On Sample-Path Staleness in Lazy Data Replication

Xiaoyong Li, Daren B.H. Cline, and Dmitri Loguinov*
Texas A&M University, College Station, TX 77843, USA
Email: xiaoyong@cse.tamu.edu, dcline@stat.tamu.edu, dmitri@cs.tamu.edu

***Abstract*—We analyze synchronization issues arising between two stochastic point processes, one of which models data churn at an information source and the other periodic downloads from its replica (e.g., search engine, web cache, distributed database). Due to lazy (pull-based) synchronization, the replica experiences recurrent staleness, which translates into some form of penalty stemming from its reduced ability to perform consistent computation and/or provide up-to-date responses to customer requests. We model this system under non-Poisson update/refresh processes and obtain sample-path averages of various metrics of staleness cost, generalizing previous results and exposing novel problems in this field.**

I. Introduction

With the massive growth of the Internet and deployment of large-scale distributed applications, mankind faces new challenges in acquiring, processing, and maintaining vast amounts of data. In response to this flood of information, companies deploy cloud-based solutions designed to provide replicated and distributed support to the skyrocketing storage and processing demand of their users.

One problem in these applications is the *highly-dynamic* nature of content, especially when perfect synchronization of sources, replicas, intermediate caches, and various computation is impossible. In fact, many large-scale distributed systems (e.g., airline reservations, online banking, web search engines, social networks) operate under constant data churn and may never see consistent snapshots of the entire network.

In traditional databases, the source opens outbound communication with the replicas whenever it detects important information changes. This enables *push-based* operation that actively expires stale content and broadcasts notifications into the system. In other cases, however, scalability and administrative autonomy require that sources operate independently and provide information only based on explicit request.

This *pull-based* replication (also called *optimistic* or *lazy*) improves both scalability of the service and availability of the data, but at the expense of increased age of manipulated content [17], [35]. This model of operation has enjoyed ubiquitous deployment in the current Internet (e.g., HTTP, DNS, network monitoring, web caching, RSS feeds, stock-ticker aggregators, certain types of CDNs, sensor networks); however, it still poses many fundamental modeling challenges. Our goal is to study them in this paper.

A. Contributions

Consider a single source driven by an update process N_U and a single replica with the corresponding download process N_D, which is independent of N_U. Our first contribution is to propose a general framework for modeling staleness under arbitrary stochastic processes (N_U, N_D), in contrast to prior work that has only considered Poisson cases [4], [5], [7], [8], [9], [10], [12], [13], [14], [16], [18], [21], [22], [24], [25], [26], [27], [34]. Since staleness age and various penalties derived from it are usually defined in terms of sample-path averages, questions arise about their existence and possible variation across multiple realizations of the system. We address this issue by identifying the weakest set of conditions for which the distribution of staleness age exists and converges to the same deterministic value in every sample path.

Armed with these results, our second contribution is to model interaction between the age processes of N_U and N_D. We show that for the results to be tractable, ages of the two processes examined at random times within a given sample path must be *independent* of each other. Interestingly, this condition does not automatically follow from independence of N_U and N_D. Instead, we show that it translates into a form of ASTA (Arrivals See Time Averages) [23], where the download process N_D must observe the sample-path distribution of update age.

Under the condition of age-independence, our third contribution is to derive the distribution of time by which the replica trails the source, the fraction of consumers that encounter a stale copy, the average number of missing updates from the replica at query time, and the general staleness cost under all suitable penalty functions $w(x)$. Our results involve simple closed-form expressions that are functions of limiting age distributions of both processes.

Our fourth contribution is to analyze conditions under which N_D produces provably optimal penalty for a given download rate. We show that penalty reduces if and only if inter-refresh delays become stochastically larger in second order. This leads to constant synchronization delays being optimal under all N_U and $w(x)$. This, however, presents problems in satisfying ASTA and creates a possibility of worst-case (i.e., 100%) staleness due to phase-lock between the source and the replica. To this end, we discuss general requirements for ensuring that N_D avoids these drawbacks while remaining optimal, or at least close to it.

We finish the paper with our last contribution that considers the practical aspects of staleness, including experimentation with Wikipedia page updates, error analysis of previous Poisson models, estimation of search-engine bandwidth requirements to maintain certain freshness, and generalization to multiple sources/replicas.

*Supported by NSF grants CNS-1017766 and CNS-1319984.

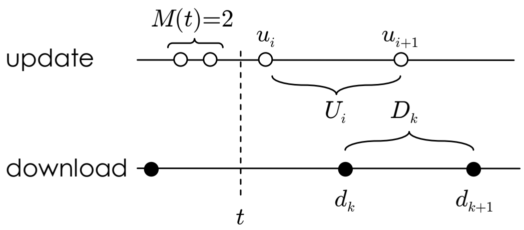
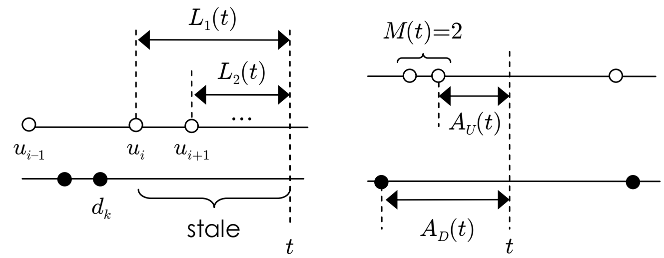


Fig. 1. Process notation.



(a) staleness lags (b) age

Fig. 2. Penalty lags and process age.

II. Staleness Formulation

We start by explaining the underlying assumptions on the system, defining the various processes that determine information flow, and specifying the metrics of interest. The omitted proofs can be found in technical report [19].

A. Updates and Synchronization

We next model interaction between a single source and a single replica, which is a prerequisite to understanding system performance. Suppose the source undergoes updates at random times $0 = u_1 < u_2 < \dots$ and define $N_U(t) = \max\{i : u_i \leq t\}$ to be a stochastic process that counts the number of updates in $[0, t]$. When referring to the entire process, rather than its value at some point, we omit t and write simply N_U.

For the replica, denote its random download instances by $0 = d_1 < d_2 < \dots$ and the corresponding point process by $N_D(t) = \max\{k : d_k \leq t\}$. This formulation neglects processing delays and treats all events as instantaneous. We additionally assume that both processes are simple (i.e., at most one point at any t) and independent of each other. Now, suppose the inter-update delays of N_U are given by a random process $\{U_i\}_{i=1}^{\infty}$ and those of N_D by $\{D_k\}_{k=1}^{\infty}$, which are illustrated in Fig. 1. Each of these sequences may be of fairly general nature, e.g., correlated and/or non-stationary.

B. Cost of Staleness

To understand the penalty of outdated content, suppose $M(t)$ counts the number of updates *missing* from the replica at time t (e.g., in Fig. 1, $M(t) = 2$). This is a discrete-state process that increments for each update and resets to zero for each synchronization.

Definition 1: A replica is called *stale* at time t if $M(t) > 0$. Otherwise, it is called *fresh*.

From the consumer's perspective, stale material reduces user satisfaction and lowers system performance, which needs to be translated into a cost metric that can be expressed via some known parameters of the system. The most basic penalty is the probability that the replica is stale at the time of request, i.e., $P(M(t) > 0)$. The second obvious metric is the expected number of missing updates $E[M(t)]$, which measures the amount of lost information during a crash and estimates the difficulty in recreating it from the most recent checkpoint.

More sophisticated cases are also possible. Suppose the source runs some computation, with updates representing certain intermediate states that are written to disk. A crash at time t requires computation to be restarted, which means that the penalty is determined not by $M(t)$, but rather by the *duration* of the computation that was lost due to staleness.

Definition 2: For a stale replica at time t, define lags $L_1(t) > L_2(t) > \dots > L_{M(t)}(t)$ to be backward delays to each unseen update.

This concept is illustrated in Fig. 2(a) for the first two lags. To keep the model general and cover the various options already seen in the literature [4], [5], [7], [8], we assume that the consumer is sensitive to either just lag $L_1(t)$, i.e., how long the *source* has been stale at time t, or the entire collection of lags $\{L_1(t), \dots, L_{M(t)}(t)\}$, i.e., how long each uncaptured *update* has been stale. Since it is usually difficult to predict the value of information freshness to each customer, one requires a mapping from staleness lags to actual cost, which we assume is given by some non-negative weight function $w(x)$.

Definition 3: At time t, the *source penalty* is given by the weight of the delay since the replica was fresh last time:

$$\eta(t) = \begin{cases} w(L_1(t)) & M(t) > 0 \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

while the *update penalty* is given by the aggregate weight of all staleness lags:

$$\rho(t) = \begin{cases} \sum_{i=1}^{M(t)} w(L_i(t)) & M(t) > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

For example, $w(x) = 1$ produces the first two metrics discussed above, i.e., $P(M(t) > 0)$ via $E[\eta(t)]$ and $E[M(t)]$ via $E[\rho(t)]$. Both (1) and (2) are random variables, which suggests that system performance should be assessed by their average values. But as neither N_U nor N_D is assumed to be stationary, the expected penalty requires additional elaboration. Instead of considering $E[\eta(t)]$ and $E[\rho(t)]$, which may depend on time t, it is more natural to replace them with sample-path averages [8]:

$$\bar{\eta} = \lim_{T\to\infty} \frac{1}{T}\int_0^T \eta(t)dt \quad \text{and} \quad \bar{\rho} = \lim_{T\to\infty} \frac{1}{T}\int_0^T \rho(t)dt, \tag{3}$$

where consumers are modeled as being equally likely to query the replica at any time in $[0, \infty)$.

C. Relationship to Prior Work

The majority of the literature on source penalty $\bar{\eta}$ is limited to Poisson N_U, either constant or exponential D, and $w(x) =$

1 or x [6], [7], [8], [10], [24], [28], [31], [34]. There has been only one attempt to model $\bar{\eta}$ under a general renewal process N_U, in which [31] assumed $w(x) = 1$ and the entire sequence of refresh instances $\{d_1, d_2, \ldots\}$ was known. While appropriate in some cases, this model is difficult to evaluate in practice when N_D is given by its statistical properties.

Update penalty $\bar{\rho}$ has received less exposure, with almost all papers considering Poisson updates and just constant D. This includes $w(x) = 1$, where $\bar{\rho}$ is usually called *divergence* [16] or *blur* [11], with analysis available in [12], [13], [26], and $w(x) = x$, where $\bar{\rho}$ is known as *additive age* [20], *aggregated age* [21], *delay* [26], or simply *cost* [12]. Finally, $\bar{\rho}$ with a general $w(x)$ was called *obsolescence cost* in [13] and analyzed under a non-stationary Poisson N_U, but no closed-form results were obtained.

The Poisson assumption on N_U allows easy computation of the various metrics of interest. Outside these special cases, superposition of non-memoryless processes produces much more complex behavior.

III. Age Model

While (3) is a convenient approach, there is a previously unnoticed obstacle with using it. Observe that (3) defines limits of sequences of random variables; however, it is unclear whether these limits exist, if they are finite, and under what conditions they are deterministic across all sample paths.

A. Main Framework

We start by performing a convenient transformation of (3) to remove the integrals. Define Q_T to be a uniform random variable in $[0, T]$, which models the random query time of consumers. Suppose Q_T is independent of N_U and N_D, in which case (3) is the limit of $E[\eta(Q_T)|N_U, N_D]$ and $E[\rho(Q_T)|N_U, N_D]$ as $T \to \infty$. We explicitly condition on processes N_U, N_D to emphasize that all expectations and probabilities involving Q_T are random variables (i.e., dependent on the pair of sample paths).

At each time t, suppose age processes $A_U(t)$ and $A_D(t)$, shown in Fig. 2(b), specify delays to the previous update and synchronization event, respectively. Using this notation and observing that $M(t) > 0$ is equivalent to $A_U(t) < A_D(t)$, define an ON/OFF staleness process:

$$S(t) = \begin{cases} 1 & A_U(t) < A_D(t) \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

whose properties at random time Q_T determine whether the consumer sees outdated information or not.

B. Assumptions

We next aim to establish a minimal set of conditions under which analysis of staleness admits closed-form results. Consider a general point process N with cycle lengths $\{X_i\}_{i=1}^{\infty}$, where each $X_i \sim F_i(x)$ is a random variable. In order for the age $A(Q_T)$ of this process to have a usable limiting distribution as $T \to \infty$, one must impose three constraints on N, which we discuss informally and motivate next, followed by a more rigorous, but functionally equivalent, definition.

The first restriction is that collection $\{X_i\}_{i=1}^{\infty}$ within each sample-path have some limiting distribution $F(x)$. The second prerequisite is that $F(x)$ be deterministic (i.e., equal in all sample-paths). Finally, the third condition is that an $o(1)$ fraction of cycles in $\{X_i\}_{i=1}^{n}$ not consume $\Omega(1)$ fraction of length as $n \to \infty$. This would be a problem because $F(x)$, being a limiting distribution, does not capture these intervals, but Q_T still lands there with a non-diminishing probability.

Let $\mathbf{1}_A$ be an indicator variable of A and $\bar{F}(x) = 1 - F(x)$ the complementary CDF (cumulative distribution function) of $F(x)$. We are now ready to summarize our discussion.

Definition 4: A process N is called *age-measurable* if:

1) For all $x \geq 0$, except possibly points of discontinuity of the limit, sample-path distribution $H_n(x)$ of variables $\{X_1, \ldots, X_n\}$ converges in probability as $n \to \infty$:

$$H_n(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i \leq x} \xrightarrow{P} F(x); \tag{5}$$

2) Function $F(x)$ is deterministic with mean $0 < \delta < \infty$;
3) The average cycle length converges to δ in probability as $n \to \infty$:

$$Z_n := \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{P} \delta = \int_0^{\infty} \bar{F}(x) dx. \tag{6}$$

Note that any renewal process $\{X_i\}_{i=1}^{\infty}$ satisfies this definition since all $F_i(x)$ are the same, which from the weak law of large numbers trivially leads to $F(x) = F_i(x)$ and $Z_n \to \delta$.

C. Distribution

Armed with the definition above, the main result of this section follows.

Theorem 1: For an age-measurable process N, the sample-path distribution of its age $A(Q_T)$ in points Q_T uniformly distributed in $[0, T]$ converges in mean as $T \to \infty$ to the residual distribution of $F(x)$:

$$G(x) := \lim_{T \to \infty} P(A(Q_T) < x | N) = \frac{1}{\delta} \int_0^x \bar{F}(y) dy. \tag{7}$$

Interestingly, (5)-(6) are not only sufficient as demonstrated by this theorem, but also necessary for $G(x)$ to exist and equal the right side of (7). Necessity is proven by well-known counter-examples in probability theory [30].

D. Expectation

While Theorem 1 establishes when $A(Q_T)$ has a limiting distribution, convergence of expectation $E[A(Q_T)|N]$ or suitability of $G(x)$ for computing it are not guaranteed. Furthermore, given that consumers may apply generic weights $w(x)$ to the various age-related metrics, it makes sense to ask when $E[w(A(Q_T))|N]$ exists as $T \to \infty$.

To build intuition for the next result, assume $X \sim F(x)$ is a non-negative variable and define its age A to be a random variable with CDF $G(x)$ in (7). Then, we are interested in the relationship between $E[w(A)]$ and X. To this end, suppose for any locally integrable function $w(x)$, we set $w_1(x) = w(x)$ and then recursively integrate the result $n-1$ times to define:

$$w_n(x) := \int_0^x w_{n-1}(y)dy. \quad (8)$$

Using integration by parts in Lebesgue-Stieltjes integrals and keeping in mind that $w_{n+1}(0) = 0$ for $n \geq 1$:

$$E[w_{n+1}(X)] = \int_0^\infty w_n(x)\bar{F}(x)dx = E[w_n(A)]E[X]. \quad (9)$$

Therefore, in order for $E[w(A)]$ to exist, one must ensure that both $E[w_2(X)]$ and $E[X]$ do. Note that the latter does so by (6), but the former requires an additional constraint.

Definition 5: A point process N is called *age-measurable by weight function* $w(x)$ if it is age-measurable and

$$\frac{1}{n}\sum_{i=1}^n w_2(X_i) \xrightarrow{P} \int_0^\infty w_2(x)dF(x) < \infty. \quad (10)$$

Note that age-measurable by a constant is equivalent to simply age-measurable since in that case (10) becomes (6). The next result is similar to Theorem 1, except convergence is weaker since $w(x)$ may be unbounded.

Theorem 2: For a process N that is age-measurable by $w(x)$, the sample-path expectation of $w(A(Q_T))$ converges in probability as $T \to \infty$:

$$\lim_{T\to\infty} E[w(A(Q_T))|N] = \int_0^\infty w(x)dG(x). \quad (11)$$

From this point on, we omit explicit conditioning on the sample-path since results do not depend on N for age-measurable processes. However, we keep in mind that all probabilities and expectations involving Q_T are still taken in the sample-path sense.

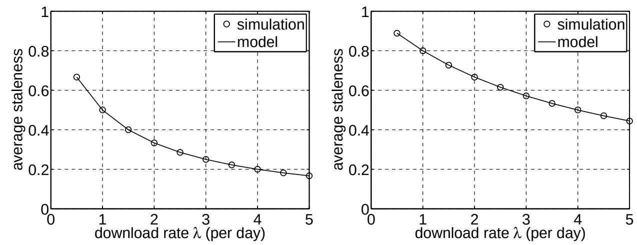
IV. Staleness Cost

This section models the probability of staleness and expected cost under both penalty metrics defined earlier.

A. Age Independence

We now return to examining (4). In order to determine when the replica is stale, one requires comparison of $A_U(Q_T)$ with $A_D(Q_T)$, which may not be independent random variables, even if N_U and N_D are. For example, suppose $\{U_i\}_{i=1}^\infty$ and $\{D_k\}_{i=1}^\infty$ are iid variables that equal 1 or 2 with probability $1/2$ each. As $T \to \infty$, the distribution of $A_D(Q_T)$ becomes a mixtures of two uniform variables in $[0,1]$ and $[0,2]$. However, conditioning on $A_U(Q_T) = y$ shifts the mass of refresh age $A_D(Q_T)$ to just three discrete points $y-1, y, y+1$, clearly showing that the two ages are dependent.

To prevent such cases, which is called *phase-lock* [3], the safest solution is to require that N_D implement a download policy that ensures independence of the two ages. In that case, conditions such as ASTA (Arrivals See Time Averages) [23] must apply to the age of one process when sampled by the arrival points of the other. This issue is delayed until a later section, but for the time being we define more clearly what independence of $A_U(Q_T)$ and $A_D(Q_T)$ means.



(a) Pareto U, constant D (b) constant U, Pareto D

Fig. 3. Examination of (14) under $\mu = 2$.

Specifically, suppose N_U and N_D are age-measurable. Then, let $F_U(x)$ and $F_D(x)$ be respectively the limiting CDF functions of interval lengths defined in (5), with the corresponding average rates μ and λ, i.e.,

$$\frac{1}{\mu} = \int_0^\infty \bar{F}_U(x)dx \quad \text{and} \quad \frac{1}{\lambda} = \int_0^\infty \bar{F}_D(x)dx. \quad (12)$$

Further, let $U \sim F_U(x)$ and $D \sim F_D(x)$ be random update and download cycle lengths. Similarly, suppose $G_U(x)$ and $G_D(x)$ are the limiting CDFs of age from (7), with lower-case functions $g_U(x)$ and $g_D(x)$ representing the corresponding PDFs. When the necessary limits exist, let $A_U \sim G_U(x)$ and $A_D \sim G_D(x)$ represent the two random ages as $T \to \infty$.

Definition 6: Two point processes N_U and N_D are called *age-independent* if they are age-measurable and $\forall x, y \geq 0$:

$$\lim_{T\to\infty} P(A_D(Q_T) < x | A_U(Q_T) = y) = G_D(x). \quad (13)$$

If either N_U or N_D is Poisson, (13) is guaranteed from PASTA (Poisson Arrivals See Time Averages) [32], which explains why prior work did not encounter these nuances.

B. Preliminaries

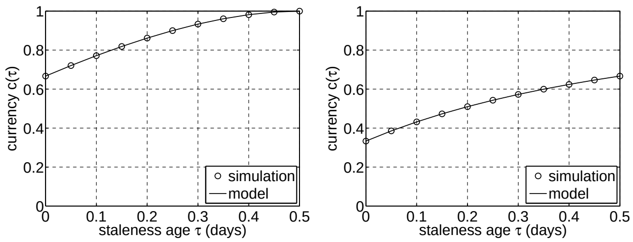
Our first objective is to derive the probability of staleness.

Theorem 3: Assuming that N_U and N_D are age-independent, the probability of staleness at time Q_T converges in mean as $T \to \infty$ to:

$$P(S(Q_T) = 1) \to p := \mu \int_0^\infty \bar{F}_U(y)\bar{G}_D(y)dy. \quad (14)$$

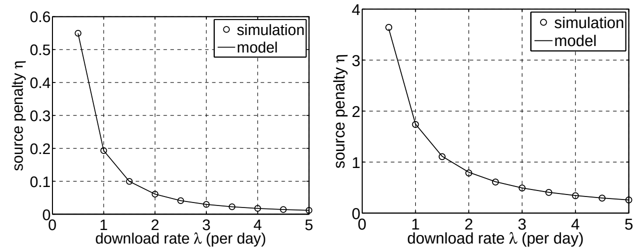
To perform a self-check against prior results with Poisson N_U, observe that (14) simplifies to $p = 1 - \lambda(1 - e^{-\mu/\lambda})/\mu$ under constant D and $\mu/(\mu+\lambda)$ under exponential D, which are consistent with [8], [10]. Simulations in Fig. 3 examine model accuracy in more interesting cases of general renewal processes. We use Pareto CDF $1 - (1 + x/\beta)^{-\alpha}$ with $\alpha = 3$ and mean $\beta/(\alpha - 1) = \beta/2$. Observe in the figure that the model matches simulations very well, with constant download intervals performing significantly better against Pareto update cycles in (a) than the other way around in (b). For example, synchronizing pages at their update rate (i.e., $\lambda = \mu = 2$) serves stale copies with probability 33% in the former case and 66% in the latter. Furthermore, for the same p, the scenario in (a) requires roughly 4 times less bandwidth than in (b).

The next intermediate result is the distribution of the first lag $L_1(Q_T)$, which relies on p in (14).



(a) Pareto U, constant D (b) constant U, Pareto D

Fig. 4. Examination of (16) under $\lambda = \mu = 2$.



(a) Pareto U, constant D (b) constant U, Pareto D

Fig. 5. Examination of (17) under $w(x) = x, \mu = 2$.

Theorem 4: If N_U and N_D are age-independent, the CDF of $L_1(Q_T)$ converges in mean as $T \to \infty$ to:

$$F_L(x) = 1 - \frac{\mu}{p} \int_0^\infty \bar{F}_U(y) \bar{G}_D(x+y) dy. \quad (15)$$

Theorem 4 allows a simple expression for the fraction of requests $c(\tau)$ that observe content outdated by less than τ time units, which was called β-currency in [4] and Δ-consistency in [28]. This can be expressed as:

$$c(\tau) = 1 - \bar{F}_L(\tau)p = \int_0^\infty g_U(y) G_D(\tau + y)) dy, \quad (16)$$

which conveniently simplifies to $P(A_D - A_U < \tau)$, where $A_D - A_U$ is the generalized lag between the replica and the source, i.e., non-positive values mean fresh states. Fig. 4 compares (16) to simulations using $\lambda = \mu$. As seen in the figure, this page retrieved at a random time is stale by less than $\tau = 0.4$ days (9.6 hours) with probability $c(\tau) = 98\%$ in the first case and 62% in the second.

C. Source Penalty

We are now ready to derive a general formula for $\bar{\eta}$.

Theorem 5: If N_U and N_D are age-independent, while N_D is age-measurable by $w(x)$, the source penalty converges in probability to:

$$\bar{\eta} = \lambda\mu \int_0^\infty \bar{F}_U(y) \int_0^\infty w(x) \bar{F}_D(x+y) dx dy. \quad (17)$$

With $w(x) = 1$, (17) reduces to staleness probability p already discussed above. For the other case $w(x) = x$ seen in the literature, we obtain the expected *staleness age* $\bar{\eta} = E[L_1(t)]p$ by which the replica trails the source. Under Poisson N_U and constant D, we get from (17):

$$\bar{\eta} = \frac{1}{2\lambda} - \frac{1}{\mu} + \frac{\lambda(1 - e^{-\mu/\lambda})}{\mu^2}, \quad (18)$$

and when both distributions are exponential:

$$\bar{\eta} = \frac{\mu}{\lambda(\lambda + \mu)}. \quad (19)$$

These special cases are consistent with [7]. Simulations in Fig. 5 additionally confirm that (17) is accurate under general renewal processes. Also observe in the figure that the combination in (b) continues to offer inferior performance to that in (a); however, the difference between the two scenarios is now more pronounced. For example, using the same $\lambda = \mu$ considered earlier, search-engine clients encounter indexing results outdated on average by 0.06 days (1.5 hours) in the left subfigure and by 0.8 days (19 hours) in the right. This example shows how drastically the cost changes based on the *shape* of $F_U(x)$ and $F_D(x)$, which emphasizes the importance of utilizing models that can accurately handle any underlying processes (N_U, N_D).

We now offer a more intuitive look at source penalty. Modifying $w(x)$ to be zero for negative x, we can rewrite (17) in a more compact form:

$$\bar{\eta} = E[w(A_D - A_U)] = \lambda E[w_2(D - A_U)]. \quad (20)$$

This result shows that $\bar{\eta}$ is determined by the *positive* deviation of the generalized lag $A_D - A_U$ from zero, or equivalently by that of $D - A_U$, where the weight applied to each deviation is given respectively by $w(x)$ and $w_2(x)$. The only caveat is that simplification (20) requires weight functions that can explicitly handle negative arguments, e.g., a constant penalty would be $w(x) = \mathbf{1}_{x \geq 0}$ rather than just $w(x) = 1$. Throughout the rest of the paper, we avoid the extra notation dealing with $x < 0$, but keep this in mind.

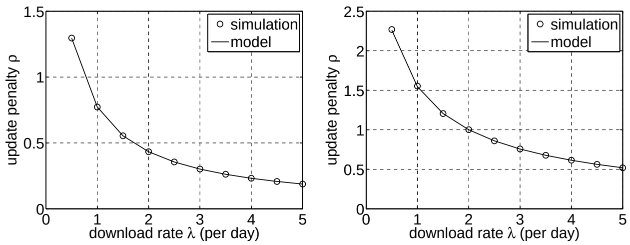
D. Update Penalty

Unlike the previous section, we next show that $\bar{\rho}$ admits a much simpler result that depends only on the mean update rate μ rather than the entire distribution $F_U(x)$. This was first observed through simulations in [12] for constant interval lengths D, but no explanation or extension to other $F_D(x)$ was offered.

Theorem 6: Assuming N_U and N_D are age-independent, while N_D is age-measurable by $w_2(x)$, the update penalty converges in probability to:

$$\bar{\rho} = \mu E[w_2(A_D)] = \lambda\mu E[w_3(D)]. \quad (21)$$

To perform a sanity check, consider Poisson N_U and constant D. Then, (21) produces $\bar{\rho} = \mu/(2\lambda)$ for $w(x) = 1$ and $\mu/(6\lambda^2)$ for $w(x) = x$, both of which match previously known results [20], [26], [34]. Generalizing these two cases to exponential D, we obtain from (21) respectively μ/λ and μ/λ^2. Interestingly, this shows that switching downloads from constant intervals to exponential *doubles* the number of missing updates and *sextuples* their combined age.



(a) Pareto U, constant D (b) constant U, Pareto D

Fig. 6. Examination of (24) and (25) under $\mu = 2$.

For $w(x) = 1$, a simple closed-form expression is possible for all D:

$$E[M(t)] = \frac{\lambda \mu E[D^2]}{2} = \frac{\mu}{2\lambda}\Big(1 + \lambda^2 Var[D]\Big). \qquad (22)$$

For example, Pareto D produces in (22):

$$E[M(t)] = \frac{\mu(\alpha - 1)}{\lambda(\alpha - 2)}, \qquad (23)$$

which for $\alpha = 3$ is quadruple that of constant D and double that of exponential D. Another peculiar case is $\alpha \to 2$, where $E[M(t)]$ tends to infinity regardless of N_U. In fact, the update process itself may exhibit $Var[U] = \infty$, but the expected number of updates by which the replica falls behind will still become unbounded as α approaches 2.

Since source penalty $\bar{\rho}$ sums up the ages of all missing updates, it allows usage of *decaying* functions $w(x)$ such that their integral is increasing. We demonstrate this effect using $w(x) = 1/(1 + x)$, for which $w_2(x) = \log(1 + x)$. This cost function increases rapidly for small x, but then becomes less sensitive to staleness as the age of replicated content grows. Since $w_3(x) = (1 + x)\log(1 + x) - x$, constant D yields:

$$\bar{\rho} = \mu[(\lambda + 1)\log(1 + 1/\lambda) - 1]. \qquad (24)$$

For $D \sim$ Pareto(α, β) with $\alpha = 3$ and $\beta = 2/\lambda$, we get:

$$\bar{\rho} = 2\mu \begin{cases} \frac{2\log(2/\lambda) - 2 + \lambda}{(\lambda - 2)^2} & \lambda \neq 2 \\ 0.25 & \lambda = 2 \end{cases}. \qquad (25)$$

Fig. 6 confirms that both models are accurate, with constant D enjoying a 60% lower penalty compared to Pareto.

V. Optimality

Motivated by (22) and consistently worse performance of Pareto D, the goal of this section is to understand the impact, if any, of $Var[D]$ on penalty and determine whether there exists an optimal distribution $F_D(x)$ that, for a fixed download budget λ, provably results in the lowest cost for all N_U and all suitable functions $w(x)$.

A. Stochastic Dominance

We start with general concepts from economics and game theory that are useful for understanding optimality. For two non-negative random variables $X \sim F_X(x)$ and $Y \sim F_Y(x)$, let their CDF difference be:

$$H(x) = F_Y(x) - F_X(x), \qquad (26)$$

whose generalization $H_n(x)$ is given by (8). Then, we have the following definition.

Definition 7: Variable X is said to stochastically dominate Y in n-th order, which we write as $X \geq_{st}^n Y$, if $H_n(x) \geq 0$ for all $x \in \mathbb{R}$.

This concept is important because desirable characteristics of D can be inferred from those of A_D, as shown next.

Lemma 1: Assume $E[X] = E[Y]$ and $n \geq 2$. Then, X stochastically dominates Y in n-th order, i.e., $X \geq_{st}^n Y$, iff the age of Y stochastically dominates the age of X in $(n-1)$-st order, i.e., $A_Y \geq_{st}^{n-1} A_X$.

As given by the next lemma, first-order stochastic dominance allows one to determine the relationship between expected utilities $E[w(X)]$ and $E[w(Y)]$. While it is possible to establish a more general version of this result using n-th order dominance, it would restrict $w(x)$ to a narrower class of functions and thus would be less useful in practice.

Lemma 2: Condition $X \geq_{st}^1 Y$ holds iff for all non-decreasing functions $w(x)$ it follows that $E[w(X)] \geq E[w(Y)]$.

B. Penalty Analysis

Returning to the topic of information staleness, our goal is to determine the condition under which both types of penalty can be reduced without changing the refresh rate λ. Define $\bar{\eta}(D_1)$ and $\bar{\eta}(D_2)$ to be the source penalties corresponding to random synchronization intervals D_1 and D_2. For the opposite problem, i.e., finding the worst update distribution for a given μ, define $\bar{\eta}(U_1)$ and $\bar{\eta}(U_2)$ to be the penalties that correspond to update intervals U_1 and U_2.

The next result shows that stochastic (rather than variance) ordering is needed to improve staleness penalty. Define $w(x)$ to be a *measure* if it is non-negative, non-decreasing, and right-continuous with $w(x) = 0$ for $x < 0$.

Theorem 7: For a given N_U and fixed download rate λ, $D_1 \geq_{st}^2 D_2$ iff $\bar{\eta}(D_1) \leq \bar{\eta}(D_2)$ for all measures $w(x)$. Similarly, with a given N_D and fixed μ, $U_1 \geq_{st}^2 U_2$ iff $\bar{\eta}(U_1) \geq \bar{\eta}(U_2)$ for all measures $w(x)$.

A similar result holds under update penalty $\bar{\rho}$. Note that $F_U(x)$ has no impact here and the result holds for all $w(x)$.

Theorem 8: For a given N_U and fixed λ, $D_1 \geq_{st}^2 D_2$ iff $\bar{\rho}(D_1) \leq \bar{\rho}(D_2)$ for all $w(x)$.

The preceding results set up motivation to ask the question of whether there exists a distribution that dominates all others in second order. We answer this next.

Theorem 9: For a given mean, a constant stochastically dominates all other random variables in second order.

This leads to the main result of this section.

Corollary 1: Constant inter-synchronization delays are optimal under both $\bar{\eta}$ and $\bar{\rho}$, all suitable weights $w(x)$, and all update processes N_U.

This allow us to resolve the relationship between the variance of D and penalty. If $E[D_1] = E[D_2]$, then $D_1 \geq_{st}^2 D_2$ implies $Var[D_1] \leq Var[D_2]$, but the opposite is not true. This shows that for a given download rate, just reducing the variance of refresh intervals, without enforcing $D_1 \geq_{st}^2 D_2$, is *insufficient* to improve the penalty across *all* functions $w(x)$.

As an example, recall the special case of $\bar{\rho}$ with $w(x) = 1$ in (22), where the penalty was reduced iff the variance of D was; however, no such causality exists for $w(x) = x$ or $\log(1+x)$. On the other hand, if reduction in penalty holds for all measures $w(x)$, then stochastic ordering between D_1 and D_2 follows and thus variance has to decrease (i.e., ordering of variances is *necessary*, but not sufficient).

C. Phase-Lock

Even though constant D is optimal from the staleness perspective, it unfortunately fails to guarantee age-independence (13) against *all* underlying N_U. In general, ensuring ASTA requires conditions (e.g., LAA, WLAA, LBA [23]) that are extremely difficult to verify, which currently precludes obtaining properties of N_D that are both sufficient and necessary for (13) to hold. Instead, we provide the following sufficient condition.

Theorem 10: In scenarios with iid $\{D_k\}$ and non-lattice $F_D(x)$, age-independence holds for *all* $N_U(x)$.

For example, randomly alternating D between intervals of size a and b, such that a is rational and b is irrational, will produce an age-independent download process. The closer the two points, the lower the resulting staleness cost. Using two irrational numbers that are not integer multiples of each other or any continuous $F_D(x)$ is possible as well.

Conversely, in certain cases, it may be known a-priori that N_U is non-lattice renewal. Then, it is not difficult to see that constant D guarantees age-independence, which means that optimality is achievable in practice under these conditions.

VI. Applications

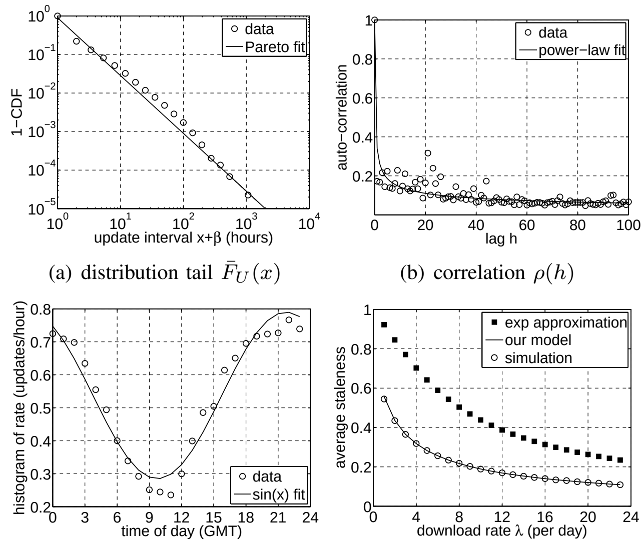
We now examine the presence of Poisson updates in real data sources and show how to apply the developed models to solve several classes of multi-source/replica problems.

A. Real-Life Update Processes

Closer examination of the origin [8] of the Poisson conclusion reveals several limitations. First, the distribution of page inter-update intervals was sampled using *incomplete observation*, meaning that some of the updates went unnoticed. As a result, bias could have been introduced in the measurements. Second, the exponential distribution was fitted to updates of *multiple* pages rather than a single page. Poisson dynamics have been known to emerge when aggregating arrival processes [1] and summing up variables [2], which does not tell us much about the individual distributions being combined. Finally, to conclude that N_U is Poisson, it is insufficient to observe an exponential distribution in $\{U_i\}_{i=1}^{\infty}$; instead, one must also show stationary independent increments [33].

B. Wikipedia

Many of these pitfalls can be avoided if model verification is performed over sources that expose information about *each* update. One particularly interesting source with public traces of all modification timestamps is Wikipedia [29]. From a search-engine perspective, this website represents a realistic example of data churn stemming from user interaction with



(a) distribution tail $\bar{F}_U(x)$ (b) correlation $\rho(h)$

(c) update rate (d) model accuracy

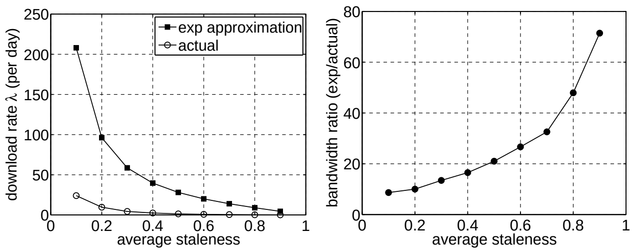
Fig. 7. George W. Bush page dynamics.

each other (e.g., edits from other people), flash crowds in response to external events, and diurnal activity patterns of the human lifecycle. Wikipedia is also well-suited for purposes of model validation and discussion.

To shed light on the complexity of real $F_U(x)$, we plot in Fig. 7(a) the tail CDF of inter-update delay for the most frequently modified article – "George W. Bush" with $44,296$ updates in 10 years (mean delay $E[U] = 1.86$ hours). The figure is a close match to Pareto tail $(1+x/\beta)^{-\alpha}$ with $\alpha = 1.4$ and $\beta = 0.93$. In Fig. 7(b), we show the corresponding auto-correlation function $\rho(h)$ with a power-law fit $h^{-0.37}$, which suggests long-range dependence (LRD) with Hurst parameter 0.81. Of course, LRD effects might be caused and/or compounded by non-stationarity. To address this question, Fig. 7(c) shows the update rate throughout the day, clearly indicating non-stationary dynamics.

This example underscores the need to keep the model general and not limit results to renewal or even stationary cases, which was our goal with assumptions (5)-(6). Approximating $F_U(x)$ as non-lattice and using constant D, we next compute the probability of staleness for this page by supplying (17) with George W. Bush' empirically computed distribution $F_U(x)$. We contrast the result against the closest Poisson formula $1 - \lambda(1 - e^{-\mu/\lambda})/\mu$ from [8]. Fig. 7(d) shows that (17) is accurate, but the Poisson approximation suffers over 100% relative error for much of the examined range.

What is more important is the performance of the model in providing an accurate assessment of the download bandwidth needed to achieve a given p. We invert the formulas to solve for λ as a function of p and plot the result in Fig. 8(a). These results show a much more dramatic difference. For example, 20% staleness requires 95 downloads/day according to previous Poisson models, while in reality this can be achieved with just 8. To illustrate this better, we show the



(a) crawl rate (b) ratio

Fig. 8. Application of staleness models to the update process of George W. Bush.

ratio of these two curves in Fig. 8(b), where the amount of Poisson overestimation varies from one to almost two orders of magnitude depending on the desired p.

C. Aggregation (Many-to-One)

When a single replica tracks M sources, performance is assessed by its ability to provide usable aggregate information to the consumer. If sources are independent, many results are relatively easy to obtain. For example, consider a system that selects a particular replica and loads it with a MapReduce job that has to execute over the data of all sources. A computation may be considered successful if at least one source is fresh at the time of job request. Then, the fraction of successful attempts is $1-\prod_{i=1}^{M} p_i$, where p_i in (14) is the probability of staleness for source i.

A more interesting problem is optimal allocation of download rates to different sources. Suppose q_i is the probability that an incoming query requests data from source i and μ_i is its update rate. Then, the goal is to allocate refresh rates λ_i so as to optimize the expected staleness cost $C(\lambda_i, \mu_i)$ for a given bandwidth budget Λ:

$$\min \sum_{i=1}^{M} q_i C(\lambda_i, \mu_i) \qquad \text{subject to} \qquad \sum_{i=1}^{M} \lambda_i \leq \Lambda, \tag{27}$$

where $C(\lambda_i, \mu_i)$ refers to either $\bar{\eta}$ or $\bar{\rho}$.

For $\Lambda \ll \sum_{i=1}^{M} \mu_i$ and certain choices of $w(x)$, solutions to (27) using cost $\bar{\eta}$ are known to completely starve frequently modified sources in favor of those that are updating slowly [8]. Since (27) does not have a closed-form solution under $\bar{\eta}$ even in the simplest cases, specific conditions for starvation are not clear. Complete loss of synchronization for sources whose μ_i is above some (typically unknown) threshold may be an unwelcome surprise for many applications. This naturally leads to the question of whether $\bar{\rho}$ suffers from the same drawback. We address this next.

Theorem 11: Assume $q_i\mu_i > q_j\mu_j > 0$ and let refresh delays be optimal (i.e., constant). Then, the solution to (27) using $\bar{\rho}$ guarantees that $\lambda_i > \lambda_j > 0$.

To explain how optimization with $\bar{\rho}$ can be used, we assume constant D and $w(x) = 1$, with the goal to maximize $\sum_{i=1}^{M} q_i E[M_i(t)]$. Solving (27), the optimal download rate of each page is proportional to the square root of $q_i\mu_i$:

$$\lambda_i = \Lambda \frac{\sqrt{q_i\mu_i}}{\sum_{j=1}^{M} \sqrt{q_j\mu_j}}. \tag{28}$$

The optimal penalty is then:

$$\sum_{i=1}^{M} q_i E[M_i(t)] = \sum_{i=1}^{M} \frac{q_i\mu_i}{2\lambda_i} = \frac{(\sum_{j=1}^{M} \sqrt{q_j\mu_j})^2}{2\Lambda}. \tag{29}$$

Define random variable μ to have the same distribution as $\{\mu_1, \ldots, \mu_M\}$. Then, for the most basic scenario where all pages are equally popular, i.e., $q_i = 1/M$, we get:

$$\sum_{i=1}^{M} q_i E[M_i(t)] = M \frac{E[\sqrt{\mu}]^2}{2\Lambda}. \tag{30}$$

For the other extreme, where pages are searched for in proportion to their modification rate, i.e., $q_i \sim \mu_i$, we have:

$$\sum_{i=1}^{M} q_i E[M_i(t)] = M \frac{E[\mu]}{2\Lambda}. \tag{31}$$

To put these models in perspective, we use Wikipedia's distribution of μ, which happens to be quite heavy-tailed (i.e., Zipf shape $\alpha = 0.6$). The average update rate across all pages is $E[\mu] = 8$ updates/day; however, 98% of them exhibit μ_i less than 1/day, 90% less than 1/week, and 50% below 8/year. Using this distribution in (30) and (31) shows that optimizing staleness of the entire Wikipedia under uniform page access $q_i = 1/M$ requires 46 times less bandwidth Λ than under Zipf. This can be explained by the fact that keeping frequently modified pages fresh costs more bandwidth. This effect is related to the variance of $\sqrt{\mu}$:

$$\frac{E[\mu]}{E[\sqrt{\mu}]^2} = \frac{Var[\sqrt{\mu}] + E[\sqrt{\mu}]^2}{E[\sqrt{\mu}]^2} = \frac{Var[\sqrt{\mu}]}{E[\sqrt{\mu}]^2} + 1. \tag{32}$$

Consider extrapolating these results to $M = 100$B sources and keeping the expected consumer lag $\sum_{i=1}^{M} q_i E[M_i(t)]$ below ω updates. We use the two models above as lower/upper bounds on the actual search-engine crawl rate. The first case requires download capability $\Lambda_1 = M \cdot E[\sqrt{\mu}]^2/2\omega = 99/\omega$ thousand pages per second (pps), while the second one $\Lambda_2 = M \cdot E[\mu]/2\omega = 4.6/\omega$ million pps. For $\omega = 10$ and 25 KB per page, these translate into 2 and 92 Gbps, respectively. Results can be easily adjusted to non-Wikipedia situations as long as $E[\sqrt{\mu}]$ and $E[\mu]$ are known.

D. Load-Balancing (One-to-Many)

The issue of redundant replication from a single source to m nodes is quite different from the opposite case considered in the previous subsection. When the source fails, suppose the goal is to deduce the expected penalty afforded by the freshest member of the entire collection of m replicas. The issue at stake is how this $1 \times m$ scenario compares to a single replica with some refresh rate λ and optimal D. To keep comparison fair, assume that each of the m replicas is allowed budget λ/m in synchronization with the source. Decentralized operation leads to much better robustness under failure, but is it possible that this causes reduced freshness?

The main caveat in solving this problem is that staleness at different replicas is no longer independent. This happens because updates at the source simultaneously make all copies outdated, which means that reliability does not benefit exponentially with increased m. To overcome this issue, let $N_D^1, \ldots, N_D^m$ be the download processes used by the individual replicas. Then, observe that the entire collection can be replaced by a single replica that implements a refresh pattern N_D^*, which is a *superposition* of all point processes $\{N_D^i\}_{i=1}^m$. Therefore, the source can be recovered during the crash with a probability determined solely by N_D^*.

If we assume centralized scheduling between the replicas, then it is possible to run the system optimally (i.e., using a perfectly spaced out round-robin) and thus keep the overall penalty exactly the same as with a single replica. Under fully decentralized (i.e., independent) replica operation and $m \to \infty$, each rate $\lambda/m \to 0$ and thus N_D^* likely converges in distribution to a Poisson process with rate λ (Palm-Khintchine theorem [15]). This creates a problem, however, because exponential D requires noticeably more overhead than constant D to achieve the same staleness penalty. For example, using our model for $\bar{\rho}$ and discussion after (22), this difference is by a factor of 2 for $w(x) = 1$ and by a factor of 6 for $w(x) = x$, which shows that a distributed cluster of replicas may need to consume $100 - 500\%$ more bandwidth than a centralized solution for a given level of QoS (quality-of-service).

VII. Conclusion

The paper introduced a novel model of sampled age under general non-Poisson update/synchronization processes and applied it to obtain many useful metrics of staleness. We additionally established that constant inter-refresh intervals were optimal for all considered cases and provided guidelines for achieving ASTA even in those cases. We finally considered a family of related problems stemming from $1 \times m$ and $M \times 1$ replication, showing that they can be easily solved from the preceding analysis of the 1×1 case.

Future work involves reducing staleness when N_D is allowed to depend on observations of N_U and/or prior knowledge of its distribution of update cycles $F_U(x)$.

References

[1] S. L. Albin, "On Poisson Approximations for Superposition Arrival Processes in Queues," *Management Sci.*, vol. 28, no. 2, pp. 126–137, 1982.

[2] R. Arratia, L. Goldstein, and L. Gordon, "Two Moments Suffice for Poisson Approximations: The Chen-Stein Method," *The Annals of Probability*, vol. 17, no. 1, pp. 9–25, Jan. 1989.

[3] F. Baccelli, S. Machiraju, D. Veitch, and J. Bolot, "The Role of PASTA in Network Measurement," in *Proc. ACM SIGCOMM*, Sep. 2006.

[4] B. E. Brewington and G. Cybenko, "How Dynamic is the Web," *Computer Networks*, no. 1-6, pp. 257–276, Jun. 2000.

[5] L. Bright, A. Gal, and L. Raschid, "Adaptive Pull-based Policies for Wide Area Data Delivery," *ACM Trans. Database Syst.*, no. 2, pp. 631–671, June 2006.

[6] D. Carney, S. Lee, and S. Zdonik, "Scalable Application-Aware Data Freshening," in *Proc. IEEE ICDE*, March 2003, pp. 481–492.

[7] J. Cho and H. Garcia-Molina, "The Evolution of the Web and Implications for an Incremental Crawler," in *Proc. VLDB*, Sep. 2000, pp. 200–209.

[8] J. Cho and H. Garcia-molina, "Synchronizing a Database to Improve Freshness," in *Proc. ACM SIGMOD*, May 2000, pp. 117–128.

[9] J. Cho and H. Garcia-Molina, "Estimating frequency of change," *ACM Trans. Internet Technol.*, vol. 3, pp. 256–290, August 2003.

[10] E. G. Coffman, Z. Liu, and R. R. Weber, "Optimal Robot Scheduling for Web Search Engines," *Journal of Scheduling*, no. 1, pp. 15–29, Jun. 1998.

[11] D. Denev, A. Mazeika, M. Spaniol, and G. Weikum, "SHARC: Framework for Quality-Conscious Web Archiving," in *Proc. VLDB*, Aug. 2009.

[12] D. Dey, Z. Zhang, and P. De, "Optimal Synchronization Policies for Data Warehouses," *INFORMS J. on Computing*, no. 2, pp. 229–242, Jan. 2006.

[13] J. Eckstein, A. Gal, and S. Reiner, "Monitoring an Information Source Under a Politeness Constraint," *INFORMS J. on Computing*, no. 1, pp. 3–20, Jan. 2008.

[14] A. Gal and J. Eckstein, "Managing Periodically Updated Data in Relational Databases: a Stochastic Modeling Approach," *J. ACM*, no. 6, pp. 1141–1183, Nov. 2001.

[15] D. Heyman and M. Sobel, *Stochastic Models in Operations Research, Volume 1*. McGraw-Hill, 1982.

[16] Y. Huang, R. H. Sloan, and O. Wolfson, "Divergence Caching in Client-Server Architectures," in *Proc. IEEE PDIS*, Sep. 1994, pp. 131–139.

[17] R. Ladin, B. Liskov, L. Shrira, and S. Ghemawat, "Providing High Availability Using Lazy Replication," *ACM Trans. Comput. Syst.*, no. 4, pp. 360–391, Nov. 1992.

[18] J.-J. Lee, K.-Y. Whang, B. S. Lee, and J.-W. Chang, "An Update-Risk Based Approach to TTL Estimation in Web Caching," in *Proc. WISE*, Dec. 2002, pp. 21–29.

[19] X. Li, D. B. Cline, and D. Loguinov, "On Sample-Path Staleness in Lazy Data Replication," Texas A&M University, Tech. Rep. 2015-1-1, Jan. 2015. [Online]. Available: http://irl.cs.tamu.edu/publications/.

[20] Y. Ling and W. Chen, "Measuring Cache Freshness by Additive Age," *SIGOPS Oper. Syst. Rev.*, vol. 38, pp. 12–17, Jul. 2004.

[21] Y. Ling and J. Mi, "An Optimal Trade-off between Content Freshness and Refresh Cost," *Applied Probability*, vol. 41, no. 3, pp. 721–734, Sep. 2004.

[22] N. Matloff, "Estimation of Internet File-access/Modification Rates from Indirect Data," *ACM Trans. Model. Comput. Simul.*, vol. 15, pp. 233–253, Jul. 2005.

[23] B. Melamed and W. Whitt, "On Arrivals That See Time Averages," *Operations Research*, vol. 38, no. 1, pp. 156–172, 1990.

[24] C. Olston and J. Widom, "Best-Effort Cache Synchronization With Source Cooperation," in *Proc. ACM SIGMOD*, May 2002, pp. 73–84.

[25] C. Olston and S. Pandey, "Recrawl Scheduling Based on Information Longevity," in *Proc. WWW*, Apr. 2008, pp. 437–446.

[26] K. C. Sia and J. Cho, "Efficient Monitoring Algorithm for Fast News Alerts," *IEEE Trans. Knowledge and Data Engineering*, no. 7, pp. 950–961, Jul. 2007.

[27] Q. Tan and P. Mitra, "Clustering-based Incremental Web Crawling," *ACM Transactions on Information Systems*, no. 4, pp. 1–27, Nov. 2010.

[28] B. Urgaonkar, A. G. Ninan, M. Salimullah, R. Shenoy, and K. Ramamritham, "Maintaining Mutual Consistency for Cached Web Objects," in *Proc. IEEE ICDCS*, Apr. 2001, pp. 371–380.

[29] Wikipedia Dumps. [Online]. Available: http://dumps.wikimedia.org/enwiki/20110317/.

[30] G. Wise and E. Hall, *Counterexamples in Probability and Real Analysis*. Oxford Univ. Press, 1993.

[31] J. L. Wolf, M. S. Squillante, P. S. Yu, J. Sethuraman, and L. Ozsen, "Optimal Crawling Strategies for Web Search Engines," in *Proc. WWW*, May 2002, pp. 136–147.

[32] R. W. Wolff, "Poisson Arrivals See Time Averages," *Operations Research*, vol. 30, no. 2, pp. 223–231, 1982.

[33] R. W. Wolff, *Stochastic Modeling and the Theory of Queues*. Prentice Hall, 1989.

[34] M. Yang, H. Wang, L. Lim, and M. Wang, "Optimizing Content Freshness of Relations Extracted From the Web Using Keyword Search," in *Proc. ACM SIGMOD*, June 2010, pp. 819–830.

[35] H. Yu and A. Vahdat, "Design and Evaluation of a Continuous Consistency Model for Replicated Services," in *Proc. USENIX OSDI*, June 2000, pp. 305–318.