Wayne State University

Client-Driven Price Selection for Scalable Video Streaming with Advertisements⋆

Musab S. Al-Hadrusi and Nabil J. Sarhan

nabil.sarhan@.wayne.edu

MMM 2012 Conference Presentation

January 5, 2012

***This work was supported in part by U.S. NSF grant CNS-0834537.**

*** This paper is published as [24].**

Outline

- Introduction
- Related Work
- Main Contributions
- Proposed Client-Driven Price Selection
- Performance Evaluation and Main Results
- Conclusions

Introduction

- The distribution of streaming media faces a significant scalability challenge.
- This challenge has been addressed by
  - Content Distribution Networks (CDNs)
  - Peer-to-Peer (P2P)
- These approaches mitigate the problem but do not eliminate it.
- The fundamental problem is unicast delivery.
- This paper considers the multicast approach.

Introduction (Cont…)

- Most prior studies considered only streaming primary media
  - No video ads
- The use of ads is important:
  - Generates revenue
  - Converts passive startup waiting times to active waiting
  - May be interesting to users
  - Leads to better request aggregation

Introduction (Cont…)

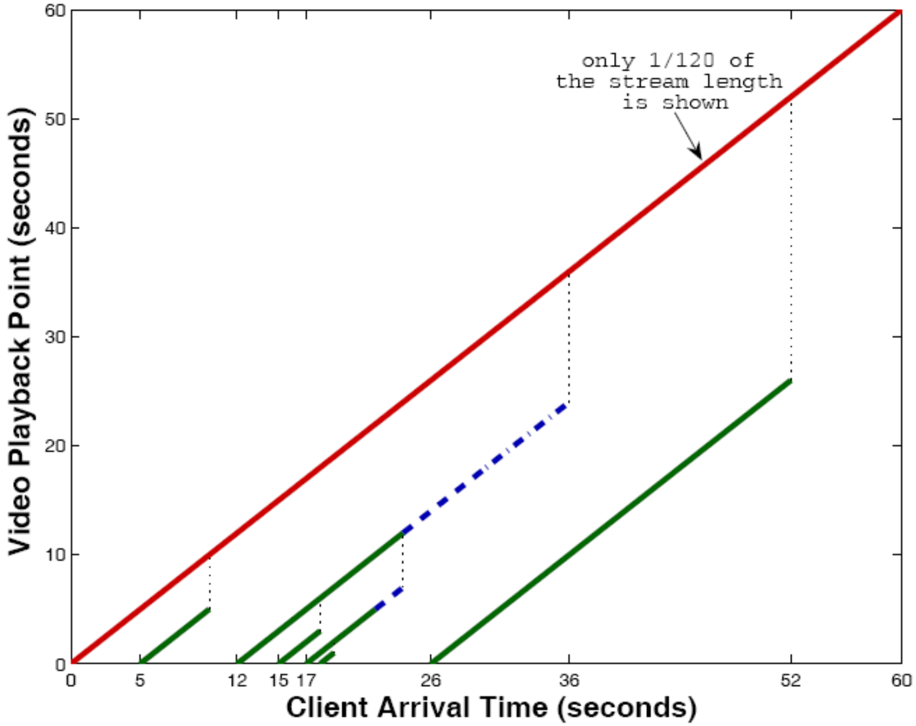
- This paper builds on our prior work
  - In [2], we proposed a scalable delivery framework of premium on-demand media content with video advertisements.
  - In [22,23], we proposed a waiting-time prediction algorithm for video streaming.
  - In [17], we proposed a waiting-time prediction algorithm to estimate the ads' viewing period in the scalable delivery framework.

Introduction (Cont…)

- This paper analyzes a solution that provides clients with the opportunity for selecting from multiple price options.
- For example, for a specific premium content, the client may be provided with three price options:
    - $2 when viewing one ad
    - $1.6 when viewing two ads
    - $1.2 when viewing three ads

Related Work

- Multicast-based delivery can be classified into
  - Stream Merging [5,22,23,18,16]
    - Earliest Reachable Merge Target (ERMT) [5,6]



ERMT

  - Periodic Broadcasting [11,20,8]

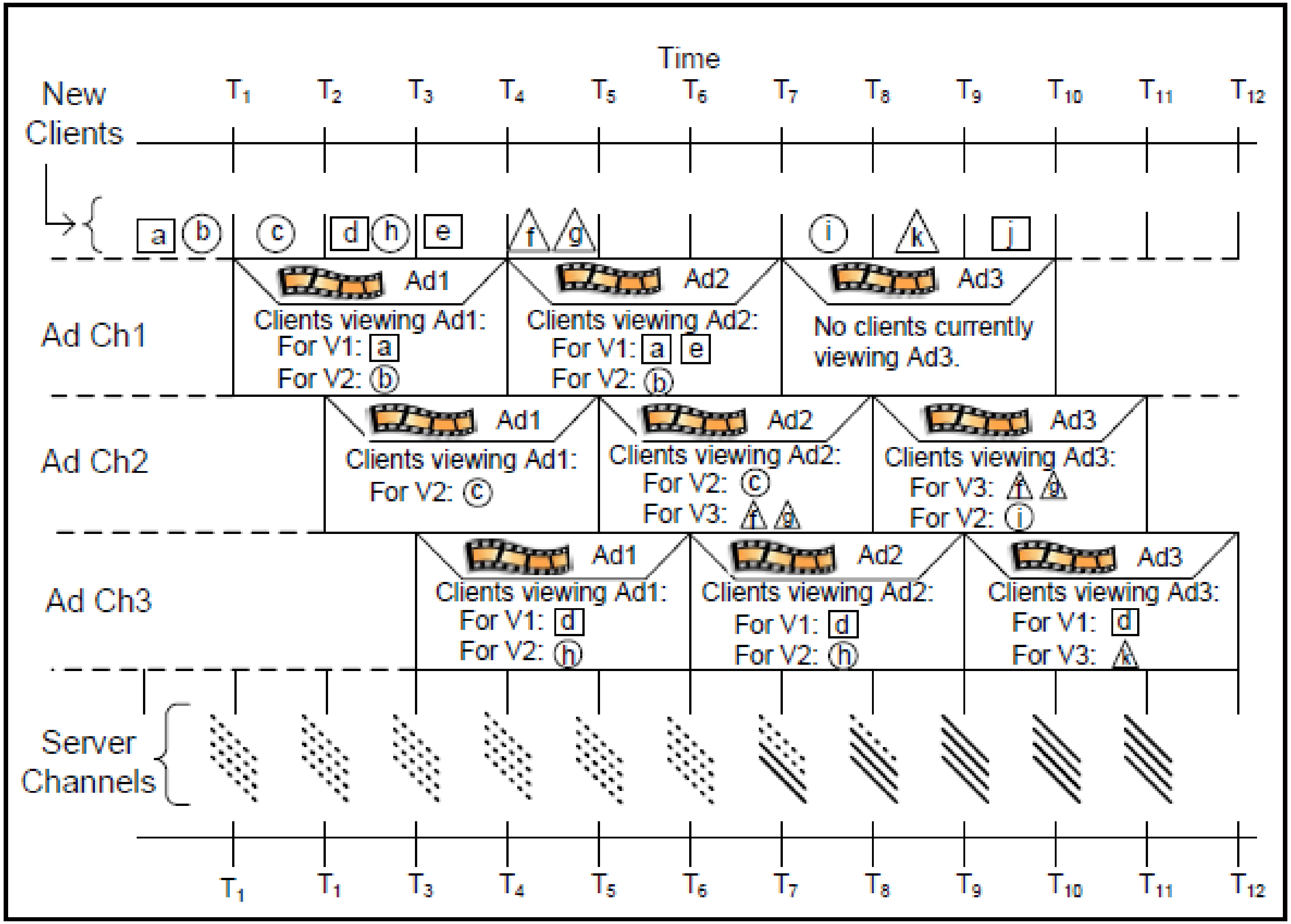
Related Work (Cont…)

- Request Scheduling
  – One waiting queue for each video
  – Minimum Cost First (MCF-P) [19] is the most efficient.
    - It selects the queue with the least cost per request.

Related Work (Cont…)

- In [2], we proposed a scalable delivery framework.
  - Ads are combined and broadcast on dedicated channels.
  - Clients start by joining an ads' broadcast channel for some time and then receive the requested video by stream merging.
  - Constraints are placed on MCF-P Scheduling
    - Any N: at least 1 client viewed N ads
    - Each N: Each client viewed at least N ads

Related Work (Cont...)



Scalable Delivery Framework with Any 2

Related Work (Cont…)

- In [17], we proposed a waiting-time prediction algorithm to estimate the ads' viewing period
  - Estimation is based on the requested video and the current system state.
  - A client is presented with an expected ads' viewing time and the associated price.
  - The revenues generated from the ads are used to subsidize the price.
  - The algorithm considers the dynamic and complex natures of stream merging and request scheduling.
- The main limitation of that work is that the client is presented with only one choice of price.

Main Contributions

- Proposing a predictive pricing scheme, called Client-Driven Price Selection (CPS), which
  - Provides clients with multiple price options, each with a certain number of expected viewed ads.
  - Enhances the revenue and profit by attracting more clients.
- Analyzing CPS under different scheduling policies and workload characteristics.

Main Contributions (Cont…)

- Combining the equation-based and willingness-based arrival rate models to assess the impacts of purchasing capacity and defection probability on the effective arrival rate.
- Experimenting with a variety of price selection criteria: *Random Price*, *Lowest Price*, *Median Price*, and *Highest Price*.

Proposed Scheme

- With the proposed Client-Driven Price Selection (CPS) scheme, the client is offered a menu with different titles each has multiple prices.
- All ads' revenues are used to subsidize the price.
- Pricing depends on
  - Chosen ads' viewing time
  - Royalty fee
  - Current delivery cost

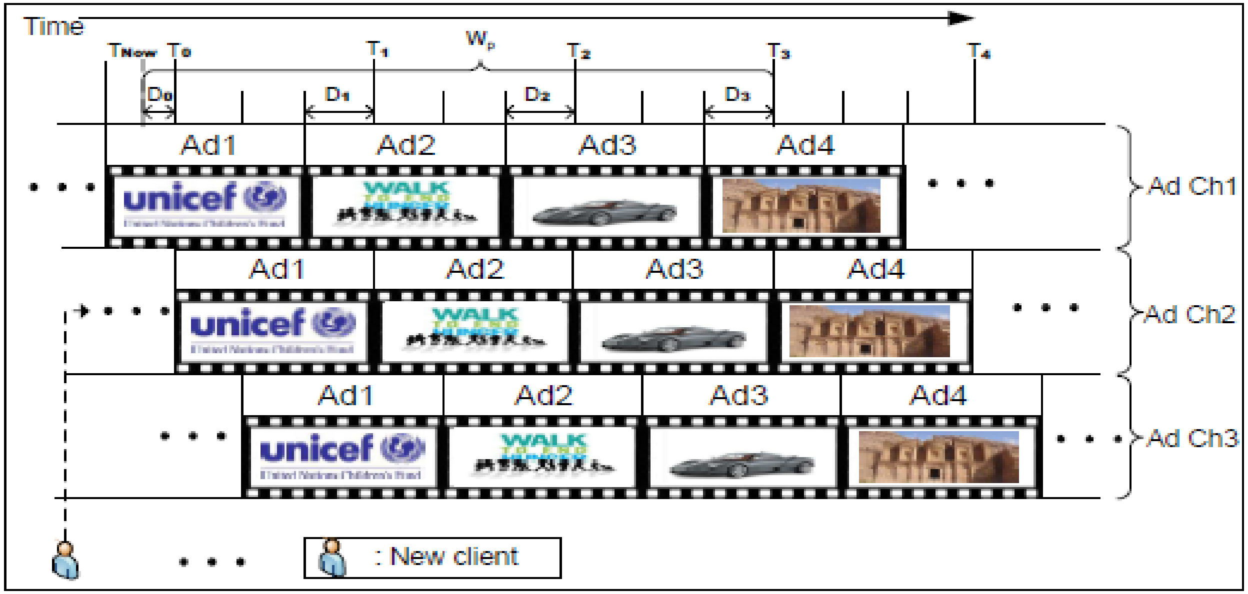
Proposed Schemes (Cont…)

- Waiting-time prediction in the scalable framework is complicated because
  - Both periodic broadcasting and stream merging are used.
  - Different clients may join different ads’ channels.
  - Clients should receive multiple numbers of ads.
    - Partial viewing is not an attractive choice.
  - Constraints on minimum ads’ viewing times must be met.
  - The number of available channels at a future time is somewhat hard to predict.
  - Multiple ads’ channels may become available simultaneously.

Overview of CPS(Cont…)

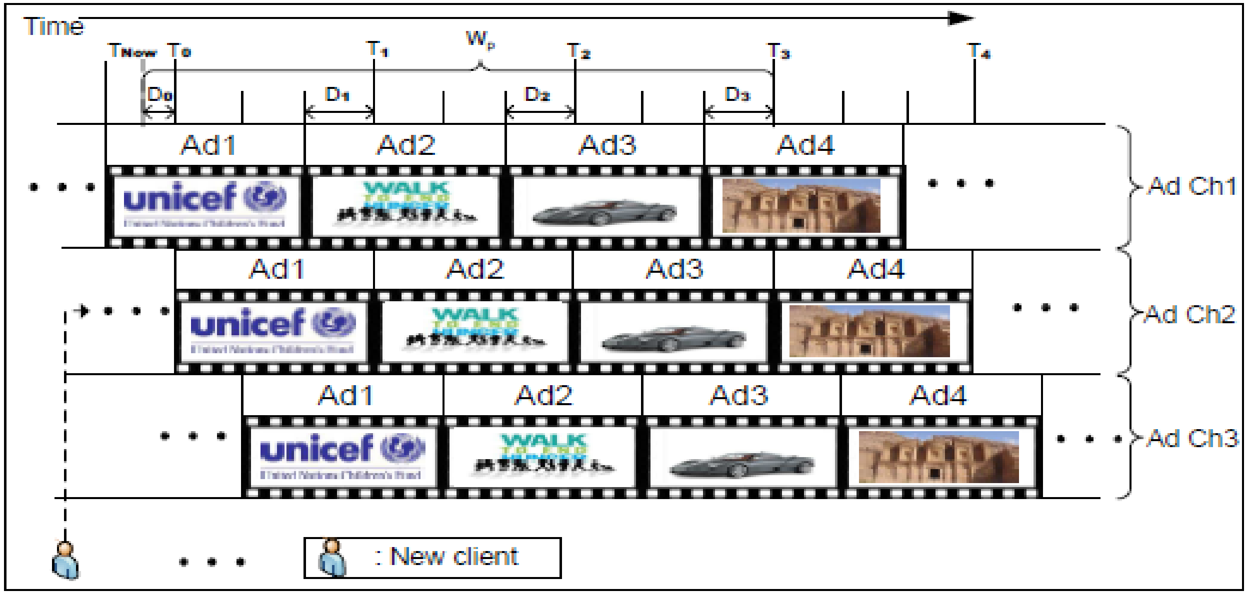
- CPS has three main parts:
  - Video Selection Prediction
  - Channel Availability Prediction
  - Scheduling Qualification Prediction

Overview of CPS(Cont…)



1. Request is mapped to the closest ads' channel.

Overview of CPS(Cont...)



2. Examine the future ad start times on that channel for possible assignment as the expected time.

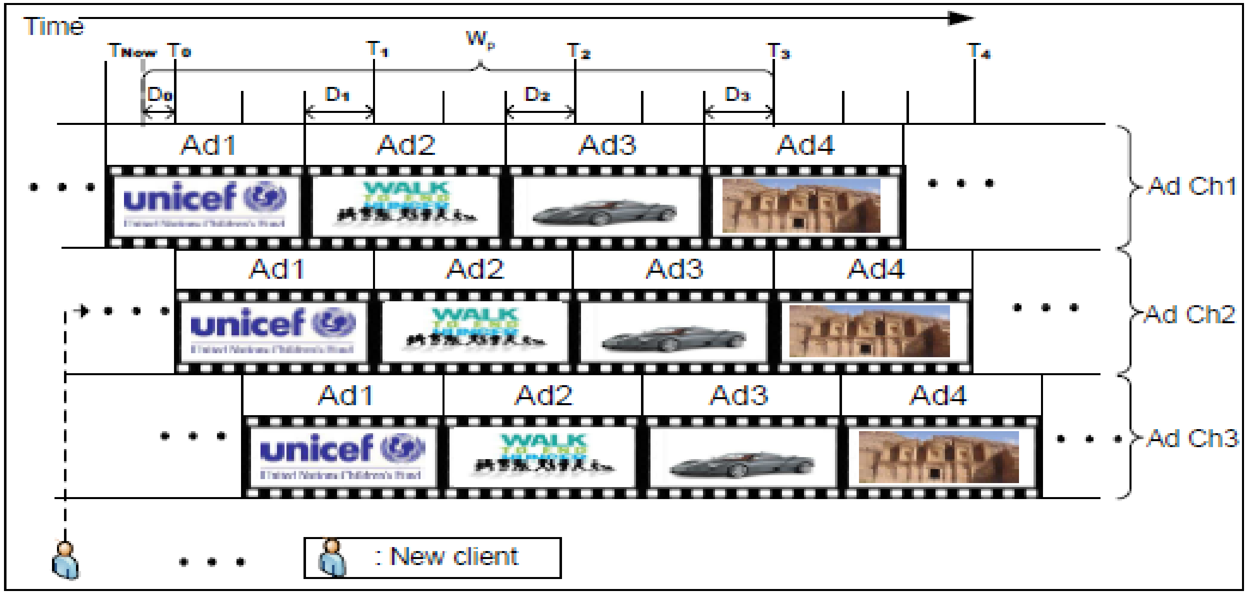
   At each ad start time,

   – Estimate the number of available channels.

   – Predict the videos that can be served at that time.

   – If the requested video is the expected video to be serviced, the corresponding ads' viewing time is added to the list of choices.

Overview of CPS(Cont…)



3. Continue until the *prediction window* (*Wp*) is exceeded.
   – *Wp* controls the implementation complexity and introduces a tradeoff between prediction accuracy and the percentage of clients receiving expected times.

Proposed CPS Algorithm

```
for (v = 0; v < N_v; v++) // Initialize
  assigned_time[v] = -1; // Not assigned expected time
adChNo = Get label # of the ads' channel with the closest start time;
T = Get next ad's start time; T_0 = T; examined_times = 0;
// Find mumber of available channels at time T
N_c = available_channels + will_be_available(T_Now, T);
while (T < T_Now + W_p) { // Loop till prediction window is exceeded
  for (v = 0; v < N_v; v++){
    if (isQualified(v, T, adChNo)) {
      if (assigne_time[v] == -1)
        expected_qlen = qlen(v, adChNo) + λ[v]×
        ((T - T_Now) + examined_times × ad_len/N_adCh);
      else // Video v has been assigned an expected time
        expected_qlen = λ[v] × (T - assigned_time[v])/N_adCh;
      objective[v] = find scheduling objective for video v;
    } // end if (isQualified(v, T, adChNo))
    else objective[v] = -1; // v is not qualified
  } // end for (v = 0; v < N_v; v++)
  while (c = 0; c ≤ N_c; c++){ //for every available channel
    // Find the expected video to serve at time T
    expected_video = find video with maximum nonzero objective;
    if(expected_video == v_j) {
      Push T into expected time que Que_ET of request R_i;
      TempPrice = CalculatePrice(T, v_j);
      Push TempPrice into price que Que_price of request R_i;
      break; }
    else { assigned_time[expected_video] = T;
        objective[v] = -1; } // -1 means can't be selected again
  } // end while (c = 0; c ≤ N_c; c++)
  T = T + ad_len; //Proceed to the next edge
  // Find mumber of available channels at time T
  N_c = left_over + will_be_available(T - ad_len/N_adCh, T);
  examined_times++;
} // end while (T < T_Now + W_p)
Present Que_Price to client and then clear it.
```

Performance Evaluation and Main Results

Summary of Workload Characteristics

Parameter	Model/Value(s)
Request Arrival	Poisson Process
Request Arrival Rate	Variable, Default = 40 Requests/min
Server Capacity	200-550
Video Access	Zipf-Like, Skew Parameter $\theta = 0.271$
Movie-Related Characteristics	80 120-min movies
Waiting Tolerance Model for clients without expected times of service	Poisson, min= 3 ads, mean= 5 ads, max= 8 ads
Waiting Tolerance Model for clients with expected times of service	Expected Service Time + Wad, Wad: Variable, Default= 2 Ad lengths
Ad-Related Characteristics	Ad Length= 30 sec, # different ads= 8, # ads channels= 3
Minimum Ads Constraint (N)	Variable, Default = 2
Prediction Window (W_p)	Variable, Default = 9
Qualification Threshold (Q_{Th})	Variable, Default = 0.25
Scale (b), Shape (α), Elasticity (δ)	Variables, Defaults= 1.0, 1, 7, resp.
Equation-Based Model Constants	$c1 = 60$, $c2 = 0.5$, $c3 = 1$

Hybrid Model

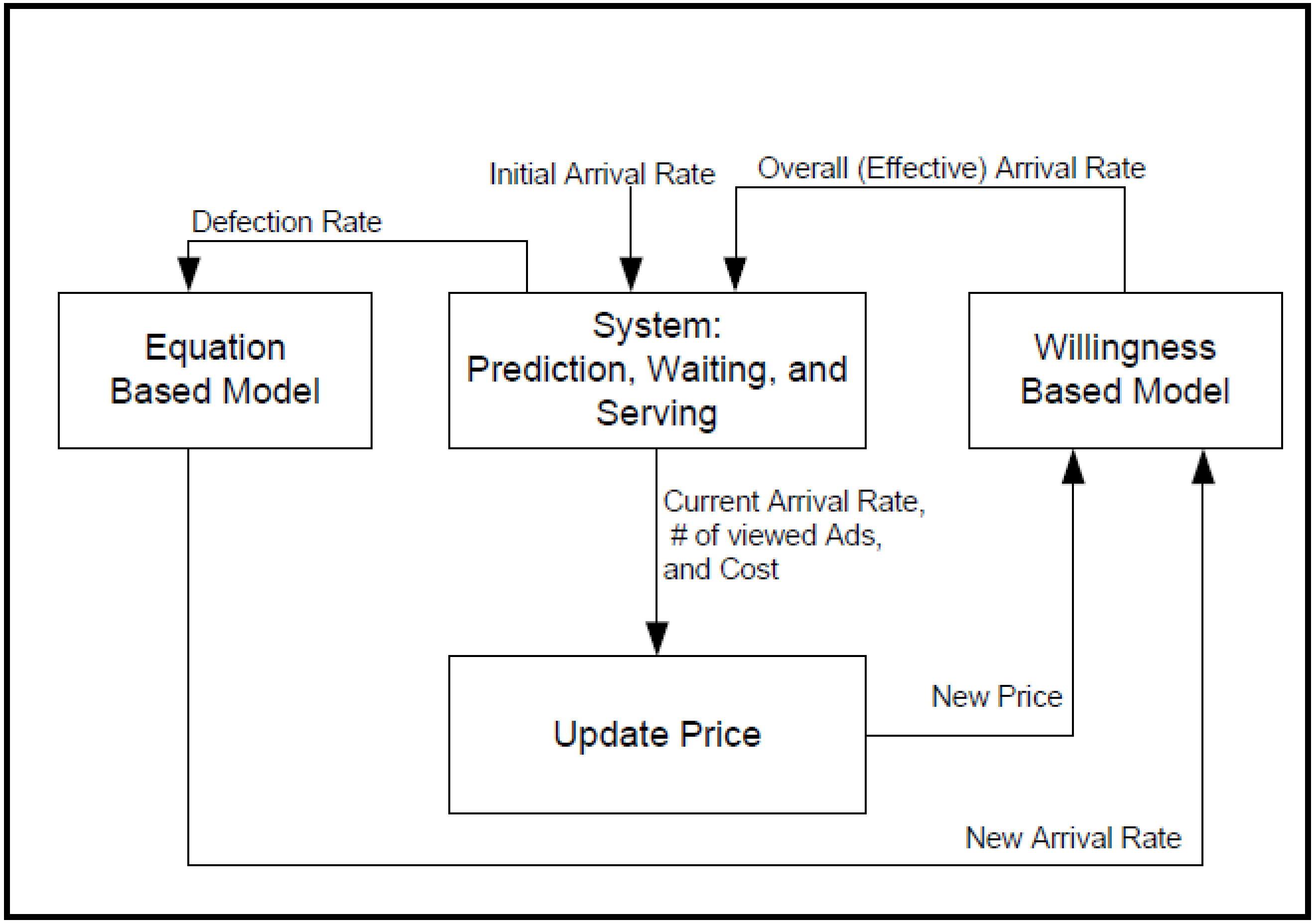
- We consider customer purchasing capacity and willingness as well as defection rate to model the effective arrival rate.
  - The capacity of clients to spend is highly skewed and follows Pareto distribution [12]

$$f_p(x) = \alpha \times b \times x^{-(\alpha+1)} \text{ for } x \geq b,$$

$$Prob(willingness) = \begin{cases} 1 - (\frac{p}{y})^{\delta} & 0 \leq p \leq y, \\ 0 & p > y. \end{cases}$$

$$\lambda = \frac{c_1(1-d)}{c_2 + c_3 d^2}$$

Hybrid Model (Cont…)

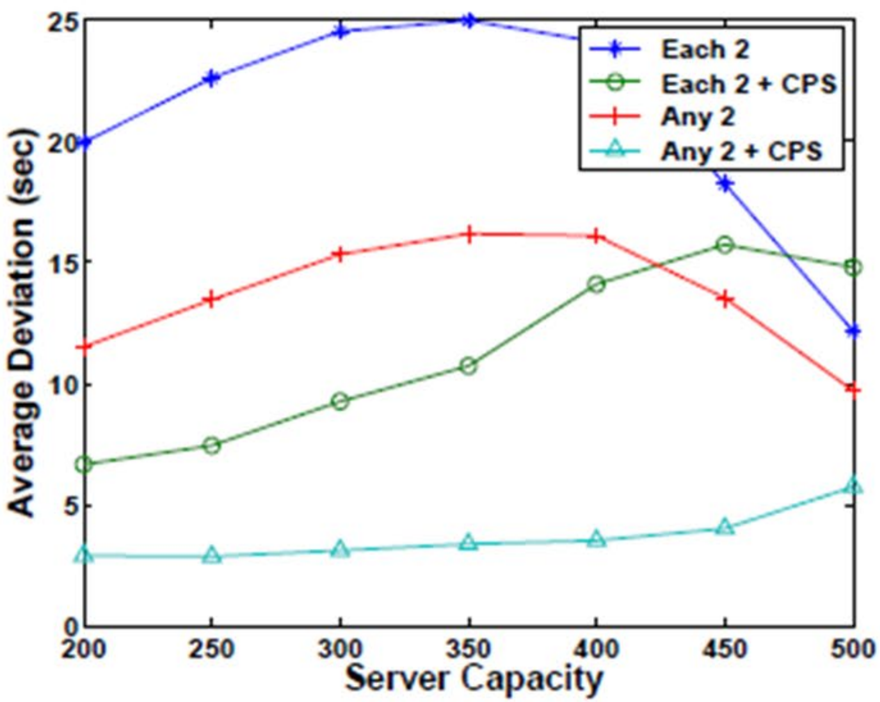


Performance Metrics

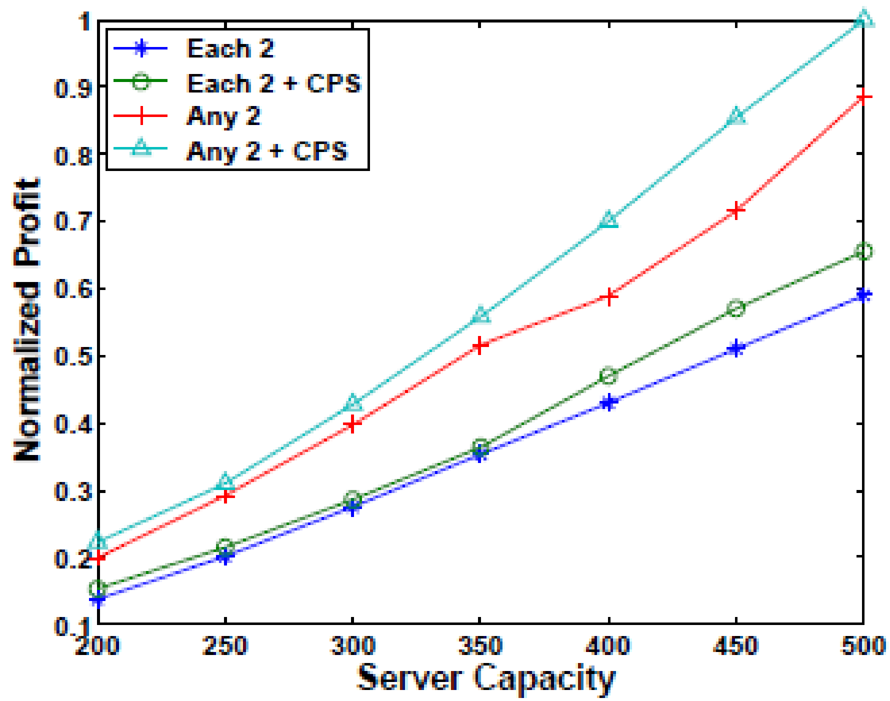
- Prediction accuracy
- Percentage of clients receiving expected times of service (PCRE)
- Client defection probability
- Average waiting time
- Price
- Arrival rate
- Profit
- Revenue

Main Results

Effectiveness of CPS (Cont…)

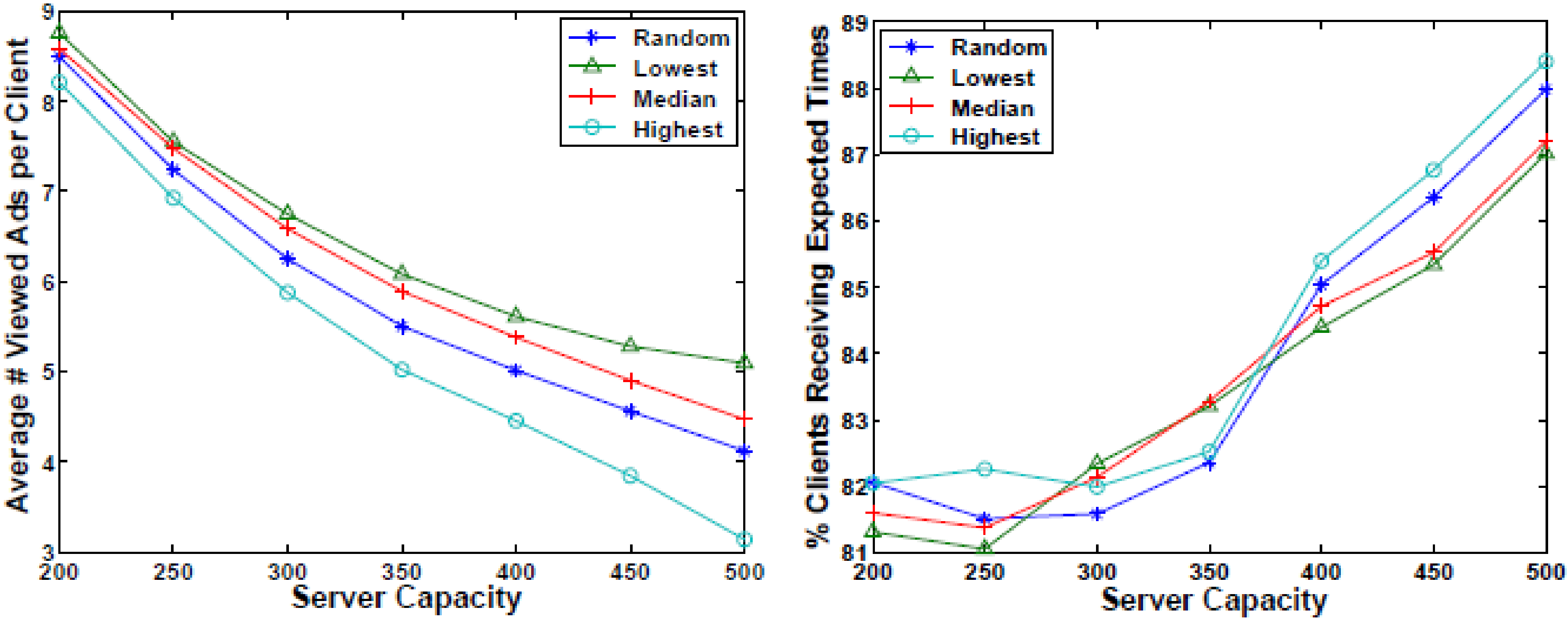


(d) Average Deviation



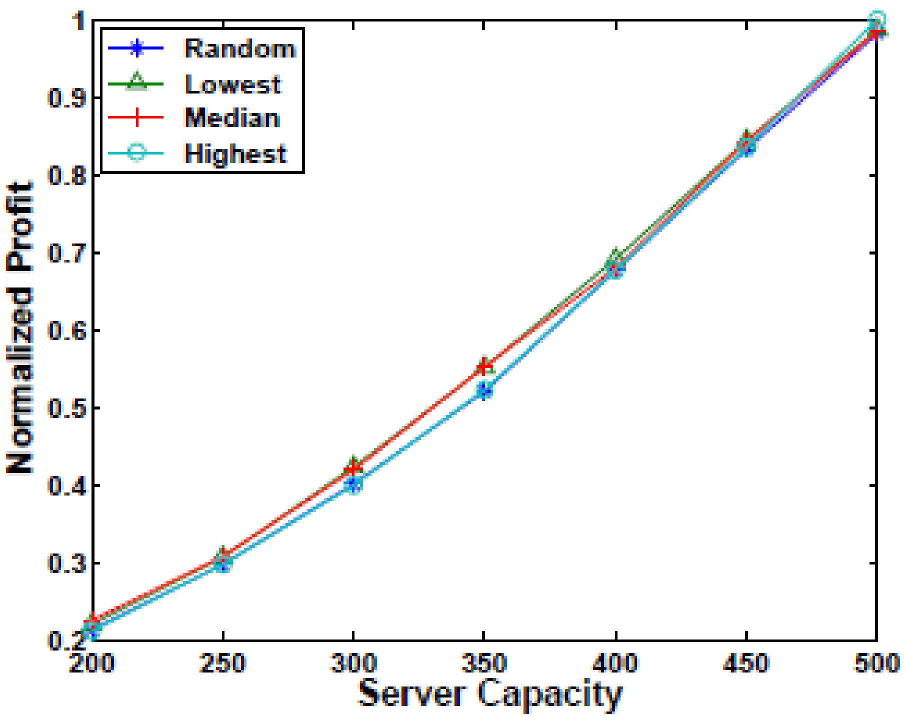
(e) Profit

Comparing Various Price-Selection Criteria

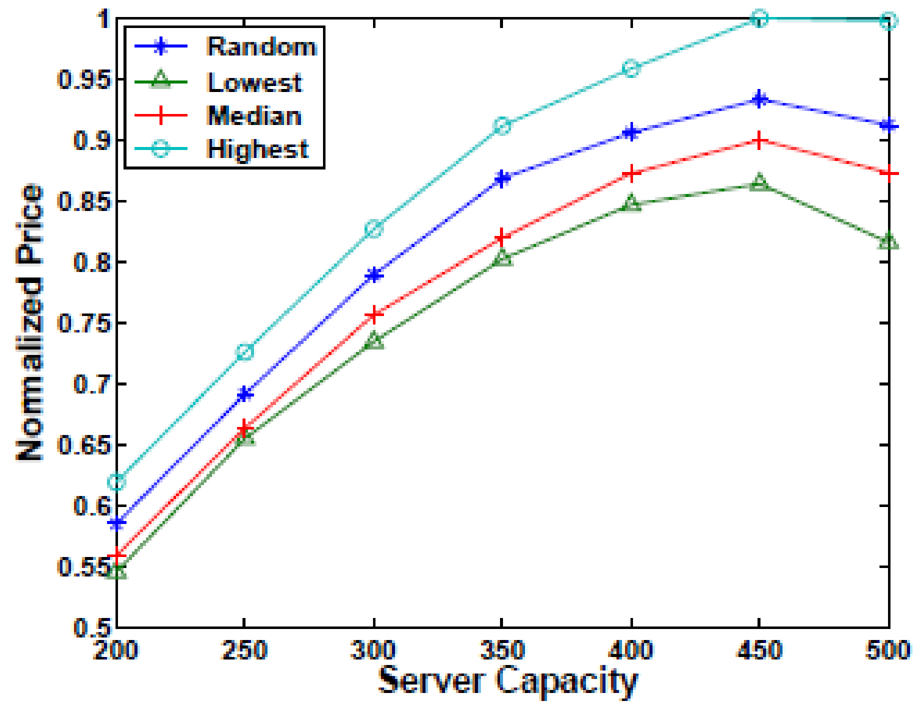


(a) Waiting Time (b) PCRE

Comparing Various Price Selection Criteria
(Cont…)



(c) Profit



(d) Price

Conclusions

- The proposed CPS approach enhances the revenue and profit by giving multiple price choices to the client, thereby attracting more clients.
- CPS is best when combined with Any N scheduling and ERMT.
- The achieved waiting time prediction accuracy with this combination is within 6 seconds (20% of an ad length).

References

[1] Aggarwal, C.C., Wolf, J.L., Yu, P.S.: The maximum factor queue length batching scheme for Video-on-Demand systems. IEEET Trans. On Computers 50(2),97–110 (Feb2001)

[2] Al-Hadrusi, M., Sarhan, N.J.: A scalable delivery framework and a pricing model for stream ing media with advertisements. In: Proc. of SPIE/ACM Multimedia Computing and Net working Conference (MMCN) (Januray 2008)

[3] Cai, Y., Hua, K.A.: An efficient bandwidth-sharing technique for true video on demand sys tems. In: Proc. of ACM Multimedia. pp. 211–214 (Oct 1999)

[4] Dan, A., Sitaram, D., Shahabuddin, P.: Scheduling policies for an on-demand video server with batching. In: Proc. of ACM Multimedia. pp. 391–398 (Oct 1994)

[5] Eager, D.L., Vernon, M.K., Zahorjan, J.: Optimal and efficient merging schedules for Video on-Demand servers. In: Proc. of ACM Multimedia. pp. 199–202 (Oct 1999)

[6] Eager, D.L., Vernon, M.K., Zahorjan, J.: Minimizing bandwidth requirements for on-demand data delivery. IEEE Trans. on Knowledge and Data Engineering 13(5), 742–757 (Sept 2001)

[7] Gao,L., Kurose, J., Towsley, D.: Efficient schemes for broadcasting popular videos. In: Proc. of the Int'l Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV) (July 1998)

References

[8] Gill, P., Shi, L., Mahanti, A., Li, Z., Eager, D.: Scalable on-demand media streaming for heterogeneous clients. ACM Transactions on Multimedia Computing, Communications, and Applications (ACM TOMCCAP) 5(1), 1–24 (Oct 2008)

[9] Hu, A.: Video-on-Demand broadcasting protocols: A comprehensive study. In: Proc. of IEEE INFOCOM(April 2001)

[10] Hua, K.A., Cai, Y., Sheu, S.: Patching: A multicast technique for true Video-on-Demand services. In: Proc. of ACM Multimedia. pp. 191–200 (1998)

[11] Huang,C.,Janakiraman,R.,Xu,L.: Loss-resilient on-demand media streaming using priority encoding. In: Proc. of ACM Multimedia. pp. 152–159 (Oct 2004)

[12] Jagannathan, S., Almeroth, K.C.: The dynamics of price, revenue, and system utilization. In: Proc. of the IFIP/IEEE International Conference on Management of Multimedia Networks and Services. pp. 329–344 (2001)

[13] Ma, H., Shin, G.K., Wu, W.: Best-effort patching for multicast true VoD service. Multimedia Tools Appl. 26(1), 101–122 (2005)

References

[14] Rayburn, D.: Streaming and Digital Media: Understanding the Business and Technology. Focal Press (2007)

[15] Rocha, M., Maia, M., Cunha, I., Almeida, J., Campos, S.: Scalable media streaming to interactive users. In: Proc. of ACM Multimedia. pp. 966–975 (Nov 2005)

[16] Sarhan, N.J., Das., C.R.: Caching and scheduling in NAD-based multimedia servers. IEEE Trans. on Parallel and Distributed Systems 15(10), 921–933 (Oct 2004)

[17] Sarhan, N.J., Al-Hadrusi, M.: Waiting-time prediction and QoS-based pricing for video streaming with advertisements. In: Proc. of IEEE International Symposium on Multimedia (ISM) (December 2010)

[18] Sarhan, N.J., Das, C.R.: A new class of scheduling policies for providing time of service guarantees in Video-On-Demand servers. In: Proc. of the 7th IFIP/IEEE Int'l Conf. on Management of Multimedia Networks and Services. pp. 127–139 (Oct 2004)

References

[19] Sarhan, N.J., Qudah, B.: Efficient cost-based scheduling for scalable media streaming. In: Proc. of Multimedia Computing and Networking Conf. (MMCN) (January 2007)

[20] Shi, L., Sessini, P., Mahanti, A., Li, Z., Eager, D.L.: Scalable streaming for heterogeneous clients. In: Proc. of ACM Multimedia. pp. 337–346 (Oct 2006)

[21] Tantaoui, M.A., Hua, K.A., Do, T.T.: Broadcatch: A periodic broadcast technique for heterogeneous Video-on-Demand. IEEE Trans. on Broadcasting 50(3) (Sept 2004)

[22] Mohammad Alsmirat, Musab Al-Hadrusi, and Nabil J. Sarhan. Analysis of Waiting-Time Predictability in Scalable Media Streaming. In Proceedings of ACM Multimedia , pages 727 - 736, Augsburg, Germany, September 2007. Acceptance Rate: 19%. DOI: https://doi.org/10.1145/1291233.1291398.

[23] Nabil J. Sarhan, Mohammad A. Alsmirat, and Musab Al-Hadrusi. Waiting-Time Prediction in Scalable On-Demand Video Streaming . ACM Transactions on Multimedia Computing, Communications, and Applications (ACM TOMCCAP), Volume 6, Issue 2, March 2010. DOI: https://doi.org/10.1145/1671962.1671967.

[24] Musab Al-Hadrusi and Nabil J. Sarhan. Client-Driven Price Selection for Scalable Video Streaming with Advertisements. In Proceedings of the International MultiMedia Modeling Conference (MMM 2012), pages 429 - 439, Klagenfurt, Austria, January 2012. Acceptance rate: 34.9%. DOI: https://doi.org/10.1007/978-3-642-27355-1_40.