Combination of satellites with terrestrial wireless and wireline access provides for aggressive and fast prestaging (caching) of vast data sets (several Gbytes) via the broadband satellite for subsequent ultrafast distribution by the local wireless or wireline LAN

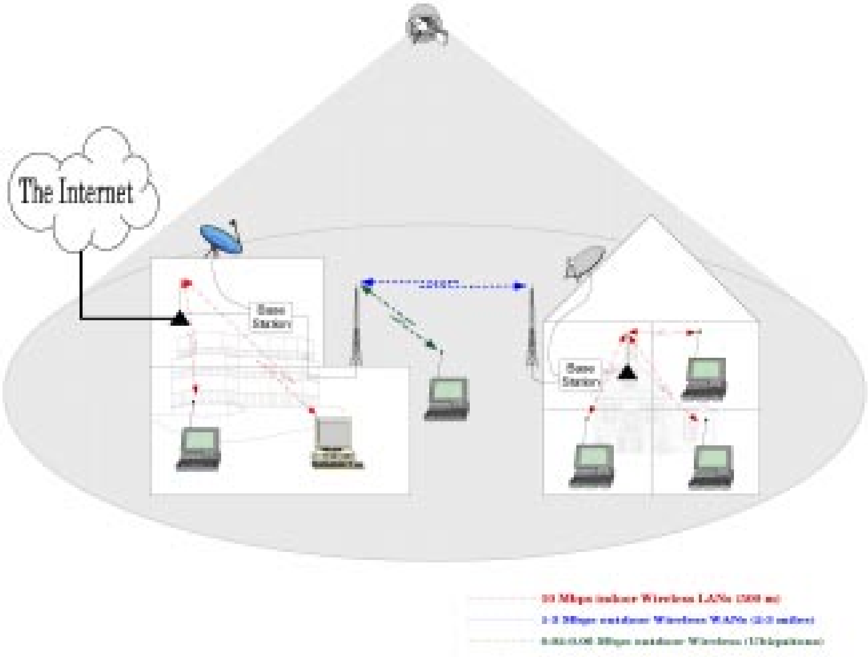
CSHCN

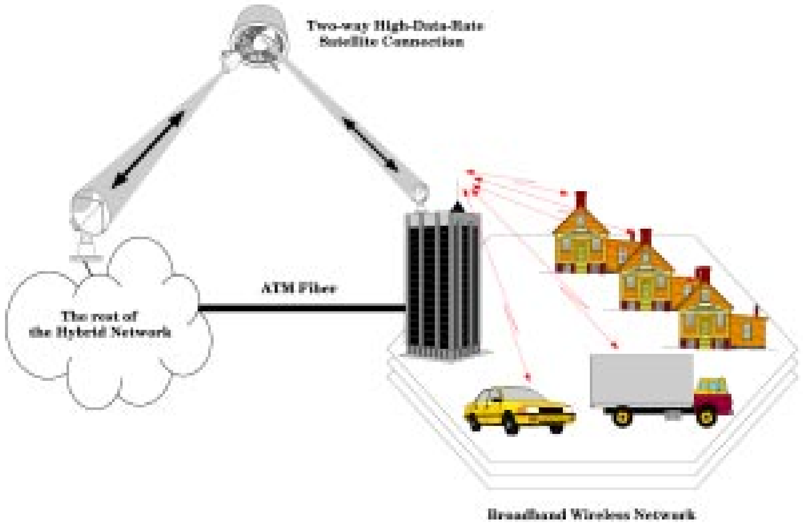
Support for NASA Missions: Objectives and Significance

- **Objectives :**
  - Develop technologies that support the transition of NASA to Internet based communications using commercial satellite networks
  - Develop efficient means to distribute NASA and space data to large number of users using hybrid networks involving broadband satellites
- **Significance :**
  - The technologies addressed here are essential for efficient broadband multimedia communications through satellites

CSHCN

Hybrid Network Architectures: HDR SatCom and Terrestrial Wireless/Wireline Networks

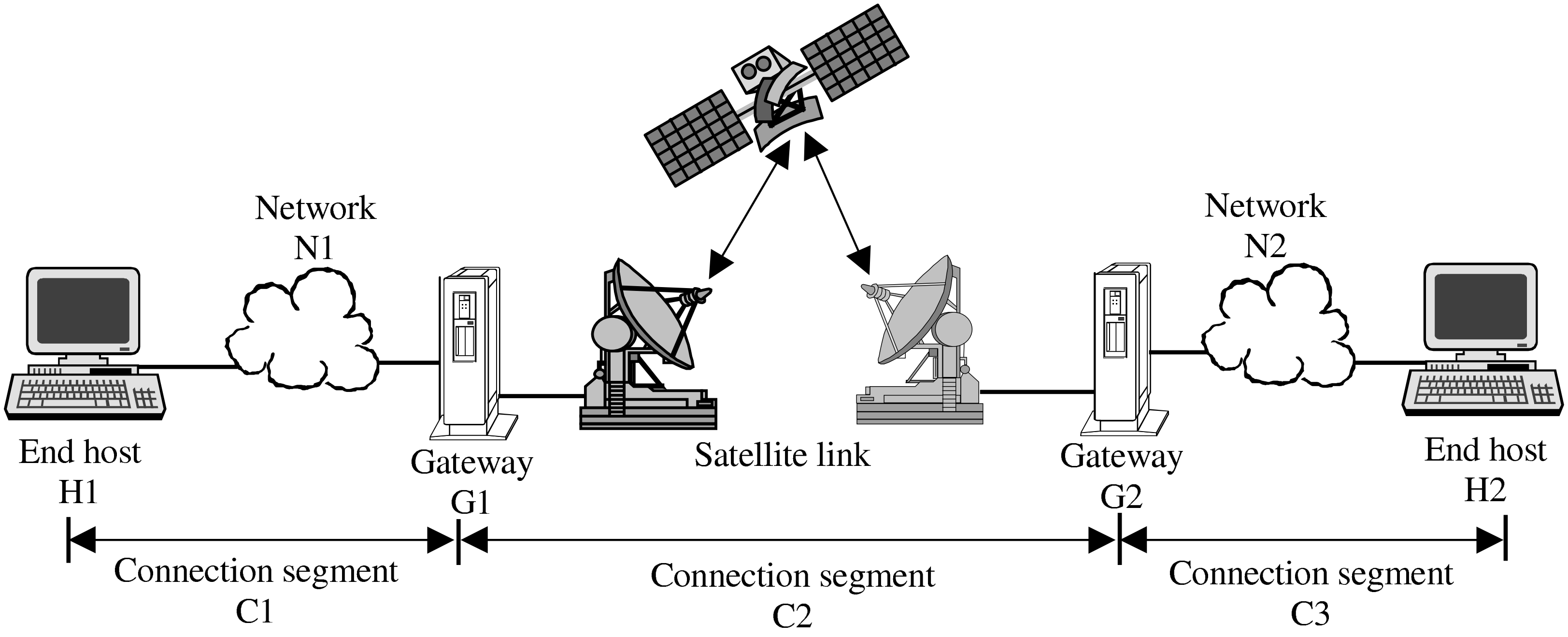

The Internet
Base Station
Base Station
10 Mbps indoor Wireless LANs (100 m)
1-2 Mbps outdoor Wireless WANs (2-3 miles)
0.01-0.06 Mbps outdoor Wireless (Ubiquitous)



DBS and Terrestrial Wireless and Mobile

HDR SatCom and LMDS

Splitting TCP Connections



- Break end-to-end connection into smaller segments
- Used optimized protocol on satellite segment
- Gateways mediate between different protocol stacks

Connection Splitting: Requirements

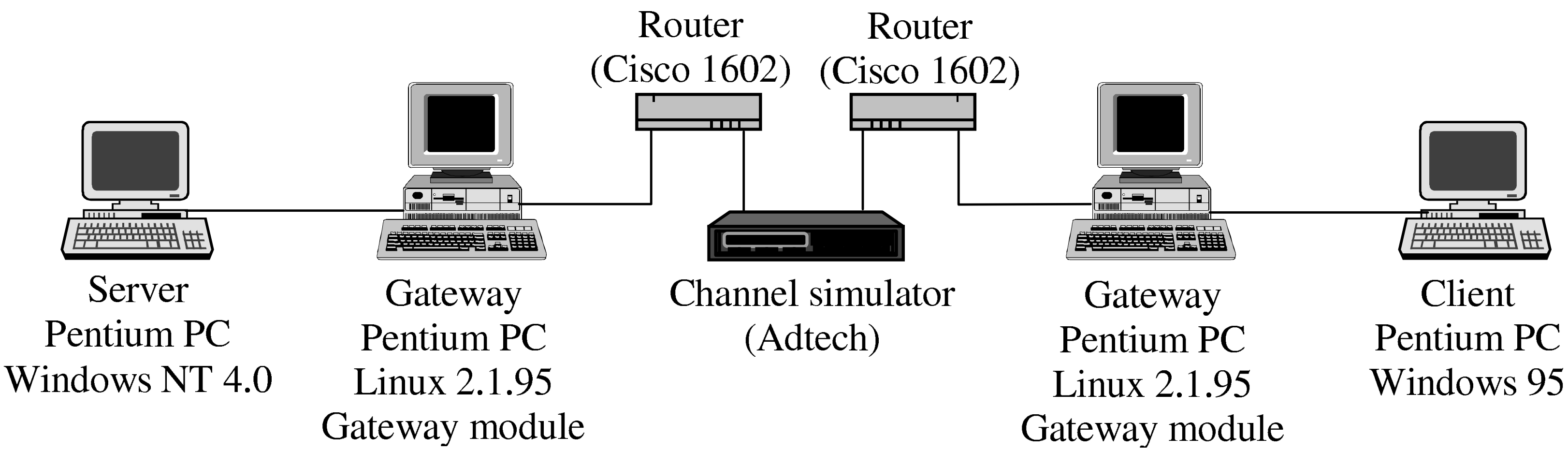
- Maintain end-to-end semantics at application level
- Propagate congestion information across gateways
- Maintain fairness by proper queue management and scheduling
- Minimize resource usage through buffer management policies
- Use flow control mechanisms optimized for satellite links

TCP Enhancements for Satellite Links

- TCP window scaling option (RFC 1323) supports windows as large as 1 GB
- TCP timestamps option (RFC 1323) protects against wrapping of sequence numbers
- TCP SACK option (RFC 2018) along with FACK algorithm corrects multiple losses per round trip
- Larger initial window (RFC 2414) reduces effect of TCP slow start, especially for short HTTP transfers

CSHCN

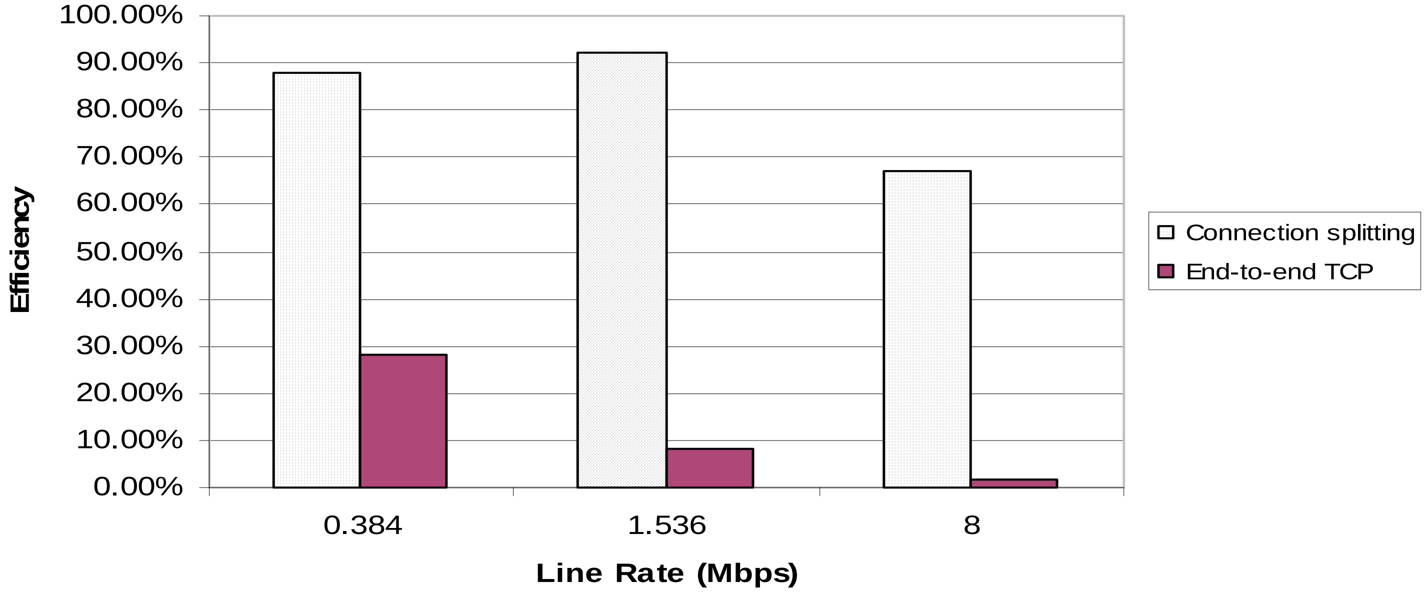
Performance Evaluation: Test Configuration



- Tested with GEO satellite delays
- Data rates from 384 kbps to 8 Mbps
- Error rates from zero to 1E-6
- Gateways in IP router and connection splitting modes

CSHCN

FTP Performance: Effects of Delay



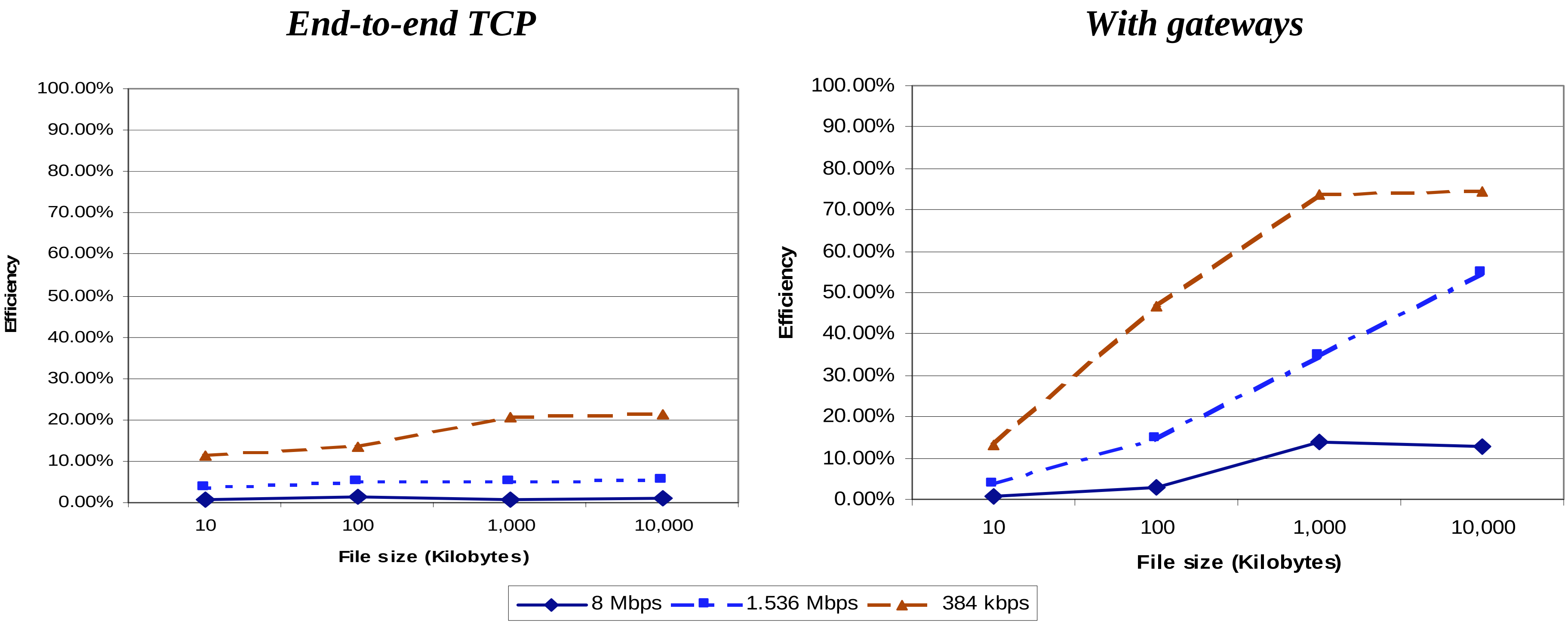
File size 10 megabytes

Link conditions: No bit errors, 250 ms delay

- End-to-end TCP shows window limitation of standard TCP, and percentage utilization falls with increasing line rate

CSHCN

FTP Performance: Delay and Bit Errors

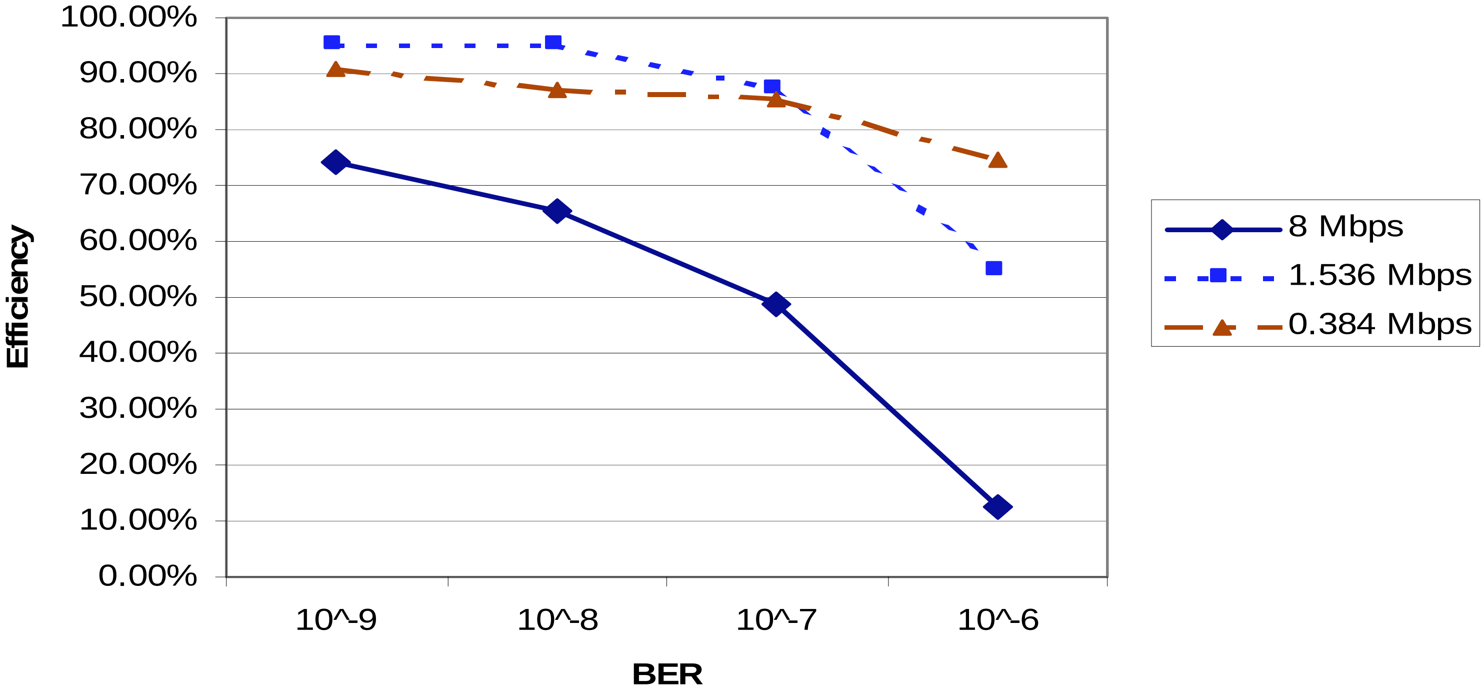


Link conditions: 250 ms delay, BER 1E-6

- Using SACK on satellite link improves performance in presence of errors. However even enhanced TCP degrades sharply with too many errors per round trip

CSHCN

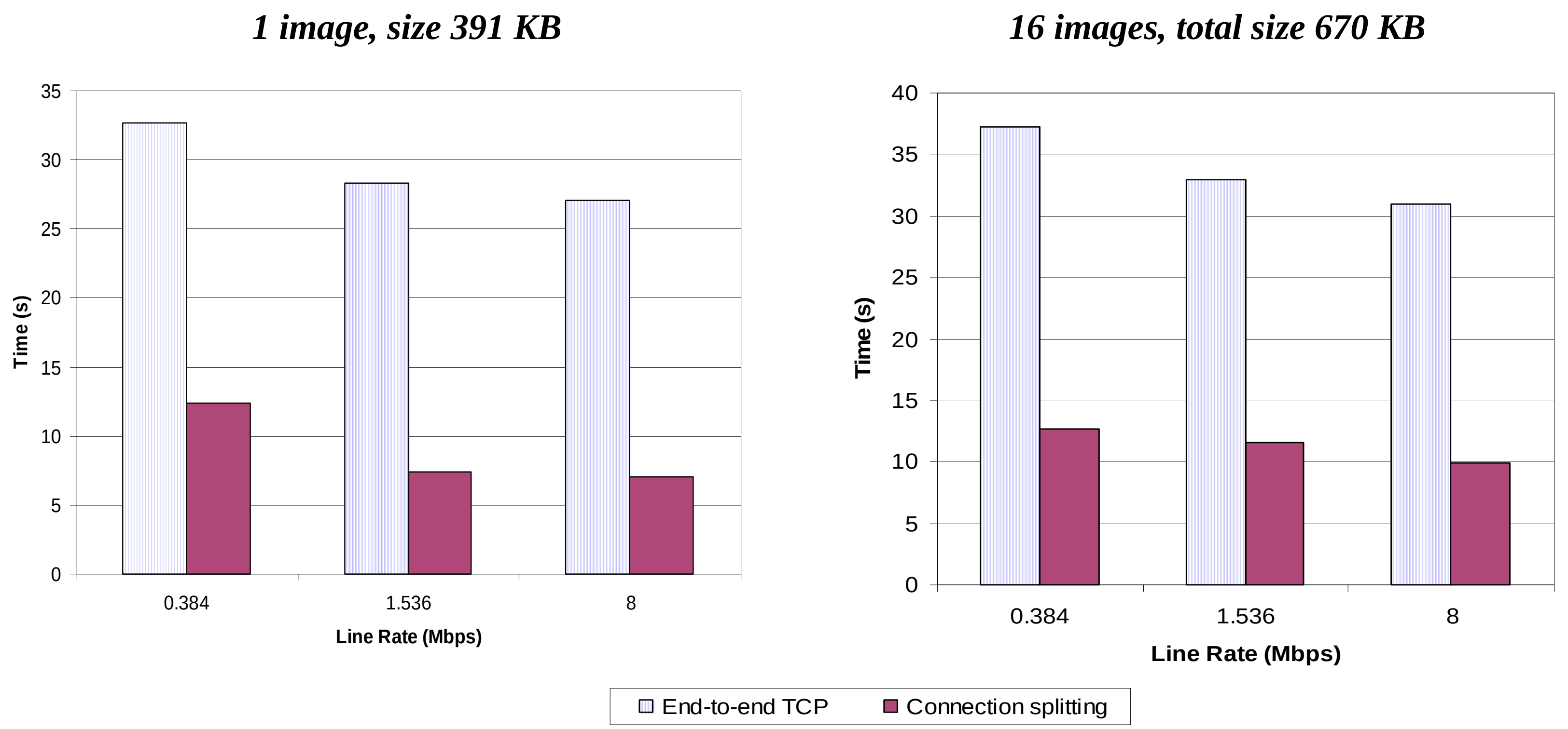
FTP Performance: Gateway with Bit Errors



File size 10 megabytes
Link delay 250 ms

- Gateway performance drops when BER ~ 1 error per round trip, as retransmissions begin to get lost

HTTP Performance: Effects on Transfer Times



1 image, size 391 KB

16 images, total size 670 KB

Link conditions: No bit errors, 250 ms delay

- Much HTTP delay is due to the request-response mechanism. Smarter HTTP should do aggregation

TCP/IP over ATM over Satellite Architecture

- **Gateway acts as multiplexer**
- **Signaling is between the ground stations ATM switches**
- **Choices for the GW to map IP flows to VCCs:**
  - Multiplex all flows into one VCC
  - Identify end-hosts IP pair addresses and multiplex these flows into one VCC
  - Establish one VCC between each transport layer entity
  - One VCC/application flow (i.e. WWW, FTP, Telnet, etc.)
- **Recommend hybrid of fixed number of medium sized connections, then add new dynamically as per demand**
- **Guaranteed rate UBR**
  - Requires no additional signaling requirements or standards changes
  - Will benefit TCP congestion control (no starvation)

CSHCN

TCP/IP over ATM over Satellite Architecture (cont.)

- **Buffer management very important**
  - Simple FIFO buffering with tail drop results in excessive waste of bandwidth
  - Also causes TCP source synchronization
  - Selective Drop and Fair Allocation are more bandwidth efficient
  - Fairness issues: Selective Drop and Fair allocation satisfactory
- **Possibilities: per VC queue management, RED, FRED**
  - Must examine and consider performance vs complexity trade-off;

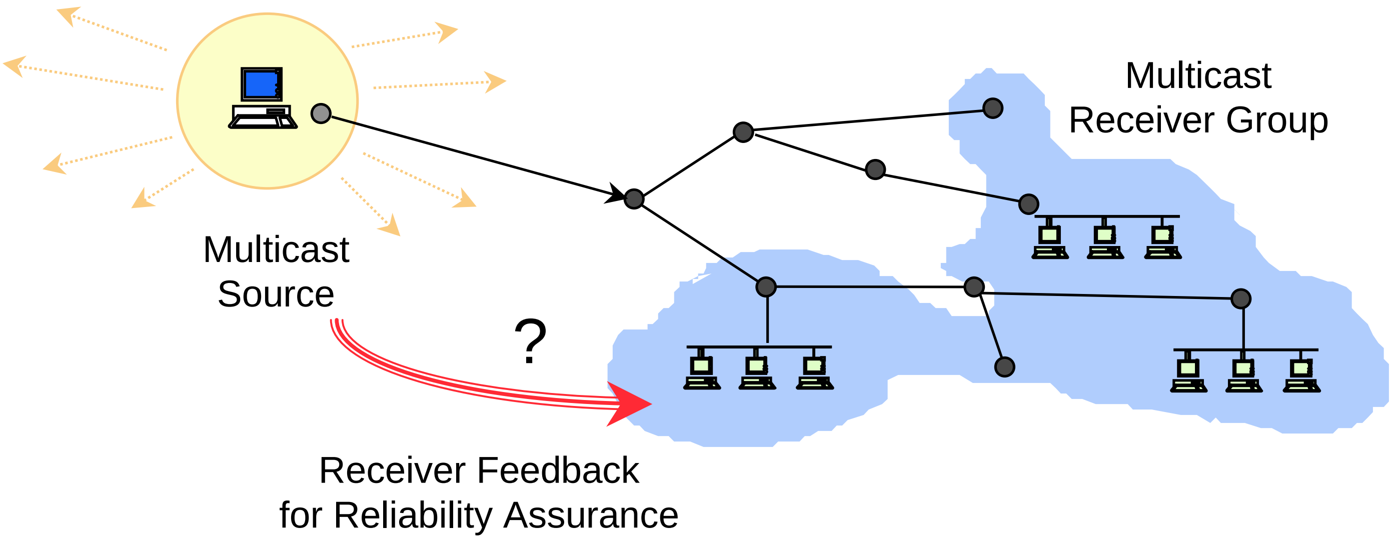
    Per VC queue management is complex while RED is simpler
- **Aggregate traffic characteristics; different for large number of TCP flows from single (self-similar)**
- **Effects of satellite delay on ATM signaling**
- **Effects of other traffic**

CSHCN

CSHCN Internet over Satellite Simulation Testbed

- Completed First Phase of Internet simulation testbed OPNET-based
  - TCP Reno, Tahoe
  - TCP SACK and FACK
  - RED, FRED Queue Management
  - TCP Spoofing/Connection Splitting
  - Hybrid Internet (entire system)
  - Timestamps
  - TCP scaled window option
  - MMPP/MMBP traffic model; Self-similar traffic generator
  - Classical IP over ATM with the gateway
- Developing extended simulation testbed with hardware-based channel simulators
- Developing several regional and national experiments for performance evaluation
- Combination of analytical and simulation techniques
  - Control approach to Internet flows
  - Emphasis on preserving the “strength” of Internet: simplicity, end-to-end, fairness

Reliable Multicast Transport



- IP Multicast delivery does not guarantee reliability (e.g., loss, errors, congestion)
- Reliability can be provided at higher layers
- Generally requires receiver group feedback (to source or other pseudo-sources) to ensure reliable delivery
- Improving the *efficiency* of this process is of interest

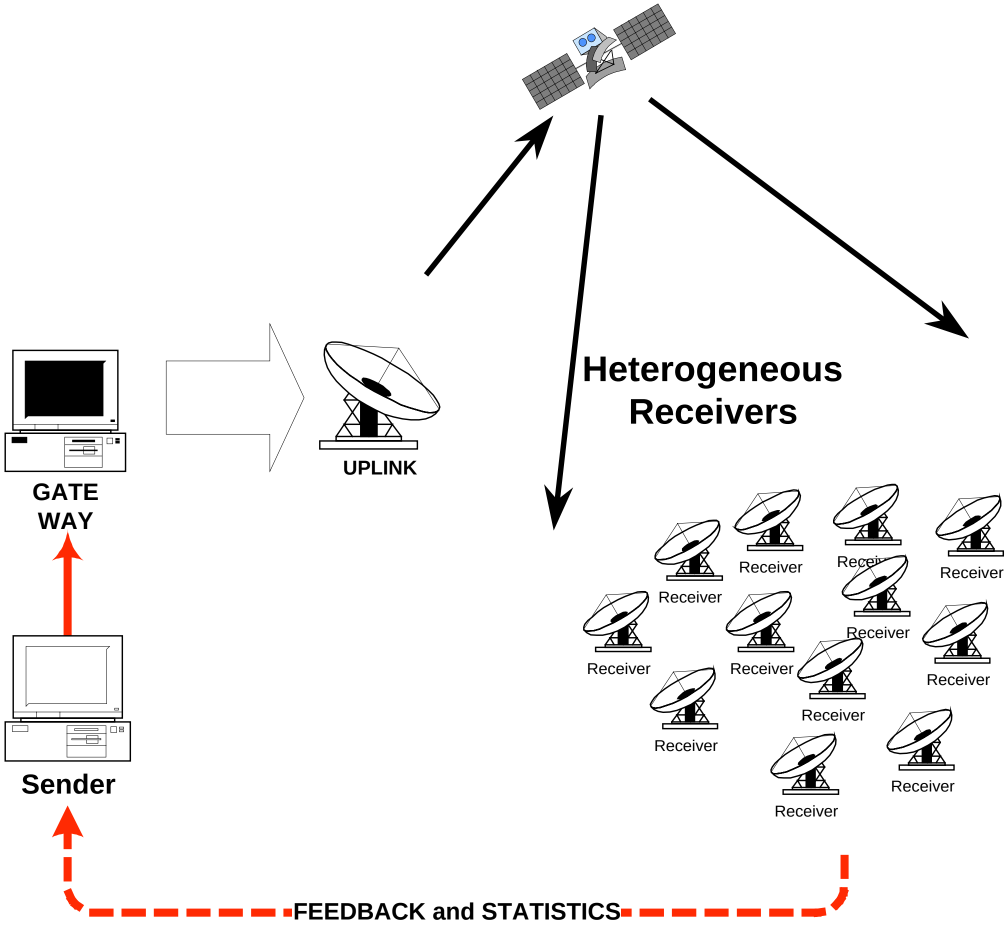
A Role for Information Coding

- **Multicast transport provides efficient group communications**
- **FEC techniques have not typically been applied to internetwork transport layers**
- **For multicast applications, FEC approaches are finding renewed interest**
- **Some reasons include:**
  - undue retransmission is expensive to a large group
  - group packet loss can be high while local receiver loss is low
  - FEC can improve end-to-end delay and robustness for real-time streaming applications
  - large receiver group asymmetric distribution systems (e.g., DBS, hybrid cable, LMDS)

Open problems in Reliable Multicast Research:

- Flow/Congestion Control (short time scale)
- Rate Control (longer time scale)
  - File Delivery
  - Object Stream Delivery

Problem: Rate Scheduling for File Delivery



Problem:

Find optimal sequence of sending rates

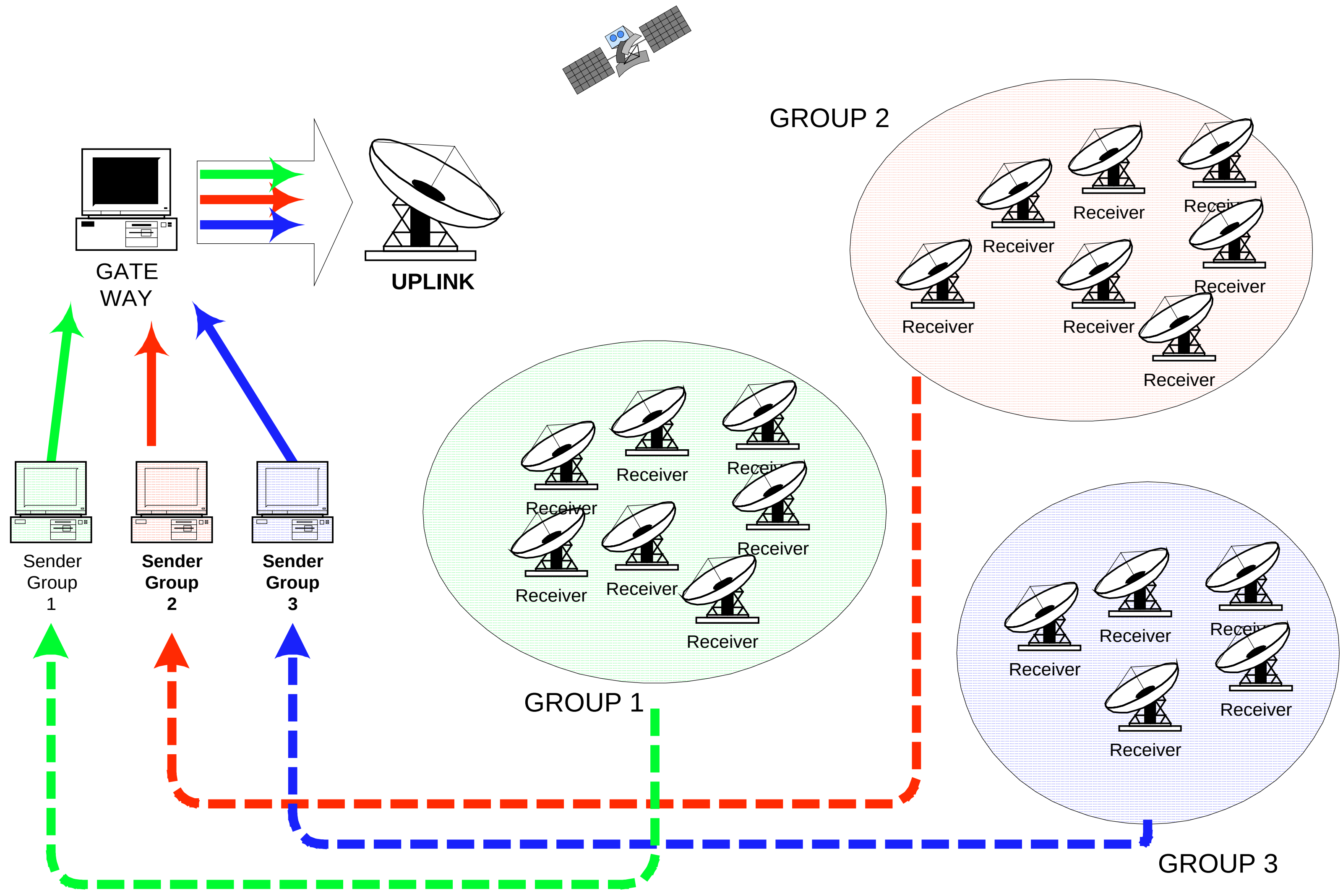
Conflicting Goals:

* minimize bandwidth

* minimize latency

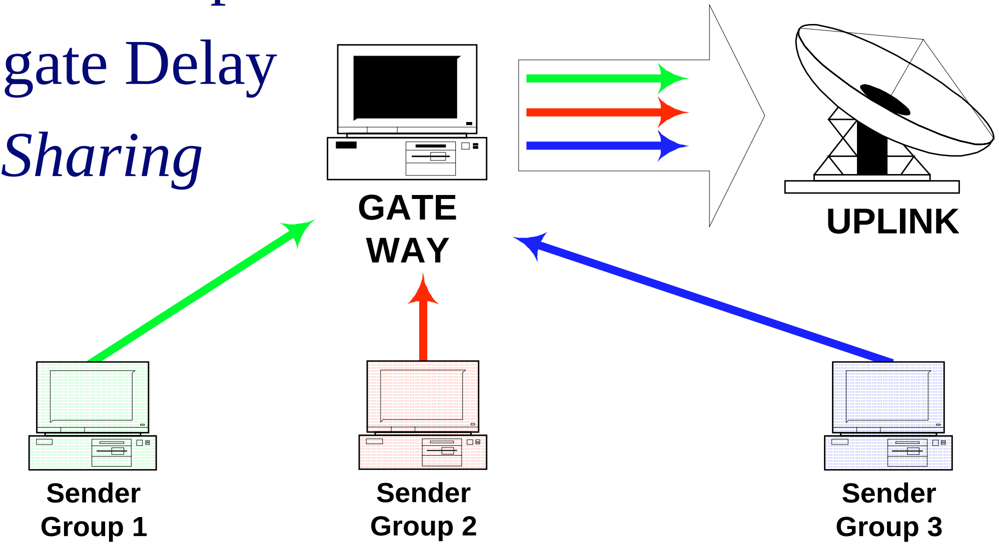
CSHCN

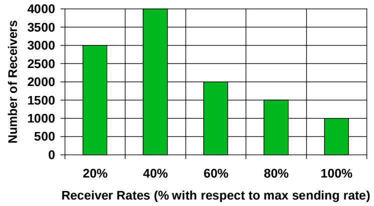
Problem: Rate Scheduling for Multiple Groups



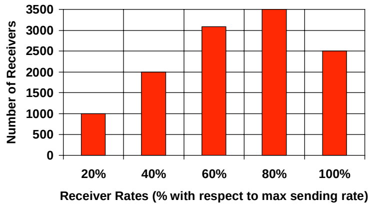
Rate Scheduling for Multiple Multicast Groups

- Minimize Bandwidth Expenditure
- Minimize Aggregate Delay
- *Fair Bandwidth Sharing*

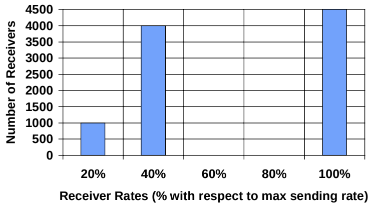




Receivers’ Rate Distribution Group 1



Receivers’ Rate Distribution Group 2



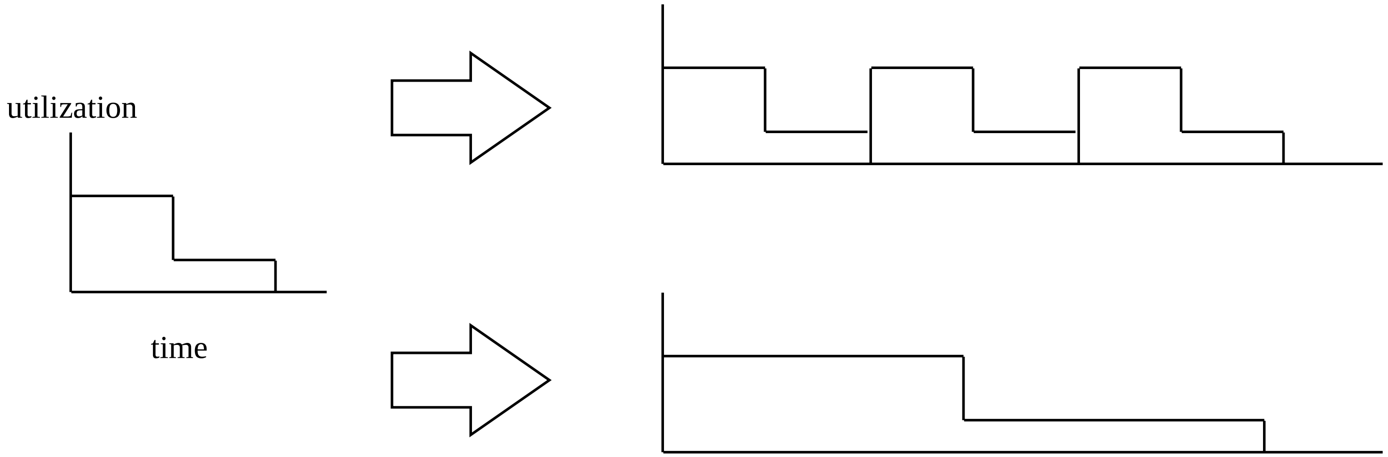
Receivers’ Rate Distribution Group 3

Multiple Groups (cont.)

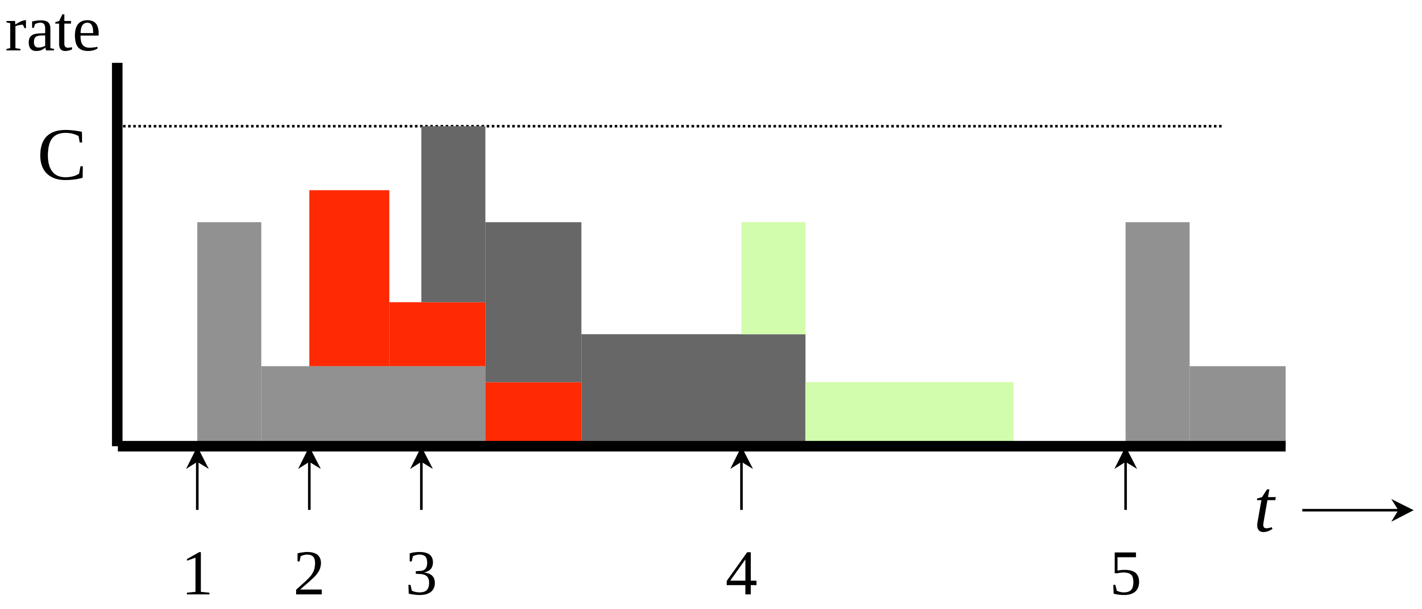
- **Key Issues:**
  - How to distribute computations among the senders and the gateway.
  - Relationships between group's buffer sizes (at senders/gateway), and receivers' population.
  - Optimal vs. Approximated Solutions

Extendable Rate Schedule

- **Idea: Compute a short rate schedule, extend it to a longer rate schedule.**
  - Short rate schedules are easier to compute.
  - Bound the longer schedule's cost in terms of the short schedule's



Scheduling Multiple Files



- Files variable size, interleaved delivery
- Rate schedule recomputed at file transmission start times

Current Issues Being Addressed

- **Simulation and Empirical Study of the Rate Control Behaviors**
  – Responsiveness to long-term network dynamics.
  – Schedule re-computation interval.
  – Stability.

- **Distributed Algorithm Solution**
  – how to converge to optimal schedule faster.

- **On-Demand Scheduling (upon file arrivals)**

Development of a Multicast Security Framework Using Non-Cryptographic Techniques

- Scalability
- High Integrity
- Heterogeneity
- Approach
- Joint Secret Generation
- Shared Key Algorithm
- Extension to ElGamal keys based on Prime field generation and Elliptic Curves

Development of a Multicast Security Framework Using Non-Cryptographic Techniques

- **Secure Key Management in Multicast Communications is a difficult problem**
  - Framework should be scalable with respect to security-related operations
    - member admission
    - member revocation
    - ACL, CRL updates
  - All the members of a secure group should share the same Traffic Encrypting Key (TEK)
  - It may be desirable to distribute key management authority, yet to do so such that cooperating members act in a *collective* fashion

PROBLEM UNDER STUDY

- **Following scenarios may force a group-level TEK update**
  - Expiration of the lifetime of the TEK
  - Member admission (protection of old traffic)
  - Member departure (protection of future traffic)
  - Member compromise
  - Revocation of Group Membership
- **How to effectively perform TEK updates and provide scalability and integrity in terms of security-related functions?**
- **How to prevent collusion among the revoked members in generating a future Key Encrypting Key (KEK)?**

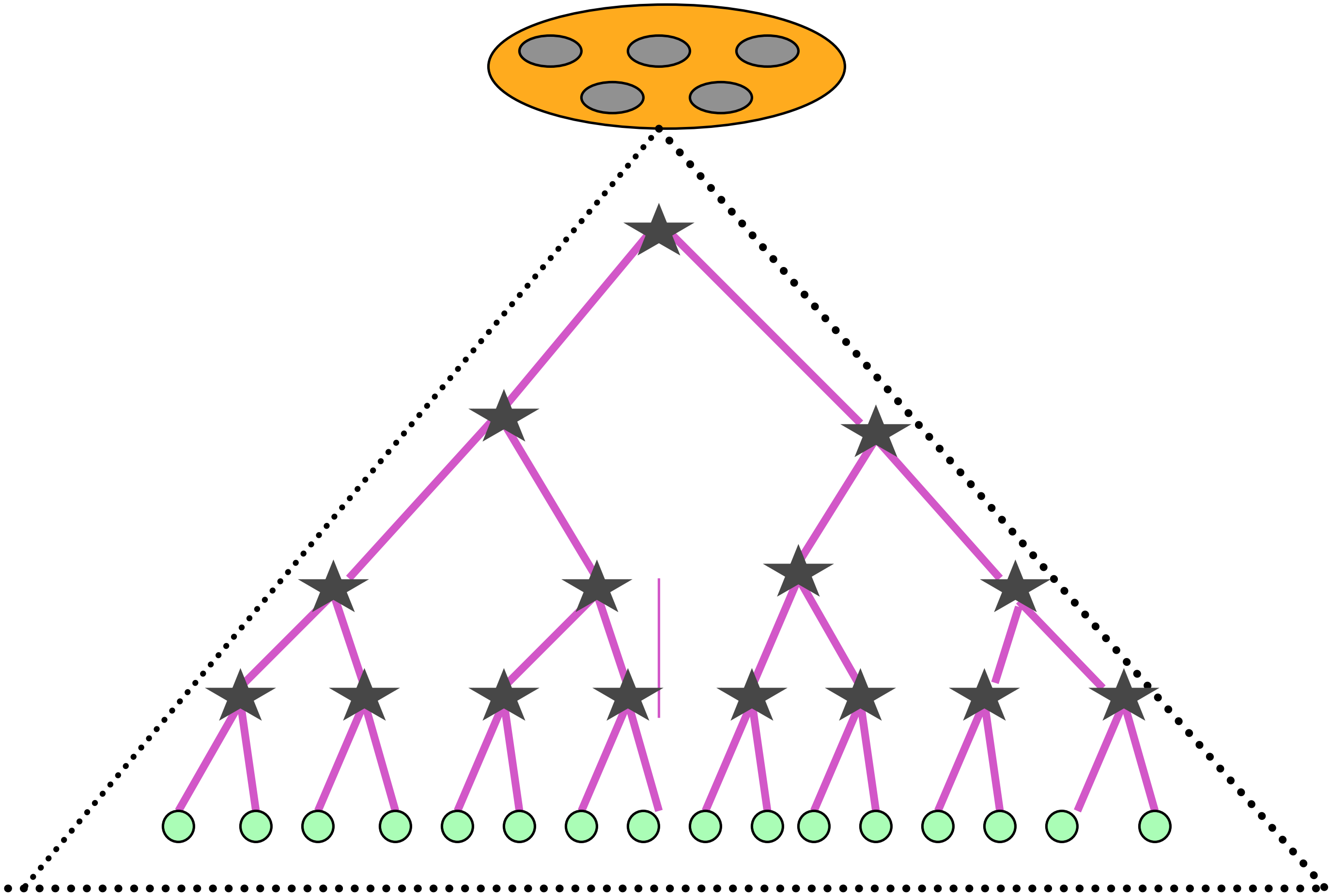
RELATED WORK ON KEY MANAGEMENT

- **GKMP family (CBT etc)**
  - uses a single key encrypting key (KEK) for updates
  - provides a simple key management approach
  - utilizes a single controller
- **Iolus**
  - partitions the group into clusters and provides different KEK for each cluster
  - provides compartmentalization of security
  - utilizes multiple, individual controllers
- **LKT**
  - uses more than one KEK per member
  - provides reduced number of messages for key updates even if a set of members are revoked
  - utilizes a single controller

Our Model

- **Group may be partitioned into clusters as in Iolus**
  - compartmentalization of key updates (a feature addressed differently by LKT)
  - increases scalability--allows heterogeneity
- **Each cluster uses a cluster-level LKT**
  - reduced number of messages for key updates in the event of revocations/compromises
  - reduced storage requirements at each cluster
- **Each cluster is controlled by a cluster panel**
  - adds high-integrity operation

A CLUSTER PANEL WITH LOCAL CLUSTER LKT



Our Model (cont.)

- **Group TEK is generated by a top-level panel**
- **Each Cluster Key is jointly generated by all the members of its respective panel**
  - removes reliance in a single, trusted, intermediate node
- **Cluster panel performs**
  - admission
  - key generation/distribution/revocation
  - updating and broadcasting local ACL, CRL to the rest of the cluster

PANEL BASED JOINT SECRET (TEK) GENERATION APPROACH

- **Computational steps involved for panel size *N***
  - Initialization
  - Individual share generation
  - Applying necessary padding to the individual shares
  - Exchanging padded shares
  - Combining the shares
  - Removing the padding effect and retrieval of the fresh random quantity

APPLICATION OF JOINT SECRET GENERATION FOR PUBLIC KEYS

- Basic joint secret generation can be modified to allow members to generate a dynamic group public key framework as well
- Individual member secret becomes the private keys of the members
- Group binding parameter becomes the group private key
- Group public key is a combination of individual public keys
- Use of joint secret sharing for public key generation
  - Group ElGamal Keys on Zp*
  - Group ElGamal Keys on Elliptic Curves

Current Issues Being Addressed

- **Software implementation of the proposed framework**
  - panel based cluster scheme
  - shared secret generation algorithm
  - group public key schemes

- **Addressing**
  - mobility and transparency
  - Optimal key allocation issues
  - performance characterization in terms of optimal key allocation issues
  - Analysis of the framework

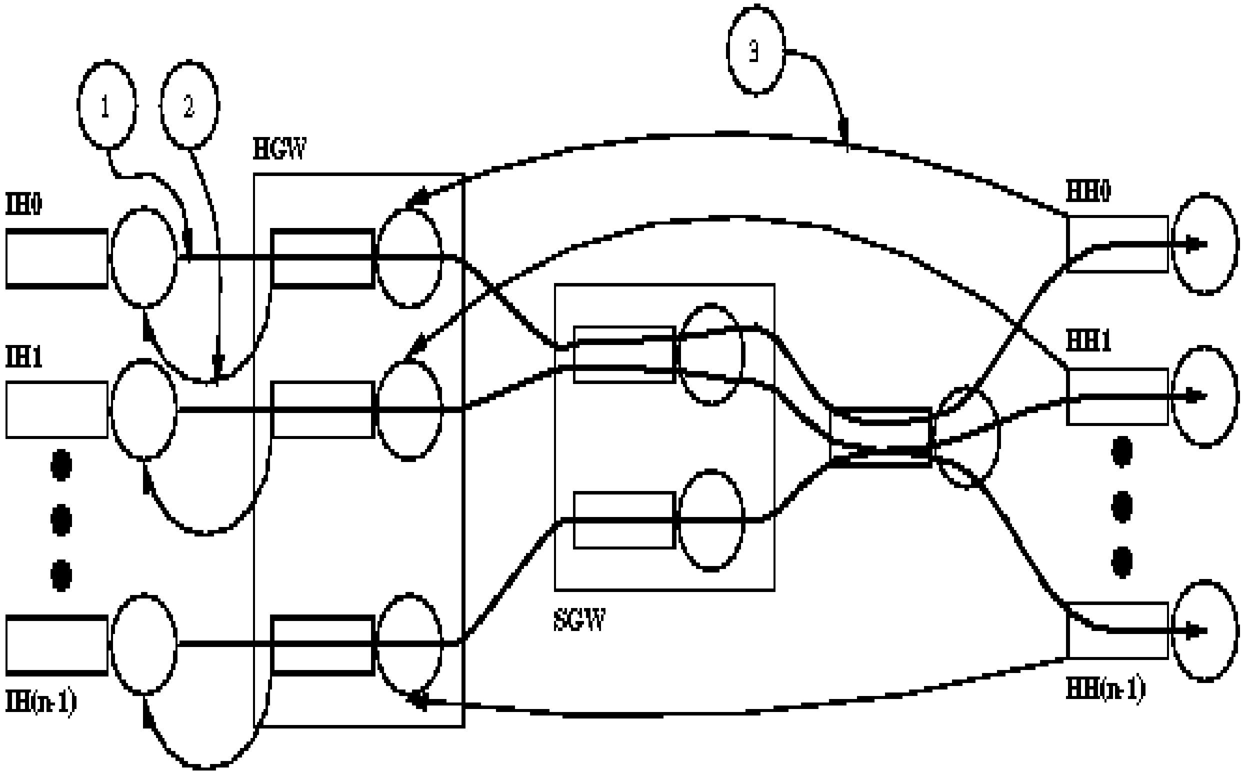
Network Operations Center (NOC) for Hybrid Internet Service

- **Congestion control :** TCP and TCP Spoofing
  Satellite channel bandwidth allocation
- **HGW : first NOC object that receives data (Router)**
  - HGW prioritizes Hybrid Internet traffic
- **SGW jobs : mixture of Internet and exogenous traffic**
  - <u>Exogenous traffic:</u> package delivery and data feed traffic

  SGW maintains four queues : two for package delivery and data fee

  two for the two priority levels of Internet
- **Exogenous traffic high priority : fluctuations in bandwidth allocated to Hybrid Internet**
- **Self-similar traffic: Interactive users as ON-OFF processes**

NOC Queueing Analysis Model



(1) Data connection:
IS sends data to corresponding HH

(2) Acknowledgments:
From HGW to IS

(3) Acknowledgments:
From HH to HGW

NOC Bandwidth Allocation Strategies: Analysis

- Model Hybrid Internet service
- Used self-similar IS traffic models
- Service is FIFO within each connection
- Demonstrated that the fair bandwidth allocation strategy provides smaller average queueing delay than the equal bandwidth allocation strategy
- Demonstrated that the optimal strategy prioritizes bandwidth allocation to the connection with largest queueing delay

Performance of NOC Schemes

- **Traffic Analysis**
  - Modeled interactive users as ON-OFF processes with Pareto idle and busy periods
  - Obtained analytical estimates of loss probability
  - Obtained optimal bandwidth allocation strategy (Most Delayed Queue Served First)
- **Analytical models and simulation used for Network Dimensioning: Estimates of No. of sources that can be in the system at the same time**
- **Extensive simulation experiments.**
  - Computed and compared: connection state, queue length, demand, bandwidth, delay, # Acked packets, # Unacked packets

Buffer per Connection	500 packets
Total Bandwidth	15 packets/unit time
Number of Connections	5 connections
Constant Arrival Rate	10 packets/unit time
Mean of the Uniform Arrival Rate	5 packets/unit time
Delay Imposed to Queued Packets	0.1 unit time

Common Input Data

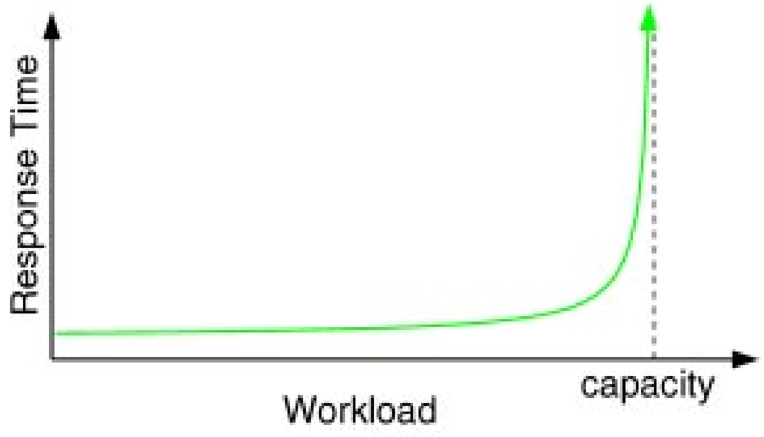
Conn1:	1.4469	1.4468	0.0
Conn2:	2.0720	2.0720	0.5298
Conn3:	1.6941	1.6689	0.204
Conn4:	2.0541	2.0524	0.0741
Conn5:	1.7182	1.7088	0.8847
	EB	FB	MDQSF

Average Delays

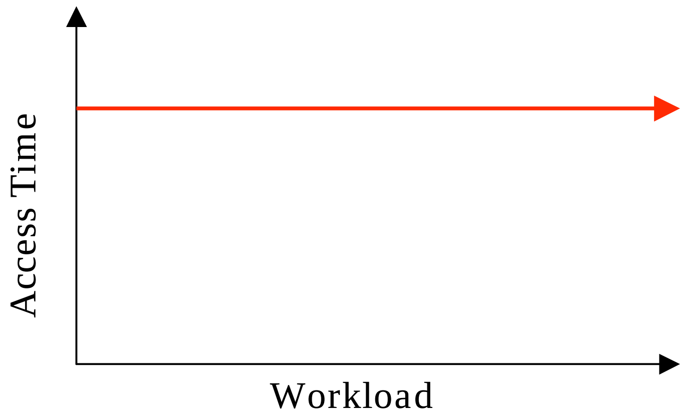
Research Problems Being Addressed

- Traffic Models and Validation
  - Development of validated traffic demand models for “push” and “pull” demand (validated for accuracy against real-life data from industry NOCs)
- Resource Allocation and Service Scheduling
  - Development of integrated resource allocation and service scheduling strategies
- Adaptive Hybrid Data Delivery
- Distributed Layered Caching
  - Development of distributed, layered caching architectures for improved delivery performance
- Network Planning, Design and Dimensioning
  - Development of quantitative methods and recommendations for network planning, design and dimensioning
- System Integration, Testing and Performance Evaluation
  - Performance evaluation and testing of the entire management and operation scheme for Hybrid Broadband Internet services

Data Pull vs Data Push
(unicast vs broadcast)

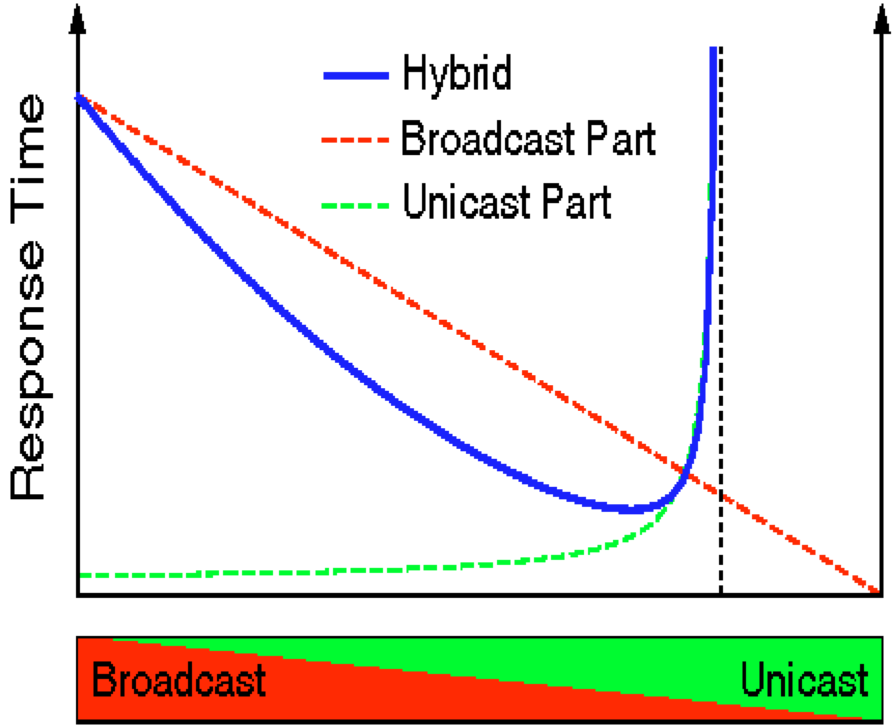


- ✔ “Simple”
- ✘ Constraint: Load < Capacity
  - Larger scale ⇒ More resources
- ✘ Redundant work & traffic for popular data



- ✔ Unlimited Scalability
- ✔ Good for asymmetric applications & popular data
- ✘ Sequential access, latency depends on broadcast data
- ✘ No user requests, good “guessing” required

Balancing Data Push & Pull Delivery

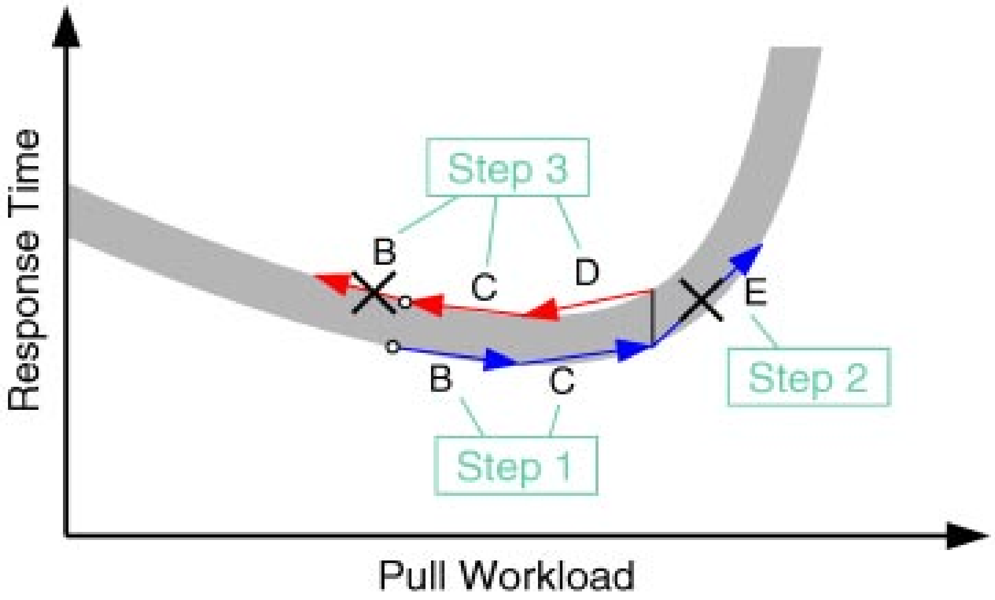


- **Problem description :**
  - Client requests for individual data objects (pages)
  - Skewed access pattern with changing hot-spots
  - Workload >> Pull Capacity
  - Flat broadcast

CSHCN

Adaptive Algorithm

Start: ... ≤ A ≤ B ≤ C ≤ D ≤ E ≤ ...
Step 1: ... ≤ A ≤ B ≤ C ≤ D ≤ E ≤ ...
Step 2: ... ≤ A ≤ B ≤ C ≤ D ≤ E ≤ ...
Step 3: ... ≤ A ≤ B ≤ C ≤ D ≤ E ≤ ...



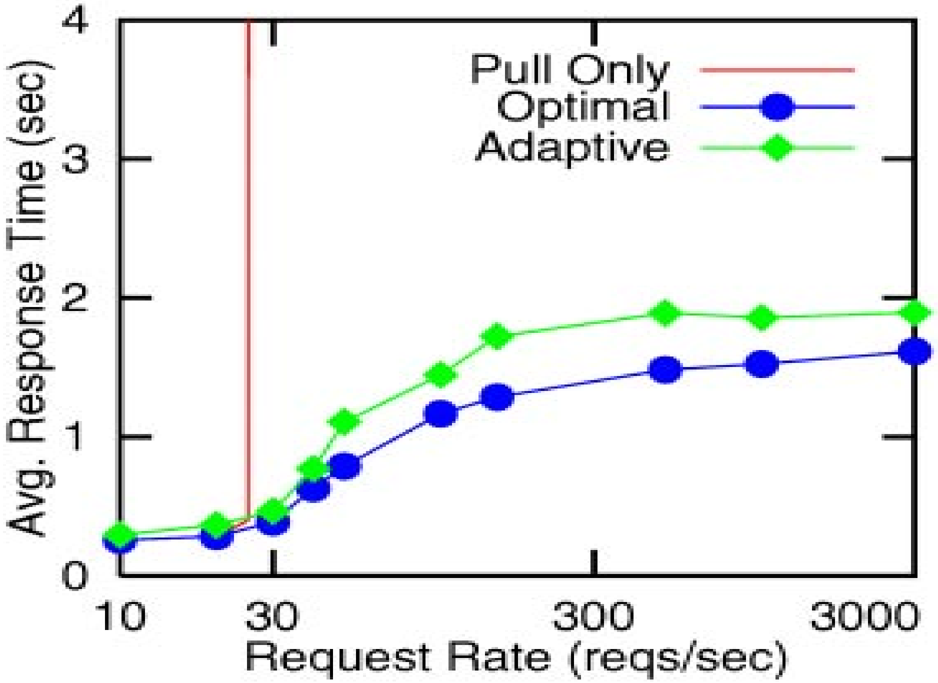
- **Adapt data states every broadcast cycle**
- **Adaptation based on expected performance marginal gain**

1. Demote all vapor items colder than the hottest liquid item
2. Demote more vapor items while there is significant gain
3. Promote liquid items while there is not significant loss

CSHCN

Experiment with Static Workload

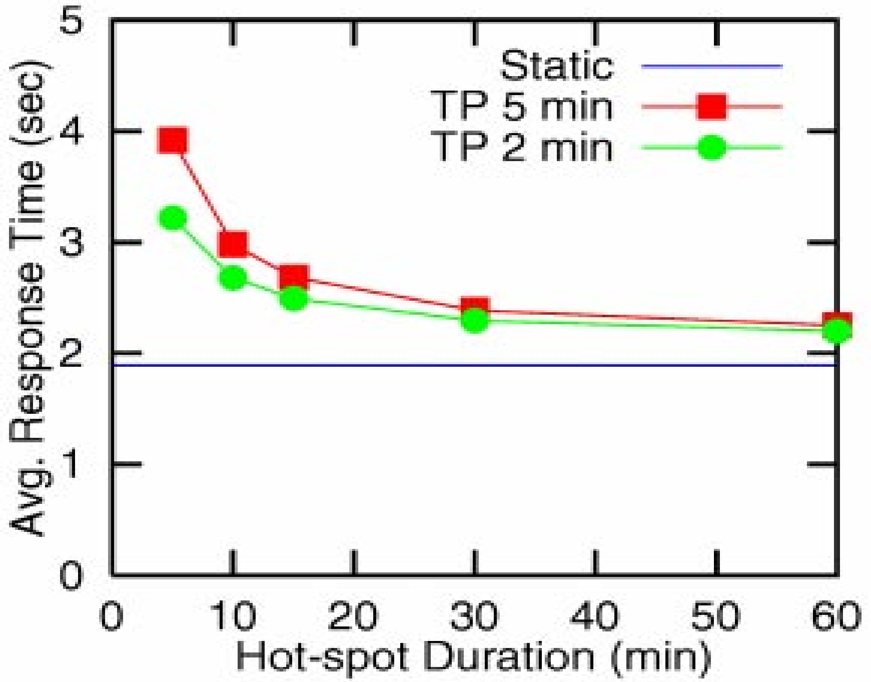
Broadcast Rate	12 Mbps
Downlink Rate	12 Mbps
Page Size	50 KB
Database Size	10000 Items
Hot-spot Size	100 Items
Access Distribution	Gaussian
Pull Capacity	**≈30 Reqs/sec**



- ✔ **Scale ≥ 100 × Pull only syst**
- ✔ **Performance depends on hot-spot size**
- ✔ **(Almost) Perfect detection of hot-spot**

Experiment with Dynamic Workload

- New hot-spot every *Duration* min
- Transition period of *TP* min
- Gaussian distribution
- Workload ≈ 20 x Pull Capacity



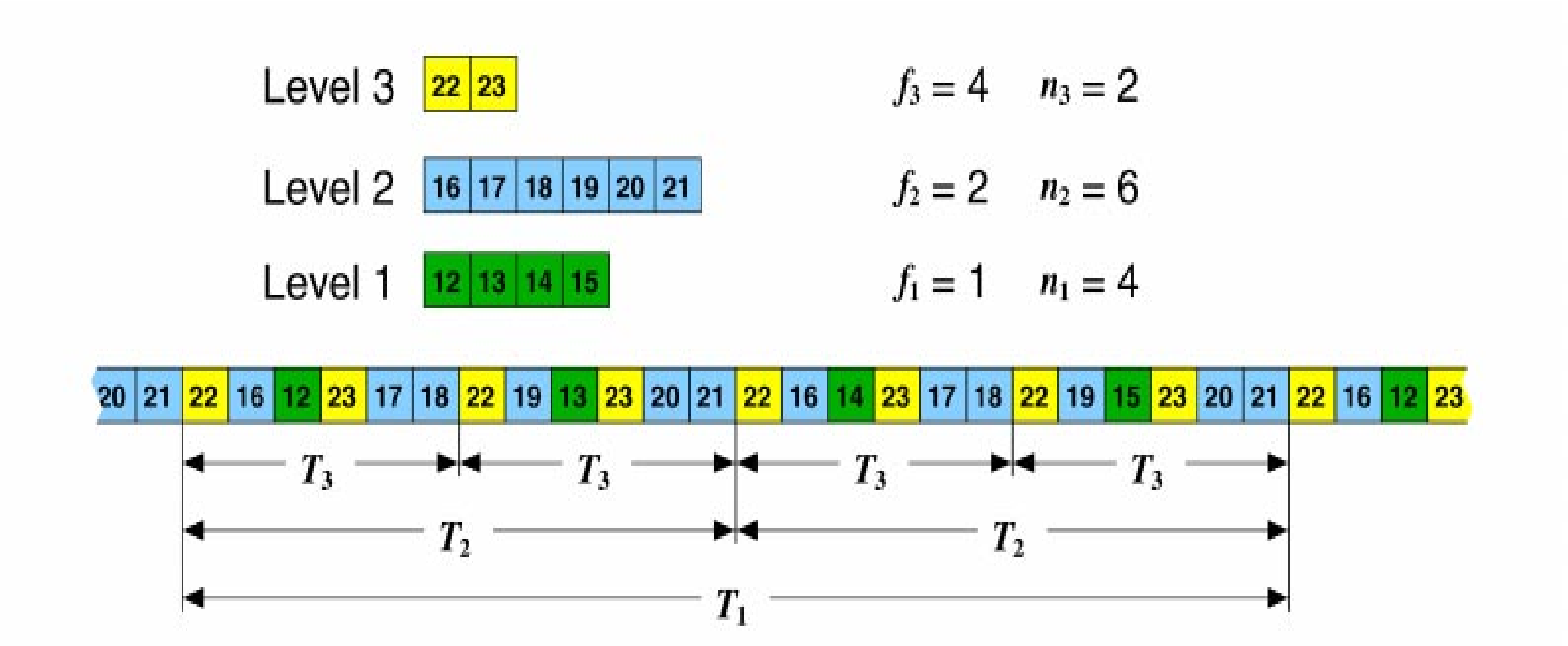
✔ **Effective detection of fast changing hot-spots**

- **Double hot-spot during TP ⇒ More broadcast**

Disseminating Logged Updates to thousands of clients

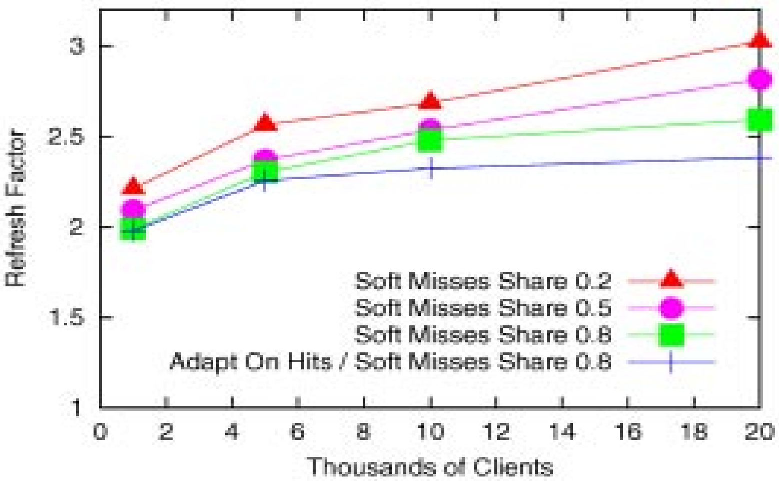
- **Mobile clients**
- **Disconnect operation**
  - sleep mode: clients do not listen
  - wake-up mode: must refresh cache
  - different refresh sizes (# of log items)
  - log item popularity decreases with age
- **Examples:**
  - Database clients (e.g. mobile workforces, sales agents)
  - Web updates, news updates
  - Software upgrades & patches
  - NASA data updates, especially from high volume experiments

Multi-Level Air-Caching



- **Fresh log pages at higher levels**
- **Problems:**
  - Decide how many log pages to air-cache
  - Decide how to partition them in different levels

Experiment with Expanding Log



- 20 new pages between adaptations (exponentially distributed)
- Clients: Half sleepers, half workaholics
- Hard and soft misses
- ✔ Effective adaptation
- ✔ Scalable dissemination of updates
- More soft misses ⇒ Better workload estimation

Current Issues Being Adressed

- **Implementation**
  - Center for Satellite and Hybrid Communication Networks
- **Air-Caching for more applications, e.g.**
  - Dissemination of database views and query results
  - Publish/subscribe applications, data channels
  - Multimedia
  - Real time applications
  - Distribution of NASA space data
- **Three-tier architectures**
  - Broadcast data pumps for proxy caches

We deliver the right data at the right place and the right moment