Online Supplement to

Analyzing Document-Duplication Effects on Policies for Browser and Proxy Caching

INFORMS Journal on Computing

Yong Tan
University of Washington Business School, Box 353200, Seattle, Washington 98195-3200, USA,
ytan@u.washington.edu

Yonghua Ji
University of Alberta School of Business, Edmonton, Alberta T6G2R6, Canada,
yji@ualberta.ca

Vijay S. Mookerjee
University of Texas at Dallas, School of Management, JO4.624, Richardson, Texas 75803-0688, USA,
vijaym@utdallas.edu

Appendix

A1. Derivation of Probability Density

The probability density of documents with strictly positive ages at time $t + h$ is given by the probability density of documents ages reduced by h at time = t, *multiplied by* the probability that there are no accesses made by the user in the interval $[t, t + h]$.

Let, $\Pi(h)$ be the probability that no accesses take place in h time units on any of the n documents with positive ages given by $x_1 - h, \cdots, x_n - h$,

$$\Pi(h) = 1 - \sum_{i=1}^{n} \int_{x_i - h}^{x_i} \theta(x_i) dx_i \,. \tag{A1}$$

In the equation above, we have used the property of a Poisson distribution that the probability of two or more access on a document in a short time interval h can be ignored. We get the following differential equation:

$$f_{t,n}(t + h, x_1, \cdots, x_n) = \Pi(h) f_{t,n}(t, x_1 - h, \cdots, x_n - h) \,. \tag{A2}$$

Dividing throughout by h and taking limits $h \to 0$, $t \to \infty$, we get the required steady state equation,

$$\sum_{i=1}^{n} \frac{\partial f_n(x_1, \cdots, x_n)}{\partial x_i} = -\sum_{i=1}^{n} \theta(x_i) f_n(x_1, \cdots, x_n) \,. \tag{A3}$$

The probability density that there is one document of zero age and the others are of positive ages is given by the probability density of a system with all documents of positive ages *multiplied by* the probability that any document in the system is accessed. Note that since a single user generates the access requests, two or more documents cannot be simultaneously accessed. We get the following boundary condition:

$$f_n(x_1, x_2, \cdots, x_i = 0, \cdots, x_n) = \int_0^\infty \theta(x_i) f_n(x_1, x_2, \cdots, x_i, \cdots, x_n) dx_i \ . \tag{A4}$$

The differential equation and boundary condition defined in (A3) and (A4) can be verified to have the following solution:

$$f_n(x_1, x_2, \cdots, x_n) = \omega \exp\left(- \sum_{i=1}^{n} \phi(x_i) \right), \tag{A5}$$

where ω is a normalizing constant that ensures that $f_n(x_1, x_2, \cdots, x_n)$ integrates to unity over the appropriate limits and $\phi(x)$ is the indefinite integral of $\theta(x)$.

A2. Evaluation of *w*(*n*, *R*)

The expression $w(n, R)$ is the expected delay when there are *n* documents, browser capacity is *R*, and no proxy server is present, explicitly,

$$w(n,R) = n! \int_0^{1/\beta} f_\theta(\theta_1) d\theta_1 \int_0^{\theta_1} f_\theta(\theta_2) d\theta_2 \cdots \int_0^{\theta_{n-1}} f_\theta(\theta_n) d\theta_n \sum_{i=R+1}^{n} Z \frac{\theta_i}{H_0} .$$

Substituting in the explicit expression of $f_\theta(\theta)$, and have $y = \beta\theta$, we have,

$$w(n,R) = n! \left(\frac{1}{\alpha} - 1 \right)^n \int_0^1 y_1^{\frac{1}{\alpha}-2} dy_1 \cdots \int_0^{y_{R-1}} y_R^{\frac{1}{\alpha}-2} dy_R \int_0^{y_R} y_{R+1}^{\frac{1}{\alpha}-2} dy_{R+1} \cdots \int_0^{y_{n-1}} y_n^{\frac{1}{\alpha}-2} dy_n \frac{1}{\beta} \sum_{i=R+1}^{n} \frac{y_i Z}{H_0} .$$

This gives,

$$w(n,R) = n! \left(\frac{1}{\alpha} - 1 \right)^n \frac{1}{\beta} \frac{Z}{H_0} \sum_{k=1}^{n-R} \prod_{j=1}^{k-1} \frac{1}{j(1/\alpha - 1)} \prod_{i=k}^{n} \frac{1}{i(1/\alpha - 1) + 1} .$$

By induction, we have,

$$\sum_{k=1}^{l} \prod_{j=1}^{k-1} \frac{1}{j(1/\alpha - 1)} \prod_{i=k}^{l} \frac{1}{i(1/\alpha - 1) + 1} = \frac{\alpha}{(1/\alpha - 1)^{l-1} (l-1)!} .$$

Therefore,

$$\begin{aligned} w(n,R) &= n! \left(\frac{1}{\alpha} - 1 \right)^n \frac{1}{\beta} \frac{Z}{H_0} \frac{\alpha}{(1/\alpha - 1)^{n-R-1} (n-R-1)!} \prod_{i=n-R+1}^{n} \frac{1}{i(1/\alpha - 1) + 1} \\ &= Z \left(1 - \frac{R}{n} \right) \prod_{i=n-R+1}^{n} \left(1 + \frac{\alpha}{i(1-\alpha)} \right)^{-1} . \end{aligned}$$

A3. Evaluation of w_1 and w_2

We first evaluate the conditional expectation,

$$\int_0^{\tilde{\theta}} \theta f_c(\theta \mid \tilde{\theta}) d\theta ,$$

which can be written as,

$$\int_0^{\tilde{\theta}} \theta \frac{f_{2,\theta}(\theta, \tilde{\theta})}{g_\theta(\tilde{\theta})} d\theta = \frac{1}{g_\theta(\tilde{\theta})} \left(\int_0^{\tilde{\theta}} \theta f_\theta(\theta) g_\theta(\tilde{\theta}_{-1} = \tilde{\theta}) d\theta + \tilde{\theta} f_\theta(\tilde{\theta}) P_\theta(\tilde{\theta}_{-1} \le \tilde{\theta}) \right) = b(m)\tilde{\theta} ,$$

where the subscript θ indicates the transformed (from x to θ) probability, or probability density, functions, and

$$b(m) = m^{-1}\left((1-\alpha)(m-1)+1\right).$$

Here, $\tilde{\theta}_{-1}$ is the maximum of access rates to that document for other $m - 1$ users.

Now, we can simplify,

$$w_1 = \frac{n!}{(n-r)!} \frac{Z_p Z}{H_0} b(m) \int_0^{\infty} g(\tilde{x}_1) d\tilde{x}_1 \cdots \int_{\tilde{x}_{r-1}}^{\infty} g(\tilde{x}_r) d\tilde{x}_r \sum_{i=1}^{r} \tilde{\theta}_i \left(P(\tilde{\theta} \le \tilde{\theta}_r) \right)^{n-r} ,$$

where the integral over documents outside the proxy cache is evaluated, using the explicit expressions given by (8) and (9) in the paper. Furthermore, we make a substitution,

$$y_i = P(\tilde{x} \ge \tilde{x}_i) = P_\theta(\tilde{\theta} \le \tilde{\theta}_i) .$$

This yields,

$$\begin{aligned} w_1 &= \frac{n!}{(n-r)!} \frac{Z_p Z}{\beta H_0} b(m) \int_0^1 dy_1 \cdots \int_0^{y_{r-1}} dy_r \sum_{i=1}^{r} y_i^{\frac{\alpha}{m(1-\alpha)}} y_r^{\,n-r} \\ &= \frac{Z_p Z}{n(1-\alpha)} b(m) Factor_P , \end{aligned}$$

where

$$Factor_P = \sum_{i=1}^{r} \prod_{j=n-r+i}^{n} \left(1 + \frac{\alpha}{jm(1-\alpha)} \right)^{-1} .$$

The expected delay due to original server, w_2, can be written as,

$$\begin{aligned} w_2 = n! \int_0^{\infty} g(\tilde{x}_1) d\tilde{x}_1 \int_{\tilde{x}_1}^{\infty} g(\tilde{x}_2) d\tilde{x}_2 \cdots \int_{\tilde{x}_{r-1}}^{\infty} g(\tilde{x}_r) d\tilde{x}_r \left(P(\tilde{x}_{-1} \ge \tilde{x}_r) \right)^{n-r} \int_{\tilde{x}_r}^{\infty} f(x_{r+1}) dx_{r+1} \\ \times \int_{x_{r+1}}^{\infty} f(x_{r+2}) dx_{r+2} \cdots \int_{x_{n-1}}^{\infty} f(x_n) dx_n \sum_{i=r+R+1}^{n} \frac{\theta(x_i) Z}{H_0} , \end{aligned}$$

where we substitute in the expression of $F_2(x, \tilde{x} \ge \bar{x})$, i.e., (10) in the paper. We can further simplify,

$$w_2 = w(n-r, R)\left(1-\frac{r}{n}\right)Factor_{\text{G-LRU}},$$

where,

$$Factor_{\text{G-LRU}} = \frac{n!}{(n-r)!}\int_0^1 d(\beta\theta_1)^{m(1-\alpha)/\alpha}\cdots\int_0^{\beta\theta_{r-1}} d(\beta\theta_r)^{m(1-\alpha)/\alpha}(\beta\theta_r)^{m(1-\alpha)(n-r)/\alpha+1}$$
$$= \frac{n!}{(n-r)!}\prod_{i=n-r+1}^{n}\left(i+\frac{\alpha}{m(1-\alpha)}\right)^{-1} = \prod_{i=n-r+1}^{n}\left(1+\frac{\alpha}{im(1-\alpha)}\right)^{-1}.$$

A4. Approximations

In the limit where n and $n - r$ are large, the following approximation holds,

$$\prod_{k=n-r+1}^{n}\left(1+\frac{q}{k}\right)^{-1} = \frac{\Gamma(n+1+q)\Gamma(n-r+1)}{\Gamma(n+1)\Gamma(n-r+1+q)} \approx \left(1-\frac{r}{n}\right)^{q},$$

where $\Gamma(z)$ is the Gamma function. To prove the above expression, we make use of the following approximation of the Gamma function,

$$\ln\Gamma(z) \approx z\ln z - z - \ln z/2,$$

for large z. The error term is of the order n^{-1}.

We have,

$$Factor_{\text{G-LRU}} = \prod_{i=n-r+1}^{n}\left(1+\frac{\alpha}{im(1-\alpha)}\right)^{-1} \approx \left(1-\frac{r}{n}\right)^{\frac{\alpha}{m(1-\alpha)}};$$

and,

$$Factor_P = \sum_{i=1}^{r}\prod_{j=n-r+i}^{n}\left(1+\frac{\alpha}{jm(1-\alpha)}\right)^{-1}$$
$$\approx \sum_{i=1}^{r}\left(1-\frac{i}{n}\right)^{\frac{\alpha}{m(1-\alpha)}} \approx \frac{nm(1-\alpha)}{\alpha+m(1-\alpha)}\left(1-\left(1-\frac{r}{n}\right)^{1+\frac{\alpha}{m(1-\alpha)}}\right).$$

The expected delay w_1 can be approximated as,

$$w_1 = \frac{Z_p Z}{n(1-\alpha)}\frac{(1-\alpha)(m-1)+1}{m}Factor_P$$
$$\approx \frac{Z_p Z}{n(1-\alpha)}\frac{(1-\alpha)m+\alpha}{m}\frac{nm(1-\alpha)}{\alpha+m(1-\alpha)}\left(1-\left(1-\frac{r}{n}\right)^{1+\frac{\alpha}{m(1-\alpha)}}\right)$$
$$= Z_P Z\left(1-\left(1-\frac{r}{n}\right)^{1+\frac{\alpha}{m(1-\alpha)}}\right).$$

A5. Evaluation of *L*-policy

Full Duplication. Given a value of j, the conditional expectation of the binary variable φ_k is,

$$\boldsymbol{E}_{\varphi;j} \circ \varphi_k = \frac{r-j}{n-R},$$

for $k = R+1, \cdots, n$. All $(n-R)$ documents are equally likely to be tagged. It can be shown that,

$$\sum_{j=0}^{R} p_j^d \frac{r-j}{n-R} = \frac{r}{n}.$$

Substituting in the expected value of φ_k and carrying out the integrals, we get,

$$\sum_{j=0}^{R} p_j^d \left(\frac{r-j}{n-R} Z_p Z + \frac{n-R-r+j}{n-R} Z \right) \frac{1}{(1-\alpha)n} \sum_{k=R+1}^{n} \prod_{l=n-k+1}^{n} \left(1 + \frac{\alpha}{l(1-\alpha)} \right)^{-1}.$$

After averaging over j, the expected delay becomes,

$$\left(\frac{r}{n} Z_p Z + \left(1 - \frac{r}{n} \right) Z \right) \frac{n-R}{n} \prod_{l=n-R+1}^{n} \left(1 + \frac{\alpha}{l(1-\alpha)} \right)^{-1}.$$

This is nothing but the expression for our base case.

No Duplication. Similarly, for a given a value of i, the conditional expectation of the binary variable φ_k is,

$$\boldsymbol{E}_{\varphi;i} \circ \varphi_k = \begin{cases} \dfrac{i}{R+i-1}, & \text{if } k \le R+i-1; \\ \dfrac{r-i}{n-R-i}, & \text{if } k \ge R+i+1. \end{cases}$$

The $(R+i)^{th}$ document is not tagged. One can verify that,

$$\sum_{i=0}^{r} p_i^{nd} \frac{i}{R+i-1} = \frac{r}{n} \text{ and } \sum_{i=0}^{r} p_i^{nd} \frac{r-i}{n-R-i} = \frac{r}{n}.$$

The expected delay can be reduced to, after evaluating the integrals,

$$\sum_{i=0}^{r} p_i^{nd} \frac{i}{R+i-1} Z_p Z \left(1 - \frac{n-R-i+1}{n} \prod_{l=n-R-i+2}^{n} \left(1 + \frac{\alpha}{l(1-\alpha)} \right)^{-1} \right)$$

$$+ \sum_{i=0}^{r} p_i^{nd} \left(\frac{r-i}{n-R-i} Z_p Z + \frac{n-R-r}{n-R-i} Z \right) \frac{n-R-i}{n} \prod_{l=n-R-i+1}^{n} \left(1 + \frac{\alpha}{l(1-\alpha)} \right)^{-1}.$$

We carry out the summation over i. This gives,

$$\frac{r}{n} Z_p Z + \frac{n-R-r}{n} \sum_{i=0}^{r} p_i^{nd} \prod_{l=n-R-i+1}^{n} \left(1 + \frac{\alpha}{l(1-\alpha)} \right)^{-1} Z.$$

Partial Duplication. We start with the probability of j tagged documents for the first L documents, $p(j)$,

$$p(j) = \sum_{i=0}^{r-j} p(j,i) = \binom{L}{j}\binom{n-L}{r-j}\binom{n}{r}^{-1}.$$

This gives the conditional probability of i tagged documents for documents ranked from the $(L+1)^{th}$ to the $(R+i-1)^{th}$,

$$p(i \mid j) = \frac{p(j,i)}{p(j)} = \binom{R-L+i-1}{i}\binom{n-R-i}{r-j-i}\binom{n-L}{r-j}^{-1}.$$

The $(R+i)^{th}$ document should not be tagged. Similarly, for given j and i, each document has a probability of,

$$\boldsymbol{E}_{\varphi;j,i} \circ \varphi_k = \begin{cases} \dfrac{i}{R-L+i-1}, & \text{if } L+1 \le k \le R+i-1; \\ \dfrac{r-j-i}{n-R-i}, & \text{if } k \ge R+i+1. \end{cases}$$

being tagged. We also have,

$$\sum_{i=0}^{r-j} p(i \mid j)\frac{i}{R-L+i-1} = \frac{r-j}{n-L} \text{ and } \sum_{i=0}^{r-j} p(i \mid j)\frac{r-j-i}{n-R-i} = \frac{r-j}{n-L}.$$

For given j and i, the expected delay can be expressed as,

$$\frac{i}{L-R+i-1} Z_p Z\left(\frac{n-L}{n}\prod_{l=n-L+1}^{n}\left(1+\frac{\alpha}{l(1-\alpha)}\right)^{-1} - \frac{n-R-i+1}{n}\prod_{l=n-R-i+2}^{n}\left(1+\frac{\alpha}{l(1-\alpha)}\right)^{-1}\right)$$

$$+\left(\frac{r-j-i}{n-R-i} Z_p Z + \frac{n-R-r-j}{n-R-i} Z\right)\frac{n-R-i}{n}\prod_{l=n-R-i+1}^{n}\left(1+\frac{\alpha}{l(1-\alpha)}\right)^{-1}.$$

We average the above expression over i. This yields,

$$\frac{r-j}{n}\prod_{l=n-L+1}^{n}\left(1+\frac{\alpha}{l(1-\alpha)}\right)^{-1} Z_p Z + \frac{n-R-r+j}{n}\sum_{i=0}^{r-j} p(i \mid j)\prod_{l=n-R-i+1}^{n}\left(1+\frac{\alpha}{l(1-\alpha)}\right)^{-1} Z,$$

which can be further reduced, after averaging over j, to,

$$\frac{r}{n}\frac{n-L}{n}\prod_{l=n-L+1}^{n}\left(1+\frac{\alpha}{l(1-\alpha)}\right)^{-1} Z_p Z + \sum_{j=0}^{L}\sum_{i=0}^{r-j} p(j,i)\frac{n-R-r+j}{n}\prod_{l=n-R-i+1}^{n}\left(1+\frac{\alpha}{l(1-\alpha)}\right)^{-1} Z.$$

A6. Approximation for Equation (35)

If the ratios r/n and R/n are small, we can expand

$$w(n, R+i) \approx 1 - \frac{1}{1-\alpha}\frac{R+i}{n} + \frac{1}{2}\frac{\alpha}{(1-\alpha)^2}\left(\frac{R+i}{n}\right)^2.$$

To carry out the summation in Equation (35), we utilize the following results,

$$\sum_{j=0}^{L}\sum_{i=0}^{r-j} p(j,i)\frac{n-R-r+j}{n-R-i} = \frac{n-r}{n};$$

$$\sum_{j=0}^{L}\sum_{i=0}^{r-j} p(j,i)\frac{n-R-r+j}{n-R-i}\frac{i}{n} = \frac{r(R-L)}{n^2};$$

$$\sum_{j=0}^{L}\sum_{i=0}^{r-j} p(j,i)\frac{n-R-r+j}{n-R-i}\left(\frac{i}{n}\right)^2 = \frac{r(R-L)\big(n-R+L+r(R-L)\big)}{n^3(n-r+1)}.$$

(36) in the paper can be obtained by simplifying and collecting terms together.

A7. Pseudo Code for Simulation

STEP 1: INITIALIZE

1.1. Populate
- A 2D array $DocAgeAll(i, j)$ with age uniformly on [0, 1] for user i and document j
- Two 1D arrays with $size$ and $delay$ log-normally distributed for each document

1.2. Store in an array $GLRUList$ the indices of all documents according to sorted (small to large) GLRU ages

1.3. Store in a 2D array $LRUAges$ the indices of all documents according to sorted (small to large) LRU ages for each user

STEP 2: GENERATE INTER-EVENT TIME τ_k FOR NEXT ACCESS k

2.1. Generate $\Delta x_{i,j}$ user i and document j using LRU age from $DocAgeAll(i, j)$

2.2. Find inter-event time $\tau_k = \min(\Delta x_{i,j})$, and the user and document for that access: $(i_{\min,}\ j_{\min}) = \arg\min_{i,j}(\Delta x_{i,j})$

STEP 3: UPDATE CACHING SYSTEM

3.1. Fill proxy cache $proxycache$ up to its capacity from $GLRUList$

3.2. Fill user $i_{\min}$'s browser cache $localcache$ up to its capacity from $LRUAges$

3.3. IF $localcache$ contains $j_{\min}$, or a local cache hit, THEN

$localhit = localhit + \tau_k$

ELSE

IF $proxycache$ contains $j_{\min}$, or a proxy cache hit, THEN

$proxyhit = proxyhit + \tau_k$

$browserdelay = browserdelay + delay(j_{\min})\tau_k Z_p$

ELSE, or a cache miss,

$browserdelay = browserdelay + delay(j_{\min})\tau_k$

END IF

Update $GLRUList$

END IF

3.4. Increase $DocAgeAll(i, j)$ by τ_k

3.5. Reset $DocAgeAll(i_{\min}, j_{\min})$ to 0

3.6. Update $LRUAges$ for user $i_{\min}$

STEP 4: COLLECT STATISTICS

4.1. Repeat Steps 2 and 3 until the steady state

4.2. Reset $localhit$ = 0, $browserdelay$ = 0, and $proxyhit$ = 0

4.3. Repeat Steps 2 and 3

4.4. Normalize $localhit$, $browserdelay$, and $proxyhit$ by the total time elapsed