Fast Web Page Downloads by Competitive Content Delivery Networks

ÖZGÜR ERÇETİN and Leandros TASSIULAS

Most of the user requests are for a few popular objects.
- Web sites may become overloaded and propagation delays may cause high retrieval delays.
- It is common practice to use caches to store popular objects to improve user latency and reduce the network load.
- Speculative (best-effort) caching:
  *Proxy caches are placed at gateways, ISPs, etc. and serve incoming user requests if the requested object is available, otherwise user request is forwarded to the WAN.
  *User requests are served without any QoS.
  *Transparent proxy, mirror sites are examples.

Content Delivery Problem
- Publishers subscribe for content delivery services from the CDNs.
- Objective: Maximize total net average publisher utility.
  *Publisher utility is a concave function of the average latency the users retrieve the publisher’s content.

Web server is the point the content first enters the Internet.
- Publishers ultimately control the content and its distribution.
  Publishers may negotiate SLA (Service Level Agreement) in terms of average user latency, storage capacity or the extend of geographical coverage of the content.
- Surrogate is a delivery server other than the origin server.
  Limited storage and transmission bandwidth resources.

CDN System Operation
1) Publisher selects a desired level of QoS.
   In terms of caching space, geographical dissemination, average user latency.
2) Disseminate content according to the desired QoS, user request distribution, location and capacity of the surrogates.
3) Direct user requests to the appropriate surrogates.
   URL re-routing, application-level routing, transparent proxies.
4) May periodically re-arrange objects and/or QoS.
5) Direct user requests to peer CDNs, if it is optimal.
6) Keep access records for accounting.

Distribution Sub-system coordinates the activity of moving publisher’s content to one or more of the surrogates.
- Request-Routing Sub-system (RRS) coordinates the activity of directing a client request to a suitable surrogate.
- Accounting Sub-system determines the methods for measurement and pricing of the distribution and delivery activities.

Content Delivery Network Model
- Publishers invest in the surrogates for improved user latency.
- Surrogates compete among each other for publisher business and sell caching and bandwidth resources.
- No cooperation between market agents other than resource price advertisements.

Optimization Problem
\[ \beta = \frac{\lambda}{\lambda/\lambda'} \] is the gain factor
\[ \lambda' = \text{total caching space allocated to publisher } i \text{ in surrogate } j \]
\[ U_i(x_i') = \sum \beta_i' (x_i')^{\lambda'/\lambda} \]
Publisher benefit, total additional delay improvement from use of surrogates

CONCLUSIONS
- Performance of non-cooperative market approaches the system optimum solution.
- Unlike system optimization, market-based operation enables improvement of benefits of the individual publishers.