University of Glasgow | School of Computing Science

I Want My Internet TV

Understanding IPTV Performance in Residential Broadband Environments

Colin Perkins

What is Internet TV?

- Television service delivered using an Internet connection, rather than using a dedicated TV distribution network
    - Replacing the TV distribution network with an IP-based infrastructure
    - Service provided directly by a TV network, by the Internet Service Provider (ISP), or by a third party
    - Viewed using a dedicated device, or a computer, smartphone, etc.

Internet TV Performance

- Quality of Internet TV can be highly variable
- The network is shared with other traffic and is not optimised for TV distribution
  - Capacity is not guaranteed
  - Last-mile residential link behaviour is not well understood

Capacity is not Guaranteed

- Best effort packet delivery service
- Links are physically shared:
  - Wide-area network shared with other customers
    - ISPs oversell capacity – network cannot support full rate for all customers
    - Network congestion likely at peak times of day
  - Last-mile link shared with other residential users
    - Last mile link shared with other traffic to/from your home
    - Web browsing, file sharing, gaming, etc.
- Variable rate transport protocols ubiquitous
  - Most applications use rate-adaptive transport, with no fixed sending rate
- Active queue management – to separate different traffic – not widely deployed

Last-mile Link Behaviour

- Highly variable infrastructure
  - ADSL, Cable, Fibre-to-the-home, Fibre-to-the-Cabinet, etc.
  - Ethernet and/or Wi-Fi in the home
  - Home gateway equipment
  - Variable end-system hardware and operating systems

- Behaviour not well understood
  - Packet loss characteristics – noisy, poor quality lines; how to model loss?
  - Buffer bloat – excess delay impacts control loop stability, but many ADSL modems are observed that can buffer multiple seconds of traffic

Why Deploy Internet TV?

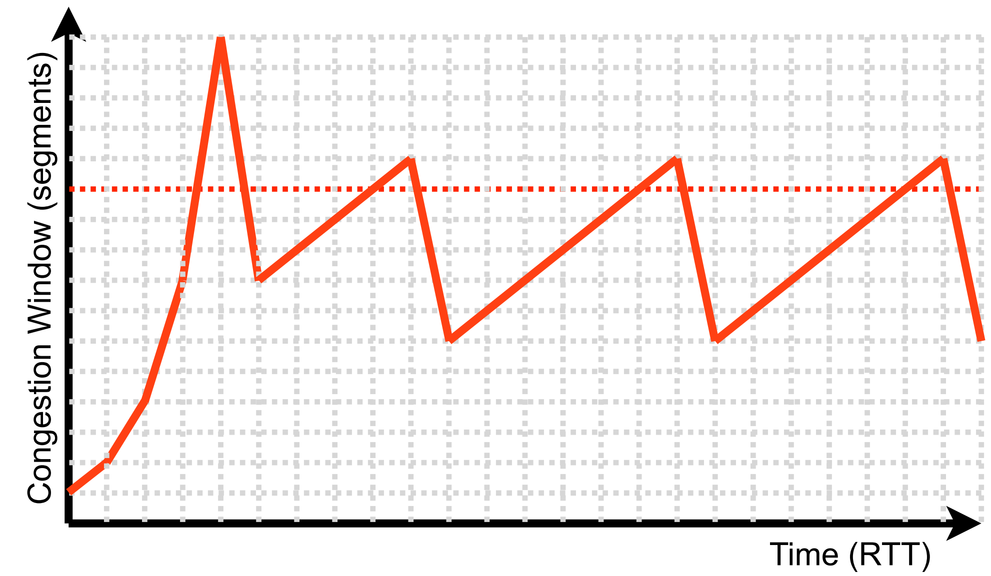
- Convergence: one network for TV, voice, and data
  - Economics of scale from combining formerly separate businesses
  - Shared infrastructure → cost savings, easier to manage

Approaches to Delivering Internet TV

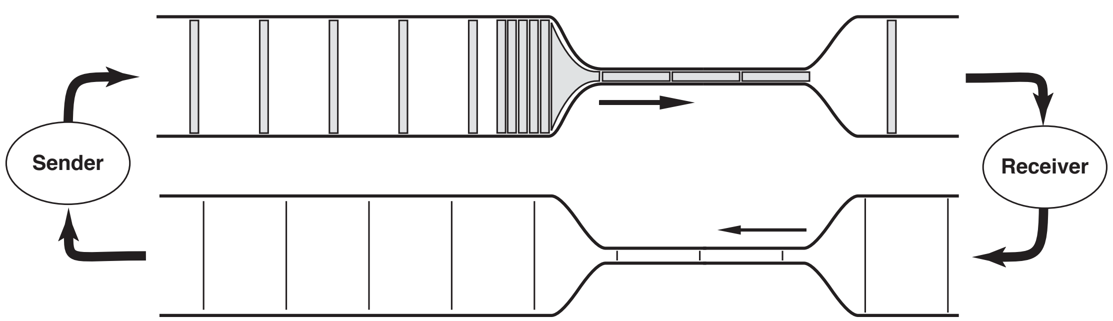
- On-demand and catch-up TV
  - For watching a single programme using a web browser, or a dedicated application/appliance
  - Choose a programme from a menu, sit back, and watch to completion
- Linear and live TV
  - Traditional TV service – choose from multiple channels
  - Content may be live or pre-recorded, but it is streamed continually, according to some schedule
  - Channel surfing commonplace
- Different constraints imply different implementation choices

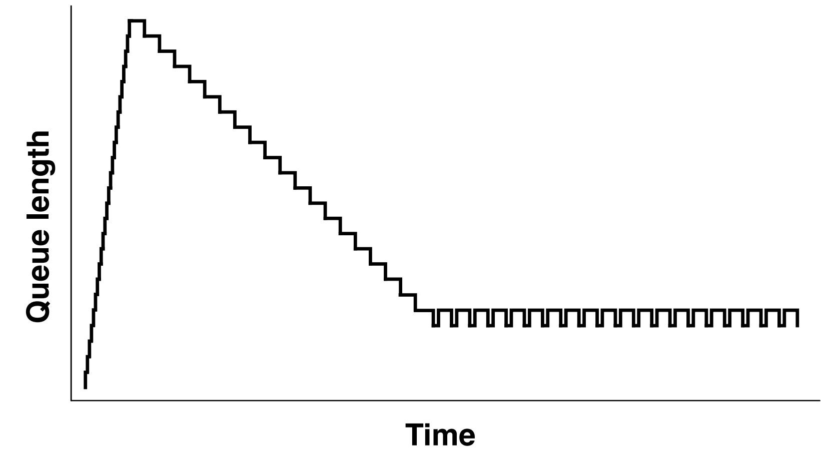
On-demand and Catch-up TV

- Start-up latency is not critical, but it is essential that playout is smooth and continuous once started
  - Sit down to watch a...buffering…movie
- Build on web infrastructure – media downloaded using HTTP on TCP/IP



- TCP causes in-network queues
  - Routers queue for outgoing data for a link
  - TCP will always probe for more capacity by increasing sending rate
  - Sending faster than link rate causes queues to build up – TCP dynamics ensures this *always* occurs to some extent
  - Queues smooth output rate from bottleneck link

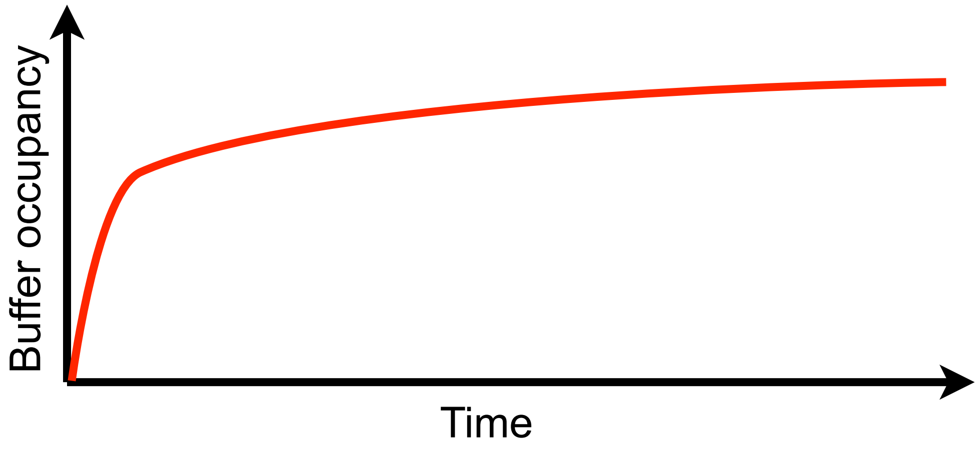




Source: Van Jacobson, IETF 84

Dynamic Adaptive Streaming over HTTP

- DASH download model:
  - Split each TV programme into a sequence of short, independently decodable, chunks
  - Encode each chunk at multiple bit rates (i.e., multiple quality levels)
  - Client measures download rate of each chunk, selects bit rate of the next chunk to match
  - All chunks fetched over HTTP, typically driven by Javascript code in a web browser, or a dedicated player app



- Initially fetch low-rate chunks to build up buffer, switch to higher-rate/quality once client buffer stabilised (~tens of seconds)
- Use in-network queues to hide transmission variability, give smooth playout
  - Excess buffering ("buffer bloat") can give multi-second queues

Dynamic Adaptive Streaming over HTTP

- Numerous open questions:
  - What is an appropriate chunk duration?
  - What chunk data rates should be chosen? How should they be spaced?
  - Use a single connection, or a new connection for each chunk?
  - Where is the intelligence to pick next chunk rate? Client or server?
  - What is the effect of TCP parameters (e.g., Google's IW=10 proposal)?
  - What is the impact on user experience of variable rate encoding?
- Lots of scope for experimentation – no need for more standardisation
- Optimality not critical
  - Enough buffering in the network that good enough is sufficient

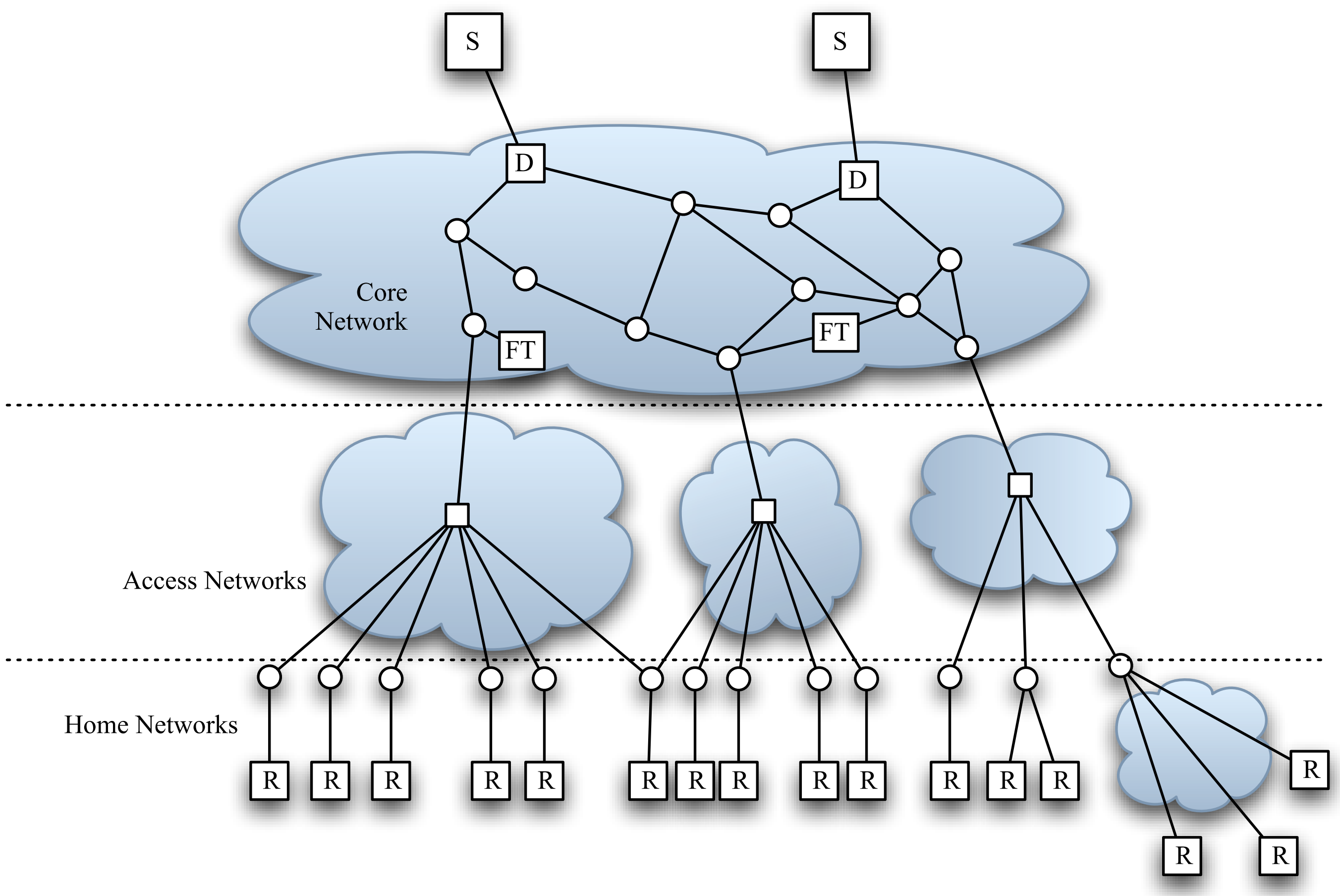
IPTV: Linear and Live TV

- DASH model doesn't work for systems that mimic traditional TV model
  - Relying on buffering to smooth over transport variation – latency – gives very slow channel change (re-buffering… ~tens of seconds!)
  - Also fails for live TV – amount of buffering depends on the way capacity varied; not all receivers playout at once – see the winning goal after you heard your neighbours cheering…

- Alternative architecture: avoid TCP/IP

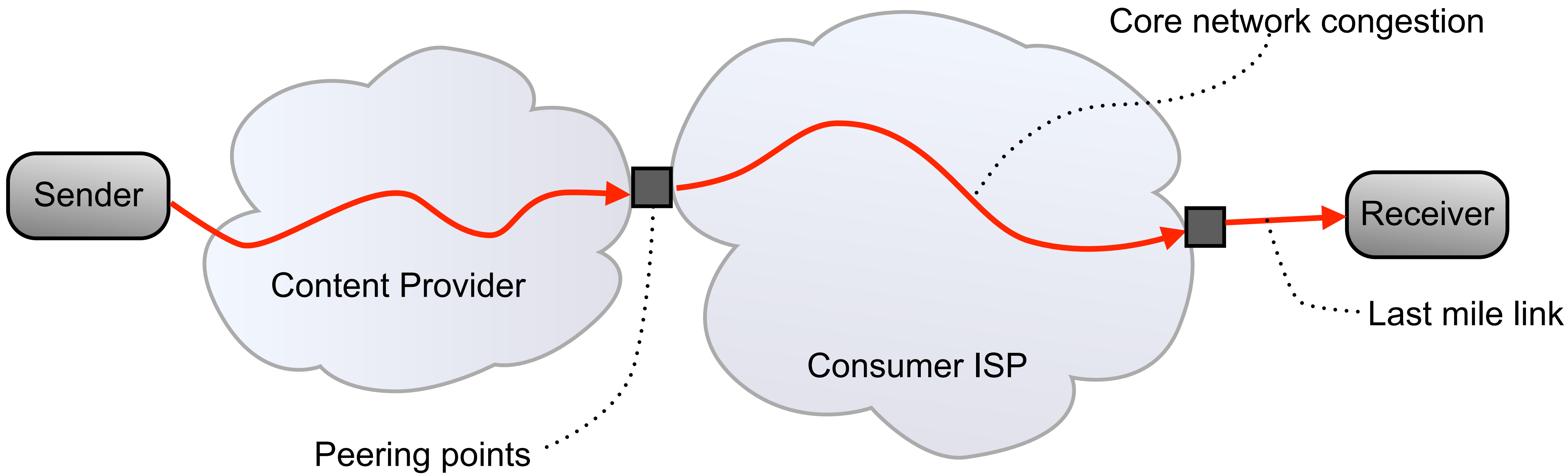
IPTV System Architecture

- MPEG-2 TS over RTP over UDP/IP
  - Choice driven by compatibility with satellite and cable TV distribution networks
- Source-specific multicast transport
  - One multicast group per TV channel
  - Efficient use of core network bandwidth
- Middleboxes for local repair and reception quality monitoring
  - Local NACK-based packet retransmission
  - Aggregation of RTCP reception reports



- ISP provisions sufficient bandwidth in the core – edges unmanaged
- No rate adaptation using UDP/IP
  - Avoids in-network queues, if sending rate below link capacity, since no probing
- Non-web-based infrastructure increases cost and complexity
- New and not well studied – where are the performance problems?

Factors Affecting Quality of Experience



- How do the different components of the path impact performance?
  - Content providers will ensure adequate provisioning at peering points: they're in the business of providing good quality
  - Consumer ISPs care about quality to the extent it prevents customer complaints – *just enough* network capacity
  - Responsibility for last mile link unclear – quality unknown and largely unmanaged
- Hypothesis: content provider and consumer ISP networks perform well enough – IPTV performance problems occur in the last mile
- Aim to conduct experiments to determine if this is true, build a model of the network behaviour

Experimental Setup

- Well-connected server at University of Glasgow
- Soekris net5501 single-board computers located in volunteers' homes as measurement clients
  - Low-power, silent, easily transported, zero-configuration
  - Run FreeBSD 7 with custom measurement software

Source: soekris.com

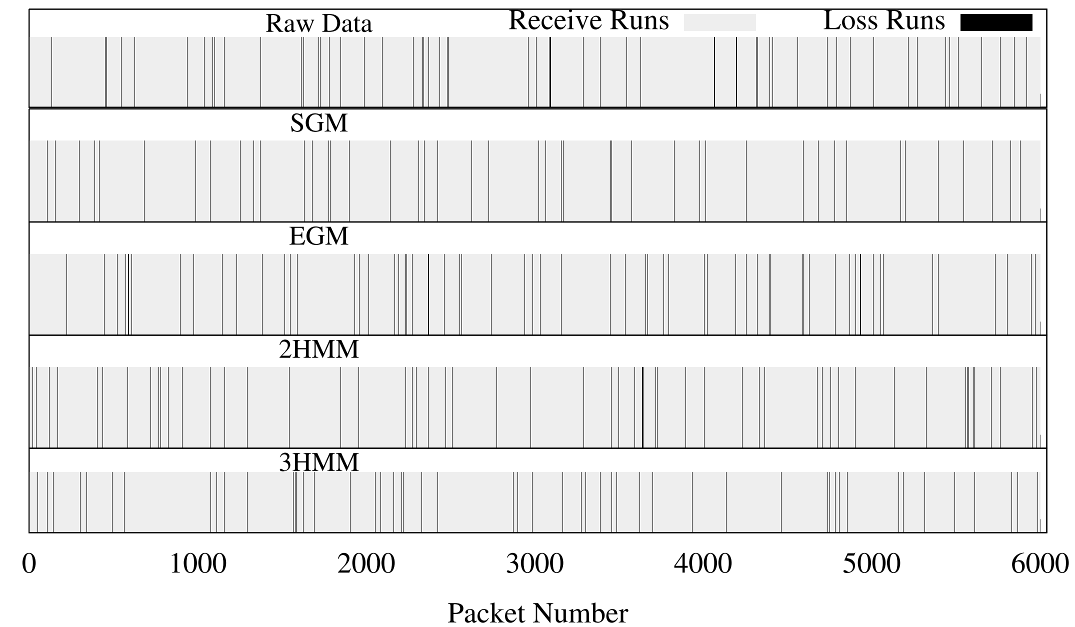
- Primarily end-to-end performance monitoring
  - RTP over UDP/IP streaming; packet sizes and rates chosen to match common IPTV systems
  - July 2009 – September 2010; 3900 traces to 16 destinations in UK and Finland; 230,000,000 packets sent
  - ≥8 measurements per host per day to capture diurnal variation
  - Capture send and receive timestamps, sequence numbers
- Additional metrics:
  - Occasional packet-pair bandwidth estimates
  - Occasional TTL-limited hop-by-hop probing to determine where loss was occurring on the path

Measurement schedule carefully chosen to avoid triggering ISP bandwidth caps

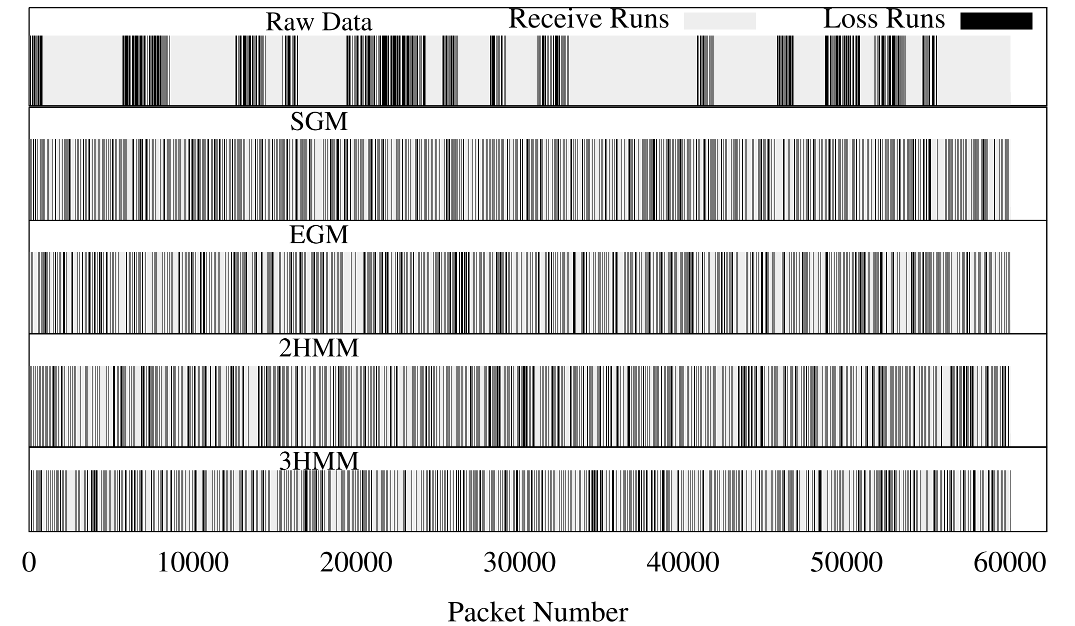
All datasets available online: http://csperkins.org/research/adaptive-iptv/ (~2.6 gigabytes compressed)

Results: Models for Packet Loss

- Clear time-of-day variation in packet loss rate and delay
  - Congestion in ISP networks clearly visible
- Loss distributions can be separated into bursty and non-bursty behaviour
  - Links exhibit both types of behaviour, with no clear time-of-day explanation
- Key finding: simple Markov models not sufficient to model loss
  - Simple and extended Gilbert models, 2- and 3-state Hidden Markov models
  - Differs from backbone networks, where they have been successful
- Need to account for different types of loss pattern
  - Believe we are seeing different loss processes in addition to time-of-day variation



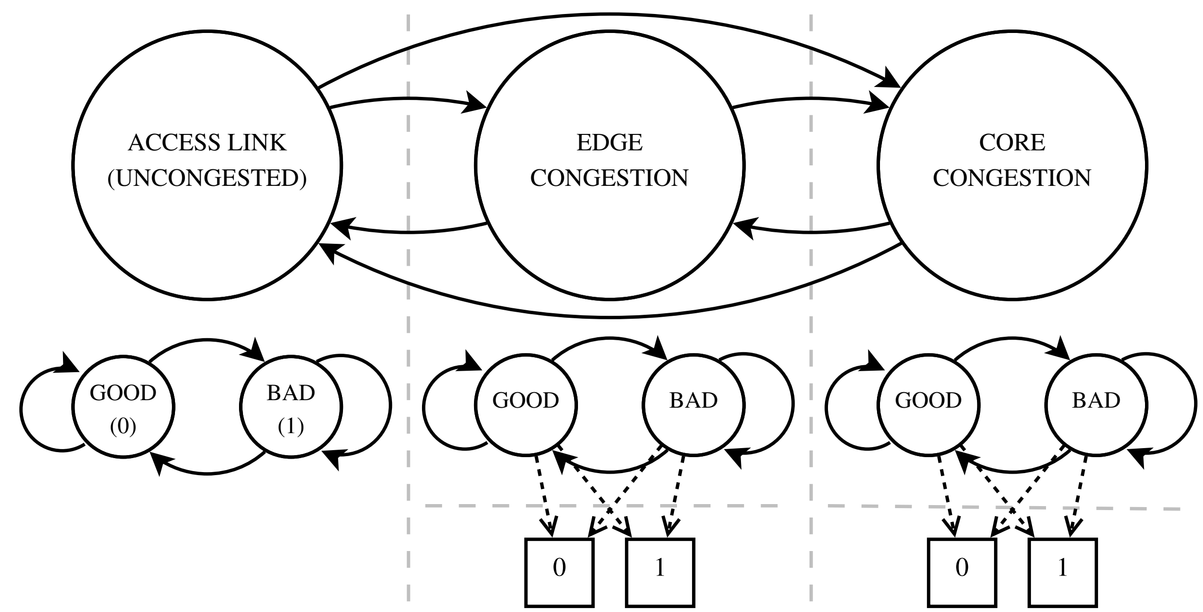
(a) "non-bursty" trace



(b) "bursty" trace

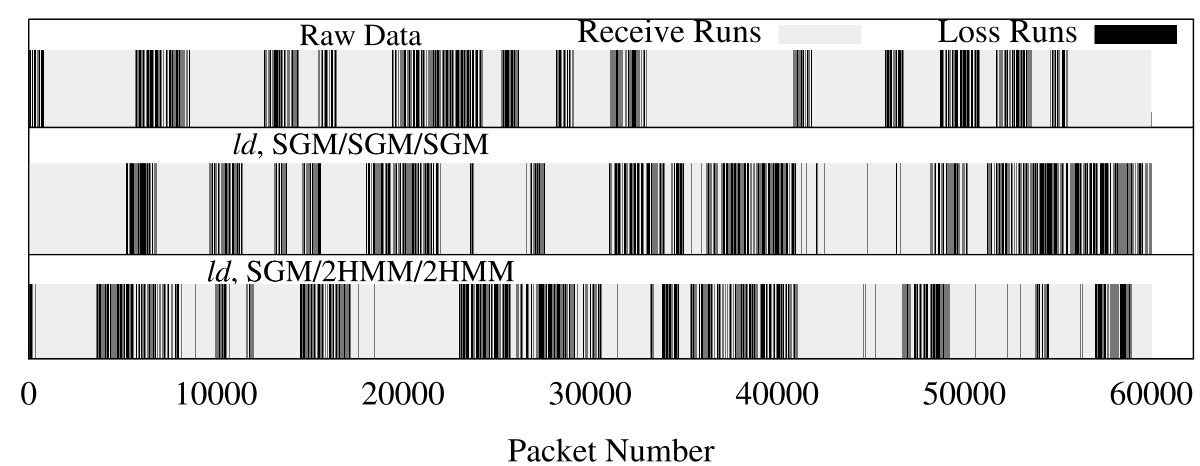
Two-level Model

- Derived a two-level Markov model
  - Top-level states denote location of loss: access link, edge network, core network
  - Lower-level states model loss process at that location



- Classifier segments packet traces according to loss location – based on loss and delay thresholds

- Lower-layer is a mixture of simple Gilbert model and 2-state Hidden Markov models
  - Use of 2HMM for core and edge models better captures bursty loss behaviour



IPTV Performance Summary

- Loss models show complex behaviour, dependant on location of loss
  - Two-level model can be trained to give an accurate model of packet loss behaviour
  - Have applied to tuning application-level FEC parameters
- Packet delay results surprisingly well behaved
  - Traces do not use TCP, and traffic was chosen to fit within the edge link bandwidth to avoid edge queueing – no evidence of buffer bloat in core

Discussion

- Our results can allow optimisation of IPTV service
- On-demand TV services also under intense study

- The two solutions work well, but conflict:
  - IPTV services require low latency → UDP-based, disrupted by TCP traffic
  - Other applications, including on-demand TV, require TCP

- Coming battleground: active queue management *vs.* TCP modifications – network neutrality?