

Optimal Plans for Aggregation

Andrei Broder
AltaVista Corporation
andrei.broder@av.com

Michael Mitzenmacher*
Harvard University
michaelm@eecs.harvard.edu

Abstract

We consider the following problem, which arises in the context of distributed Web computations. An *aggregator* aims to combine specific data from n *sources*. The aggregator contacts all sources at once. The time for each source to return its data to the aggregator is independent and identically distributed according to a known distribution. The aggregator at some point stops waiting for data and returns an answer depending only on the data received so far. If the aggregator returns the aggregated information from k of the n sources at time t it obtains a reward $R_k(t)$ that grows with k and decreases with t . The goal of the aggregator is to maximize its expected reward.

We prove that for certain broad families of distributions and broad classes of reward functions, the optimal plan for the aggregator has a simple form and hence can be easily computed.

*Supported in part by an Alfred P. Sloan Research Fellowship and NSF grants CCR-9983832, CCR-0118701, and CCR-0121154.

1 Introduction

We consider the following problem: an *aggregator* aims to combine specific data from n *sources*. The aggregator contacts all sources simultaneously. The time for each source to return its data to the aggregator is independent and identically distributed according to a known distribution. The aggregator at some point stops waiting for data and returns an answer depending only on the data received so far. If the aggregator returns the aggregated information from k of the n sources at time t it obtains a reward $R_k(t)$ that grows with k and decreases with t .

At first blush this problem might seem artificial; however it is quite common in the Web context where results are composed from many sources and there is often a conflict between returning fast results versus returning good results.

Here are some examples:

- *Search engines*. Current general search engines (AltaVista, Google, Inktomi, etc.) index a corpus of hundreds of millions of Web pages by maintaining a large collection of inverted files distributed in various ways among multiple machines. Often, when a query is received, a central controller must aggregate data from several machines. The goal of the controller is to optimize the user experience, which depends on two factors. The first is the quantity of information received; we will start with the simplifying assumption that this depends only on the number of machines who have responded to the controller's query. (Practically, ignoring some machines is equivalent to a reduction in the size of the corpus.) The second is the amount of time the user has to wait for the response. We wish to design an algorithm for determining an optimal plan for the controller. Of course in reality the situation is more complex: different machines may be experiencing different loads, or some machines may be down altogether, without the controller's knowledge.
- *Metasearch engines and peer-to-peer (P2P) search*. The situation here is similar except that the controller now waits for results from multiple search engines. In the P2P scenario the number of sources may not even be known a priori.
- *Portal pages*. Personalized portal pages (Excite, Lycos, Yahoo) integrate all sort of personalized information, each provided by a different server: stock quotes, news, weather, horoscopes, movie schedules, etc. There is a conflict here between returning the page quickly and having the most updated information from each of these servers. The aggregator may decide to return out-of-date information and/or not fill certain fields in the interest of responding in a timely fashion.
- *Page construction from cached components*. Companies such as Akamai provide network caching services whereby certain objects such as images are cached on multiple servers located all over the world. When a user requests a page including an "Akamaized" object, the object is returned from the most appropriate cache. In more advanced implementations most of the page is stored on an Akamai server in the form of a collection of objects and only minimal requests are made to a central server. The issue here is that

the objects have an expiration date. In some situations the cache server must decide whether to return a page containing some expired objects, wait for updates, or time out the user request altogether.

Again, the general problem situation uniting the above scenarios is that an *aggregator* simultaneously makes requests to n *sources* and at some time t decides to return a response in order to obtain a *reward* $R_k(t)$ that depends on the number k of sources that have responded by time t . The goal of the aggregator is to maximize the expected reward obtained. In full generality, this problem is quite complex; the form of the solution depends on the reward functions, the rules governing machine responses, and the type of solution desired (approximation or exact). For this paper, we have focused on several specific goals that guide us. First, in order to provide a mathematical framework for the problem, we initially focus on a natural probabilistic model, where the return time for each of the n sources is independent and identically distributed (i.i.d.) according to a cumulative distribution function F known to the controller.

Second, we are interested in situations where we can develop plans that are *optimal* for the aggregator. The problem of designing fast, on-line general approximation algorithms for this scenario is clearly interesting and a worthy problem for future work. Our hope, however, is to find a broad class of instances where computing an optimal solution is possible.

Third, we are interested in situations where we can develop a plan for the aggregator that is simple. While simplicity is a relative notion, we may describe it formally here as follows. The plan can be thought of as a series of binary functions $P_k(t)$, where $P_k(t)$ is 1 if the aggregator should return with k responses at time t and 0 otherwise. One natural measure of the complexity of $P_k(t)$ is the number of *transition points*, where t_0 is a transition point if for every $\epsilon > 0$ there exist points x, y in the neighborhood $[t_0 - \epsilon, t_0 + \epsilon]$ such that $P_k(x)$ and $P_k(y)$ take on different values. Intuitively, the optimal plan changes around t_0 . A priori there is no reason that the number of transition points must be finite (indeed, it is not clear that the set of transition points needs to be countable). A simple plan will have few transition points. In this work we find situations where each $P_k(t)$ has a single transition point.

Our work relies on using statistical properties from the literature of systems reliability: *increasing failure rate* and *decreasing failure rate*. We provide a weak example of our results here, although the statement may be more meaningful only after all the definitions and framework have been established. Suppose each reward function $R_k(t)$ equals $r_k(1 - Z(t))$, where the constants $r_k > 0$ represent the undiscounted rewards and $1 - Z(t)$ is a discount factor related to the time waited. Also, let the function $Z(t)$ be a cumulative distribution function for some probability distribution. Note that $Z(t)$ is not actually used as the distribution of a random variable; this is just a convenient way to specify the *shape* of the function. Indeed, one of our important contributions is to recognize that interesting results arise in this setting by considering the shape of the reward function as a probability distribution. As an example, a natural special case that is often used in Markov decision processes [4, 13] is where the reward has *exponential decay*; that is, $R_k(t) = r_k e^{-\gamma t}$ for some fixed γ . Recall that F is the cumulative distribution function of the return time of each source. Let F_0 represent the cumulative distribution for the first time of any of the n sources. One result we obtain is the

following:

Theorem: If F has increasing failure rate, Z has decreasing failure rate, and the sum of the failure rates for F_0 and Z is decreasing, then in the optimal plan for each k there is a single transition point. Specifically, on receiving the k th response the aggregator should either return immediately or wait for the next response. Similarly, if F has decreasing failure rate, Z has increasing failure rate, and the sum of the failure rates for F_0 and Z is increasing, then in the optimal plan for each k there is a single transition point. With k responses, the receiver should wait up to some time t_k for another response before giving up.

Hence, for extremely large classes of return distributions and reward functions, the optimal plan is incredibly simple. Using this fact, one can generally calculate the optimal plan numerically if given the return distribution in an appropriate form. Another possible application of these results is to allow an aggregator to learn a near-optimal plan effectively even if the underlying distributions are unknown. The aggregator uses a learning algorithm to discover the best plan; it can narrow the space of plans to consider by focusing only on plans with a single transition point if the aggregator knows or suspects the conditions of the theorem are satisfied.

We also provide several additional theorems. For example, we provide a result that has fewer restrictions on the distributions, but instead restricts the behavior of the values r_k . We also examine the cases where different sources may have different values or different response times. In these settings, the optimal plan also has a simple form, with one transition point per subset of sources. Finally, we present an argument showing that making another natural weaker assumption on the distributions of the response times is insufficient for our results.

1.1 Previous work

The concepts of increasing failure rate and decreasing failure rate are utilized primarily in the literature of systems reliability, where they are used to describe the lifetime of system components [12, 15, 16, 19]. The lifetime distribution may affect the proper strategy for scheduling maintenance. In this sense, the idea that these distributions can yield algorithmic implications has been known for some time [16].

The problem we examine here fits naturally into the scheme of Markov decision processes [4, 13]; there is an underlying Markov reward process, and one wants to maximize the reward. Indeed, it specifically fits into the framework of optimal stopping theory, where one wishes to find the optimal time to stop a process in order to maximize a reward [6, 8]. For example, the well-known secretary problem, where an employer interviews n secretaries and must decide after each interview whether or not to hire that person on the spot, exemplifies the problem framework of optimal stopping theory. The reward function for the secretary problem depends on the variation; the employer may wish to maximize the probability of finding the best secretary, or optimize for some other criterion (see, e.g., [1]). We note that in order to keep this work self-contained, we eschew using notation or language specific to optimal stopping theory.

What is novel in our work is the connection between reliability theory and this natural

stopping problem, and in particular the interesting relationships between the return time distribution and the shape of the reward discounting function. There is very little other prior work that builds on this natural connection. Most recently, properties of increasing failure rate distributions were used by Boyan and Mitzenmacher in the context of another scheduling and planning problem, crossing town by bus [5]. This paper extended previous work by Datar and Ranade [7], which considered the question under the assumption that the arrival distribution for buses to a stop were exponential distributions. The main result of Boyan and Mitzenmacher is to show that the weaker assumption that bus arrival distributions have increasing failure rate results in simple plans for the optimal solution, which can then be calculated using numerical methods.

After completing this work, we found additional prior work from the statistical community that utilized the connection between stopping times for aggregation problems and reliability properties of distributions [17, 18, 9].¹ In all of this prior work, however, the reward function used in the analysis was $R_k(t) = r_k - ct$ for some constant c or in some cases $R_k(t) = r_k - c(t)$ for a convex $c(t)$. That is, their discount was additive over time, while ours is multiplicative.² While the flavor of the work is similar, all of our results are new. Moreover, some of our generalizations appear to have no parallel in this previous work. While it is arguable which model is more accurate, we believe that each may be appropriate in different situations.

Our work also has some of the flavor of problems for on-line algorithms in the face of uncertainty, including machine breakdown [2, 3, 10, 11]. Our aggregator must decide when to stop in the face of uncertain response time. Again, our goals here are different than in this previous work, as we focus on what probabilistic assumptions are necessary for the optimal plan to have a simple, calculable form.

2 Definitions and Basic Lemmas

We begin by providing standard probabilistic definitions and some simple lemmas. Consider a cumulative probability distribution $F(t)$ with density function $f(t) = F'(t)$. For convenience we will generally assume that F and f are non-zero, continuous, and differentiable over all positive real numbers, although our results hold more generally.

We consider the case where F has *increasing failure rate*.³ At an intuitive level, an event has increasing failure rate if the longer you’ve waited for it, the more likely it’s just about to happen. More formally, we have the following.

Definition 1 *For a nonnegative random variable X with cumulative distribution function (cdf) $F(t)$, we define the corresponding survival function to be $\bar{F}(t) = 1 - F(t)$. The failure*

¹We suggest that this prior work was fairly obscure, and it only arose after a lengthy Web search.

²It might seem one could move from multiplicative to additive discounts by placing an appropriate logarithm, but because we seek to maximize the expected reward, this does not appear to be the case.

³Here we follow the perhaps unfortunate but apparently standard practice and use “increasing” to mean “non-decreasing” and “decreasing” to mean “non-increasing” throughout. So a constant function is both increasing and decreasing, and having increasing failure rate really means the failure rate is non-decreasing, even though IFR is the standard term.

rate of X is defined as $r(t) = f(t)/\bar{F}(t)$.

Definition 2 A nonnegative random variable X with function (cdf) $F(t)$ is said to have increasing failure rate (or be IFR) if the failure rate $r(t)$ is increasing. Equivalently, X is IFR if $\log \bar{F}(t)$ is concave on the support of \bar{F} . That is, $\bar{F}(t)$ is logconcave.

We may say that F is IFR instead of X is IFR where the meaning is clear. Also note the function $r(t)$ satisfies

$$r(t) = \frac{f(t)}{\bar{F}(t)} = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t < X \leq t + \Delta t \mid X > t)}{\Delta t};$$

that is, $r(t)$ represent the probability that the event corresponding to X is about to occur, given that it has not occurred already. This explains our previously given intuition. Exponential, normal, and uniform distributions are for example all IFR, as are gamma distributions corresponding to the sum of exponential distributions.

Similarly, we have the following:

Definition 3 A nonnegative random variable X with function (cdf) $F(t)$ is said to have decreasing failure rate (or be DFR) if the failure rate $r(t)$ is decreasing. Equivalently, X is DFR if $\log \bar{F}(t)$ is convex on the support of \bar{F} , or $\bar{F}(t)$ is logconvex.

Exponential distributions are also DFR; since the failure rate for exponential distributions is constant (another way of saying the distribution is memoryless), they are both IFR and DFR. Another class of interesting distributions that fall into the DFR class is certain power-law distributions, for example distributions where $\bar{F}(t) = \frac{1}{(1+x)^\alpha}$ for $\alpha > 0$. Power-law distributions have appeared in several contexts in recent work, as they model phenomena with heavy tails. A heavy-tailed response time distribution may make sense in situations where there is some probability a machine is temporarily down, in which case there is a non-trivial possibility that it may be a long time before a response is heard.

Before proceeding, we state a few short lemmas about IFR and DFR distributions that prove useful in the sequel.

Lemma 1 If X_1, X_2, \dots, X_n are IFR (resp. DFR) random variables, then $\min(X_1, X_2, \dots, X_n)$ is IFR (resp. DFR).

Proof: If $r_j(t)$ is the failure rate for X_j , then $\sum_{j=1}^n r_j(t)$ is the failure rate for $\min(X_1, X_2, \dots, X_n)$. The result follows. \square

Lemma 2 If F is IFR (resp. DFR), then $\bar{F}(t+x)/\bar{F}(t)$ is decreasing (resp. increasing) in t for a fixed $x \geq 0$.

Proof: We prove for the case where F is IFR; the other case is entirely similar. The derivative of $\bar{F}(t+x)/\bar{F}(t)$ with respect to t is

$$\frac{\bar{F}(t+x)f(t) - \bar{F}(t)f(t+x)}{(\bar{F}(t))^2}.$$

If F is IFR, then

$$\frac{f(t+x)}{\bar{F}(t+x)} \geq \frac{f(t)}{\bar{F}(t)},$$

from which we may conclude the derivative is non-positive and hence the theorem holds. \square

Finally, we recall that we are interested in the complexity of the plan, where the complexity is measured in terms of *transition points*.

Definition 4 *A plan has a transition point at time t with k responding sources if for every $\epsilon > 0$ there exist times x, y in the neighborhood $[t - \epsilon, t + \epsilon]$ such that with k responding sources the plan returns at time x but waits at time y .*

3 Simple Plans for Classes of Distributions and Rewards

In this section, we prove our result in the case where the cumulative distribution function F for the response time from each individual source is IFR. For convenience, from here on we denote by F_j the cumulative distribution function for the next response given that j sources have already responded. Of course f and f_j are the corresponding density functions. For notational convenience we may think of \bar{F}_n as being 1 everywhere; that is, the corresponding random value is infinite.

Lemma 3 *If F is IFR (respectively DFR), then F_j is also IFR (respectively DFR).*

Proof: This follows immediately from Lemma 1, since $f_j(t)/\bar{F}_j(t) = (n-j)f(t)/\bar{F}(t)$. \square

Let $R_j(t) = r_j \bar{Z}(t)$, where $r_0 \leq r_1 \leq r_2 \dots \leq r_n$ are constants and $\bar{Z}(t)$ is the survival function of a probability distribution. Recall that the r_j can be considered the undiscounted reward for returning when j sources have been heard from, and $\bar{Z}(t)$ is a multiplicative discount accounting for time. We emphasize again that because $\bar{Z}(t)$ is a survival function does not imply the reward $R_j(t)$ is random; we are merely specifying the shape of the reward function in a convenient manner. For convenience we assume the support of all distributions is all positive real numbers. Let $z(t)$ be the corresponding density function for the cumulative distribution function $Z(t)$.

Finally, note that if $\bar{F}(t)$ and $\bar{Z}(t)$ are survival functions, then so is their product $\bar{F}(t)\bar{Z}(t)$. Abusing notation, we let $F \cdot Z$ be the cumulative distribution function associated with the survival function $\bar{F}(t)\bar{Z}(t)$. The failure rate associated with the distribution $F_j \cdot Z$ is the sum $\frac{f_j(t)}{F_j(t)} + \frac{z(t)}{Z(t)}$.

Theorem 1 *If F has increasing failure rate, Z has decreasing failure rate, and $F_m \cdot Z$ has decreasing failure rate, then in the optimal plan for each number of responses $j \geq m - 1$ there is a single transition point. Moreover, on receiving the j th response the aggregator should either return immediately or wait for the next response.*

We have stated Theorem 1 in a general way, so that it may apply once some number of responses have been obtained even if it does not apply for fewer responses. In particular, since $\bar{F}_n(t)$ is identically one, the theorem always holds for $n - 1$ responses if F is IFR and Z is DFR.

Proof:

We begin by noting a useful fact. Since $\frac{f_j(t)}{F_j(t)} = (n - j) \frac{f(t)}{F(t)}$, F is IFR, Z is DFR, and the failure rate associated with the distribution $F_j \cdot Z$ is the sum $\frac{f_j(t)}{F_j(t)} + \frac{z(t)}{Z(t)}$, we have that if $F_j \cdot Z$ is DFR, so is $F_k \cdot Z$ for all $k \geq j$. This fact allows us to do a backwards induction.

Let $V_j(t)$ be the expected payoff to the aggregator using the optimal strategy from time t given that j sources have responded. Clearly $V_n(t) = R_n(t)$; the optimal strategy when all sources have been heard must be to return immediately. We show via a backward induction from $j = n$ down to m the following:

- At any time t , the optimal strategy having j responses is either to wait for the next arrival, or to return immediately.
- The function $W_j(t) = V_j(t)/\bar{Z}(t)$ is increasing in t .

The above are both true for $j = n$. The first part of the induction will also go through for $j = m - 1$, and hence the induction yields the theorem.

To begin, for any fixed t consider $v_t(u)$ to be the value obtained by the aggregator if, starting at time t , it chooses the policy of waiting until time $u \geq t$ and then returning when it has already heard from j sources. The optimal policy must be of this form (since the distribution is known ahead of time, randomness cannot help the aggregator). Then

$$v_t(u) = \int_{x=t}^u V_{j+1}(x) \frac{f_j(x)}{F_j(t)} dx + \frac{\bar{F}_j(u)}{F_j(t)} R_j(u).$$

The first term on the right hand side corresponds to the contribution should an arrival happen before u , and the second term corresponds to the contribution if no arrival happens before u .

Now

$$\begin{aligned} \frac{dv_t}{du} &= V_{j+1}(u) \frac{f_j(u)}{F_j(t)} - \frac{f_j(u)}{F_j(t)} R_j(u) - \frac{\bar{F}_j(u)}{F_j(t)} r_j z(u) \\ &= \frac{f_j(u) \bar{Z}(u) [W_{j+1}(u) - r_j] - r_j \bar{F}_j(u) z(u)}{\bar{F}_j(t)} \end{aligned}$$

We show that the numerator of the final right hand side above is non-positive for all u in some interval $t \leq u \leq u^*$, and non-negative for all $u \geq u^*$ (where u^* may take on the value

∞). It follows that the maximum value for v is achieved either when $u = t$ or in the limit as u goes to infinity. That is, the optimal strategy is either to return immediately or wait for the next arrival.

Suppose u^* is the infimum of all values such that

$$f_j(u^*)\bar{Z}(u^*)[W_{j+1}(u^*) - r_j] - r_j\bar{F}_j(u^*)z(u^*) \geq 0.$$

Equivalently, u^* is the infimum of all values such that

$$\frac{f_j(u^*)}{\bar{F}_j(u^*)}[W_{j+1}(u^*) - r_j] \geq r_j \frac{z(u^*)}{\bar{Z}(u^*)}.$$

The left hand side above is increasing in u^* , as F_j is IFR, $W_{j+1}(u^*) \geq r_{j+1} \geq r_j$, and W_{j+1} is increasing by the inductive hypothesis. The right hand side is decreasing, as Z is DFR. Hence for all $u \geq u^*$,

$$\frac{f_j(u)}{\bar{F}_j(u)}[W_{j+1}(u) - r_j] \geq r_j \frac{z(u)}{\bar{Z}(u)},$$

or equivalently, for $u \geq u^*$,

$$f_j(u)\bar{Z}(u)[W_{j+1}(u) - r_j] - r_j\bar{F}_j(u)z(u) \geq 0.$$

Hence the derivative of dv_t/du is non-negative over some interval $u \geq u^*$ and non-positive for $u \leq u^*$, as was to be shown.

The above argument says that at any time t , the optimal strategy is either to return immediately or wait for another response. Let t_j be the infimum of all values of t such that it is better to wait for the $(j + 1)$ st source at time t . It must be the case that for every $t > t_j$ the aggregator should wait for another response and for every $t < t_j$ the aggregator should return immediately.

It follows that

$$V_j(t) = \max \left(R_j(t), \int_{x=t}^{\infty} V_{j+1}(x) \frac{f_j(x)}{\bar{F}_j(t)} dx \right).$$

Equivalently,

$$\begin{aligned} W_j(t) &= \max \left(r_j, \int_{x=t}^{\infty} \frac{V_{j+1}(x)}{\bar{Z}(t)} \frac{f_j(x)}{\bar{F}_j(t)} dx \right) \\ &= \max \left(r_j, \int_{x=0}^{\infty} \frac{V_{j+1}(t+x)}{\bar{Z}(t)} \frac{f_j(t+x)}{\bar{F}_j(t)} dx \right) \\ &= \max \left(r_j, \int_{x=0}^{\infty} W_{j+1}(t+x) \frac{\bar{Z}(t+x)}{\bar{Z}(t)} \frac{\bar{F}_j(t+x)}{\bar{F}_j(t)} \frac{f_j(t+x)}{\bar{F}_j(t+x)} dx \right). \end{aligned}$$

Now the expression within the integral is increasing in t : $W_{j+1}(t+x)$ is by the induction hypothesis; $\frac{\bar{Z}(t+x)}{\bar{Z}(t)} \frac{\bar{F}_j(t+x)}{\bar{F}_j(t)}$ is by Lemma 2 and the fact that $F_j \cdot Z$ is DFR; and $\frac{f_j(t+x)}{\bar{F}_j(t+x)}$ is since F_j is IFR. It follows that the integral is increasing in t and hence so is $W_j(t)$, completing the induction. \square

3.1 Decreasing Failure Rate Return Times

Naturally, we ask whether the above result can be extended in the alternative case where sources have a decreasing failure rate. Such a model is reasonable in this setting; it may be that once we have not heard from a source, there is reason to believe it is busy or broken and might not respond for some time. We prove a result similar to Theorem 1. We also provide a second result that loosens the restrictions on F and Z , but requires additional restrictions on the values r_j .

Theorem 2 *If F has decreasing failure rate, Z has increasing failure rate, and $F_m \cdot Z$ has increasing failure rate, then in the optimal plan for each number of responses $j \geq m - 1$ there is a single transition point. Specifically, an aggregator with j responses should wait only until some time t_j for another response.*

Proof:

The proof follows Theorem 1, with the directions “reversed.” The goal of the backwards induction is now to show:

- With j sources, the optimal strategy is to wait only until some time t_j for another response.
- The function $W_j(t) = V_j(t)/\bar{Z}(t)$ is decreasing in t .

The above are both true for $j = n$ (here $t_n = -\infty$). The first part of the induction will also go through for $j = m - 1$, and hence the induction yields the theorem.

Similarly to Theorem 1, we have that if $F_j \cdot Z$ is IFR, so is $F_k \cdot Z$ for all $k \geq j$ under these conditions.

Following Theorem 1, for any fixed t again let $v_t(u)$ be the value obtained by the aggregator if, starting at time t , it chooses the policy of waiting until time $u \geq t$ and then returning. As in Theorem 1 we find

$$\frac{dv_t}{du} = \frac{f_j(u)\bar{Z}(u)[W_{j+1}(u) - r_j] - r_j\bar{F}_j(u)z(u)}{F_j(t)}$$

Suppose t_j is the infimum of all values such that $\frac{dv_t}{du} \leq 0$. Equivalently,

$$\frac{f_j(t_j)}{\bar{F}_j(t_j)}[W_{j+1}(t_j) - r_j] \geq r_j \frac{z(t_j)}{\bar{Z}(t_j)}.$$

The left hand side above is decreasing in t_j , as F_j is DFR, $W_{j+1} \geq r_{j+1} \geq r_j$, and W_{j+1} is decreasing by the inductive hypothesis. The right hand side is increasing, as Z is IFR. Hence for all $u \geq t_j$,

$$f_j(u)\bar{Z}(u)[W_{j+1}(u) - r_j] - r_j\bar{F}_j(u)z(u) \geq 0.$$

Hence the derivative of dv_t/du is non-negative over some interval $u \geq t_j$ and non-positive for $u \leq t_j$. Hence v_t first increases to t_j and then decreases, from which we conclude there

is a point t_j that the aggregator should wait until before returning. Further note that the numerator of dv_t/du is independent of t , and hence the value t_j is in fact independent of t ; that is, the value t_j completely summarizes the optimal plan for the aggregator when it has j responses.

We now need to show that $W_j(t)$ is decreasing up to t_j . To determine the sign of the derivative of $W_j(t)$, we begin by considering the value of $W_j(t + \Delta) - W_j(t)$, where t and $t + \Delta$ are less than t_j ; indeed, we will think of $\Delta > 0$ as going to 0.

$$\begin{aligned} W_j(t + \Delta) - W_j(t) &= \int_{x=t+\Delta}^{t_j} \frac{f_j(x)V_{j+1}(x)}{\bar{F}_j(t + \Delta)\bar{Z}(t + \Delta)} dx - \int_{x=t}^{t_j} \frac{f_j(x)V_{j+1}(x)}{\bar{F}_j(t)\bar{Z}(t)} dx \\ &\quad + R_j(t_j)\bar{F}_j(t_j) \left(\frac{1}{\bar{F}_j(t + \Delta)\bar{Z}(t + \Delta)} - \frac{1}{\bar{F}_j(t)\bar{Z}(t)} \right). \end{aligned}$$

Let us examine the first two terms on the right hand side. Note

$$\int_{x=t}^{t_j} \frac{f_j(x)V_{j+1}(x)}{\bar{F}_j(t)\bar{Z}(t)} dx = \int_{x=0}^{t_j-t} W_{j+1}(x+t) \frac{f_j(x+t)}{\bar{F}_j(x+t)} \frac{\bar{F}_j(x+t)\bar{Z}(x+t)}{\bar{F}_j(t)\bar{Z}(t)}.$$

The product inside the integral is decreasing in t for a fixed x by Lemma 2 and our assumptions on F_j and Z . Hence

$$\begin{aligned} \int_{x=t+\Delta}^{t_j} \frac{f_j(x)V_{j+1}(x)}{\bar{F}_j(t+\Delta)\bar{Z}(t+\Delta)} dx - \int_{x=t}^{t_j} \frac{f_j(x)V_{j+1}(x)}{\bar{F}_j(t)\bar{Z}(t)} dx &= \int_{x=0}^{t_j-t-\Delta} \frac{f_j(x+t+\Delta)W_{j+1}(x+t+\Delta)\bar{Z}(x+t+\Delta)}{\bar{F}_j(t+\Delta)\bar{Z}(t+\Delta)} dx \\ &\quad - \int_{x=0}^{t_j-t} \frac{f_j(x+t)W_{j+1}(x+t)\bar{Z}(x+t)}{\bar{F}_j(t)\bar{Z}(t)} dx \\ &\leq - \int_{x=t_j-t-\Delta}^{t_j-t} \frac{f_j(x+t)W_{j+1}(x+t)\bar{Z}(x+t)}{\bar{F}_j(t)\bar{Z}(t)} dx \\ &= - \int_{x=t_j-\Delta}^{t_j} \frac{f_j(x)W_{j+1}(x)\bar{Z}(x)}{\bar{F}_j(t)\bar{Z}(t)} dx. \end{aligned}$$

Now consider $(W_j(t + \Delta) - W_j(t))/\Delta$ in the limit as Δ goes to 0 (from the right).

$$\frac{W_j(t + \Delta) - W_j(t)}{\Delta} \leq - \int_{x=t_j-\Delta}^{t_j} \frac{f_j(x)W_{j+1}(x)\bar{Z}(x)}{\bar{F}_j(t)\bar{Z}(t)\Delta} dx + \frac{1}{\Delta} \left(\frac{R_j(t_j)\bar{F}_j(t_j)}{\bar{F}_j(t+\Delta)\bar{Z}(t+\Delta)} - \frac{R_j(t_j)\bar{F}_j(t_j)}{\bar{F}_j(t)\bar{Z}(t)} \right).$$

Taking the limit gives

$$\frac{dW_j(t)}{dt} \leq - \frac{f_j(t_j)W_{j+1}(t_j)\bar{Z}(t_j)}{\bar{F}_j(t)\bar{Z}(t)} + \frac{(R_j(t_j)\bar{F}_j(t_j))(f_j(t)\bar{Z}(t) + \bar{F}_j(t)z(t))}{(\bar{F}_j(t)\bar{Z}(t))^2},$$

where we have replaced the limiting terms by the appropriate derivatives. Recall that

$$f_j(t_j)W_{j+1}(t_j)\bar{Z}(t_j) = r_j f_j(t_j)\bar{Z}(t_j) + r_j z(t_j)\bar{F}_j(t_j),$$

and $R_j(t) = r_j \bar{Z}(t)$, so simplifying the above yields

$$\frac{dW_j(t)}{dt} \leq \frac{r_j \bar{F}_j(t_j) \bar{Z}(t_j)}{\bar{F}_j(t) \bar{Z}(t)} \cdot \left(\frac{f_j(t)}{\bar{F}_j(t)} + \frac{z(t)}{\bar{Z}(t)} - \frac{f_j(t_j)}{\bar{F}_j(t_j)} - \frac{z(t_j)}{\bar{Z}(t_j)} \right).$$

Hence $\frac{dW_j(t)}{dt}$ is non-positive if

$$\frac{f_j(t_j)}{\bar{F}_j(t_j)} + \frac{z(t_j)}{\bar{Z}(t_j)} - \frac{f_j(t)}{\bar{F}_j(t)} - \frac{z(t)}{\bar{Z}(t)} \geq 0.$$

Since the failure rate of $F_j \cdot Z$ is increasing, the above holds, and the induction goes through. \square

If the reward function has exponential decay, then Theorem 1 and Theorem 2 both hold for any number of responses if the return distribution times are exponential. We therefore obtain the following corollary:

Corollary 1 *If $Z(t)$ and $F(t)$ are given by exponential distributions, then the optimal strategy for the aggregator is to always wait for k sources to respond for some k .*

Proof: In this case, Theorem 1 says that for each number of responding sources we return immediately or wait forever, while Theorem 2 says that for each number of responding sources we wait until some fixed time and return. Both can only hold if for j responding sources we either always return immediately or always wait for another source, independent of the time passed. Hence the aggregator always waits for a fixed number of sources to respond. \square

This corollary can also be proved directly using the algebraic properties of the exponential distribution, but the above proof based on first principles is appealing.

Our second theorem for this case (inspired by [18]) places restrictions on the values r_k . In stopping time nomenclature, it is an example of an infinitesimal stopping rule [14], where the stopping rule is easily expressed in terms of the input parameters.

Theorem 3 *If F has decreasing failure rate and Z has increasing failure rate, and moreover $\frac{r_{j+1}}{r_j}$ is decreasing in j for $j \geq k$, then in the optimal plan there is a time t_j for each $j \geq k$ so that an aggregator with j responses should wait only until time t_j for another response. Here t_j is the first time such that $\frac{f_j(t_j)}{\bar{F}_j(t_j)} \left(\frac{r_{j+1}}{r_j} - 1 \right) \leq \frac{z(t_j)}{\bar{Z}(t_j)}$, or $t_j = \infty$ if no such time exists.*

Proof: Again the proof follows a reverse induction. Recall

$$\frac{dv_t}{du} = \frac{f_j(u) \bar{Z}(u) [W_{j+1}(u) - r_j] - r_j \bar{F}_j(u) z(u)}{\bar{F}_j(t)}.$$

This is positive whenever

$$\frac{f_j(u)}{\bar{F}_j(u)} \left(\frac{W_{j+1}(u)}{r_j} - 1 \right) \geq \frac{z(u)}{\bar{Z}(u)}.$$

Let t_j be the first time where

$$\frac{f_j(t_j)}{\bar{F}_j(t_j)} \left(\frac{r_{j+1}}{r_j} - 1 \right) \geq \frac{z(t_j)}{\bar{Z}(t_j)}.$$

(If no such time exists, we take $t_j = \infty$.) We claim that the optimal plan is to wait until time t_j and then stop. Since $W_{j+1}(u) \geq r_{j+1}$, F_j has decreasing failure rate, and Z has increasing failure rate, at all times $u \leq t_j$ we have that $\frac{dW_t}{du}$ is positive. Hence we should always wait until time t_j .

We now claim that $t_{j+1} \leq t_j$. This holds trivially when $j = n - 1$, and holds for other values of j from the definition as $\frac{r_{j+1}}{r_j}$ is decreasing in j and $\frac{f_j(x)}{F_j(x)} \geq \frac{f_{j+1}(x)}{F_{j+1}(x)}$. Since $t_{j+1} \leq t_j$, by a reverse induction $W_{j+1}(t_j) = r_{j+1}$. This implies the aggregator need not wait past time t_j for an optimal schedule. \square

4 Related Cases

Up to this point, we have considered the case where all sources were indistinguishable, both in terms of the distribution of their response time and the value of their response. It is natural to ask if we can generalize to less stringent cases. We show that Theorems 1 and 2 can be generalized naturally if either the distributions are different or the value corresponding to individual sources are different, as we show below. In these cases, however, there may be a transition point for each possible subset of sources, so the size to represent the optimal plan may grow exponentially in the number of sources n . This may still be suitable in some situations, where the number of sources is small (in the tens) but there may be some variability in either response time or the value of answers. For convenience, we state our results as variations of Theorem 1 only.

To begin, consider the case where for a subset S of sources there is an associated value $R_S(t) = r_S \bar{Z}(t)$. For example, each source e could have its own corresponding value r_e , and the value r_S would be the sum of the r_e for the set. Of course the r_S should be increasing, in the sense that if $S \subset T$ then $r_S \leq r_T$. All sources are governed by the same distribution for the response time.

Theorem 4 *Consider the setting where $R_S(t) = r_S \bar{Z}(t)$, the r_S are increasing, F has increasing failure rate, Z has decreasing failure rate, and $F_m \cdot Z$ has decreasing failure rate. In the optimal plan for each set of respondents S with $|S| \geq m - 1$ there is a single transition point. On receiving a response the aggregator should either return immediately or wait for the next response.*

Proof: Let $V_S(t)$ be the expected payoff to the aggregator using the optimal strategy given that at time t all the sources in S have been heard from. We use a similar backward induction as in Theorem 1. Consider any set S of sources of size $j < n$. Let \mathcal{T} be the set of all sets of the form $\{S\} \cup e$ for some $e \notin S$, and $V_{\mathcal{T}}(t) = \sum_{T \in \mathcal{T}} V_T(t) / |\mathcal{T}|$. Then, following Theorem 1,

consider $v_t(u)$ to be the value obtained by the aggregator if, starting at time t , it chooses the policy of waiting until time $u \geq t$ for some fixed set S . We have

$$v_t(u) = \int_{x=t}^u V_{\mathcal{T}}(x) \frac{f_j(x)}{\bar{F}_j(t)} dx + \frac{\bar{F}_j(u)}{\bar{F}_j(t)} R_S(u).$$

The proof now follows Theorem 1, with $V_{\mathcal{T}}(x)$ and $W_{\mathcal{T}}(x)$ replacing $V_{j+1}(x)$ and $W_{j+1}(x)$ as appropriate. The point is that if each $W_{\mathcal{T}}(t)$ is decreasing, so is their average. \square

In the second case, the return value function $R_j(t) = r_j \bar{Z}(t)$ depends only on the number of sources, but the return time distribution F_e for each source e may vary, under the constraint that all distributions are IFR. We let F_S be the distribution of the time until the first response for all sources excluding those in S , and similarly define f_S appropriately.

Theorem 5 *Consider the setting where each return distribution F_e has increasing failure rate, Z has decreasing failure rate, and $F_S \cdot Z$ has decreasing failure rate for some subset S of the sources. In the optimal plan for each set of respondents T satisfying $S \subseteq T$ there is a single transition point. Moreover, on receiving a response the aggregator should either return immediately or wait for the next response.*

Proof: Following Theorem 1, consider $v_t(u)$ to be the value obtained by the aggregator if, starting at time t , it chooses the policy of waiting until time $u \geq t$ for some fixed set Q . We have

$$v_t(u) = \int_{x=t}^u V_{j+1}(x) \frac{f_Q(x)}{\bar{F}_Q(t)} dx + \frac{\bar{F}_Q(u)}{\bar{F}_Q(t)} R_j(u).$$

There is now a backwards induction on supersets of S that goes through exactly as in Theorem 1. \square

We do not have a proof for the case when both the response times and the values may vary across sources. The problem in this case is that as time changes which source is most likely to return may change, and this may then affect the value of waiting for the next response. Hence it is not clear if in fact both response times and response values vary that the optimal plan continues to have such a nice form; a stronger condition may be required.

The above theorems may be more useful in settings where we have many sources but a small number of types of sources. For example, suppose we have a small set of extremely valuable sources and a larger set of less valuable sources, where all sources have the same response time distribution. If the value to the aggregator is just a function of the number of extremely valuable and less valuable sources that have been heard from, then we may use Theorem 4. In this case, the optimal plan will have one transition point associated with each pair (x, y) , where x is the number of extremely valuable sources and y is the number of less valuable sources that have responded. In this case the size of the optimal plan is no longer exponential in the number of sources.

5 A Counterexample for a Weaker Distribution Property

For completeness, we provide a counterexample to show that some strong condition is necessary for our theorems to apply. In all of our theorems, the case where we have heard from $n - 1$ sources is the easy first step of our backward induction. In Theorem 1, for example, just the fact that F has increasing failure rate and Z has decreasing failure rate is sufficient to guarantee that there is only one transition point when $n - 1$ sources have responded. Our counterexample shows that a weaker condition on the response times than Theorem 1 fails to guarantee that optimal plans have only a single transition point even with $n - 1$ responses.

The mean residual life of X at time t is defined as $m_X(t) = E[X - t \mid X > t]$. We define $m_X(t)$ to be 0 where $\bar{F}(t) = 0$. The random variable X is said to have *decreasing mean residual life* or be DMRL if $m_X(t)$ is decreasing. It is easy to show that if X is IFR then it is DMRL, but the reverse need not hold. We here show that if the response time distributions for sources have decreasing mean residual life, the optimal plan may have more than one transition point for each number of sources heard from.

For our counterexample, we use the fact that the uniform distribution over the range $[0, 2] \cup [4, 12]$ is DMRL. Call this distribution D . We consider a simple problem with two sources with response time distribution D . The reward is governed by an exponential decay over time. With no responses, the reward is always 0. With one response, at time t the reward is e^{-t} . With two responses, at time t the reward is $10e^{-t}$. The idea of this counterexample is that by making the reward large for receiving two response, we encourage the aggregator to wait if it has received a single response up to the gap where responses will not arrive. At the same time, the gap dissuades an aggregator from waiting if one response comes in early enough.

The following four facts are easy to verify using computation and a case analysis, as shown in the Appendix.

- At time 4 and higher, if the aggregator has one response, it should always wait for a second.
- At time 2, if the aggregator has one response, it should return rather than wait at least two time units for the next response.
- For times in the range $[0.5, 2]$, if the aggregator receives a first response, the expectation for the aggregator is larger if it waits until time 2 for the second response and then returns at time 2 if it fails to arrive.
- For times in the range $[0, 0.4]$, if the aggregator receives a first response, the expectation for the aggregator is larger if it returns immediately rather than waits until time 2 for a second response.

More specifically, there are in fact two transition points, one at about 0.406 and one at 2.

Although the support of the response distribution is not all positive numbers in our example, we could easily modify the distribution to have this property, keeping the density in portions outside the intervals suitably small. Also, we could use distributions consisting of more disjoint intervals to make the optimal schedule even more complex.

6 Conclusion

We have introduced the natural problem of aggregation, which arises in Web search engines as well as other distributed systems. The aggregation paradigm appears quite general; we believe it will prove a foundational framework for several applications. Our focus in this work has been to determine appropriate probabilistic assumptions under which the optimal plan for the aggregator has a simple form.

Our work suggests many further questions for future study; we list some possibilities. It would be interesting to know how complex an optimal solution can be, or if there is a natural way to tie the complexity of the response distributions to the complexity of the optimal plan. Proving the existence of and designing algorithms for finding approximately optimal plans with low complexity in terms of the number of transition points would also be useful. Designing on-line systems that deal properly with widely heterogeneous sources, either exactly or with approximate algorithms, may be possible; the full plan would not need to be computed in advance, but proper responses might be calculated on the fly. Extending results to cases where response time distributions may be correlated in non-trivial ways might be important for practical applications.

References

- [1] M. Ajtai, N. Megiddo, and O. Waarts. Improved Algorithms and Analysis for Secretary Problems and Generalizations. *SIAM Journal on Discrete Mathematics*, 14(1):1-27, 2001.
- [2] S. Albers and G. Schmidt. Scheduling with Unexpected Machine Breakdowns. In *Lecture Notes in Computer Science 1671*, pages 269-280, 1999.
- [3] B. Awerbuch, Y. Azar, A. Fiat, and T. Leighton. Making Commitments in the Face of Uncertainty: How to Pick a Winner Almost Every Time. In *Proceedings of the 28th Annual ACM Symposium on Theory of Computing*, pages 519-530, 1996.
- [4] D. Bertsekas and J. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, Belmont, MA, 1996.
- [5] J. A. Boyan and M. Mitzenmacher. Improved Results for Stochastic Transportation Networks. In *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 895-902, 2001.
- [6] Y. S. Chow, H. Robbins, and D. Siegmund. *Great Expectations: The Theory of Optimal Stopping*, Houghton Mifflin Company, Boston, 1971.

- [7] M. Datar and A. Ranade. Commuting with Delay Prone Buses. In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 22–29, 2000.
- [8] E. B. Dynkin. *Markov processes and related problems of analysis*, Cambridge University Press, Cambridge, 1981.
- [9] T. Ferguson. A Poisson Fishing Model. In *Festschrift for Lucien Le Cam*, pp. 234-244, Springer, 1997.
- [10] B. Kalyanasundaram and K. R. Pruhs. Fault-Tolerant Scheduling. In *Proceedings of the 26th Annual ACM Symposium on Theory of Computing*, pages 115–124, 1994.
- [11] B. Kalyanasundaram and K. R. Pruhs. Fault-Tolerant Real Time Scheduling. In *Proceedings of the 1997 European Symposium on Algorithms*, pages 296–307, 1997.
- [12] K. C. Kapur and L. A. Lamberson. *Reliability in Engineering Design*. John Wiley and Sons, New York, 1977.
- [13] M. L. Puterman. *Markov Decision Processes—Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, NY, 1994.
- [14] S. Ross. Infinitesimal Look-Ahead Stopping Rules. *The Annals of Mathematical Statistics*, 42(1):297-303, 1971.
- [15] S. Ross. *Stochastic Processes*. John Wiley and Sons, New York, 1983.
- [16] M. Shaked and J. G. Shanthikumar. *Stochastic Orders and Their Applications*. Academic Press, San Diego, 1994.
- [17] N. Starr and M. Woodroffe. Gone fishin’: Optimal stopping based on catch times. University of Michigan Technical Report, Department of Statistics, 1974.
- [18] N. Starr, R. Wardrop, and M. Woodroffe. Estimating a mean from delayed observations. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebeite*, 35:103-116, 1976.
- [19] D. Stoyan. *Comparison Methods for Queues and Other Stochastic Models*. John Wiley and Sons, Berlin, 1983.

Appendix: A Counterexample

For our counterexample, we use the fact that the uniform distribution over the range $[0, 2] \cup [4, 12]$ is DMRL. Call this distribution D . We consider a simple problem with two sources with response time distribution D . The reward is governed by an exponential decay over time. With no responses, the reward is always 0. With one response, at time t the reward is e^{-t} . With two responses, at time t the reward is $10e^{-t}$. The idea of this counterexample is that by making the reward large for receiving two response, we encourage the aggregator to wait up

to the gap where responses will not arrive if it has received a single response. The gap will dissuade an aggregator from waiting, however, if one response comes in early enough. There are two transition points, one at about 0.406 and one at 2. We simply show that there are at least two transition points.

First, we show that at time 4 and higher, if the aggregator has one response, it should always wait for a second. In the language of Theorem 1, for $t \geq 4$ the aggregator should wait if

$$\frac{f_j(t)}{F_j(t)}[W_{j+1}(t) - r_j] \geq \frac{z(t)}{Z(t)}.$$

In this case $\frac{f_j(t)}{F_j(t)} \geq \frac{1}{8}$, $W_{j+1}(t) - r_j \geq 9$, and $\frac{z(t)}{Z(t)} = 1$. Hence the aggregator should wait.

Second, at time 2, if the aggregator has one response, it should return rather than waiting at least two time units for the next response. The reward by returning is e^{-2} . If the aggregator waits, from the above paragraph it should wait until the second arrival. Comparing the rewards we find

$$e^{-2} \geq \int_{t=4}^{12} \frac{10}{8} e^{-t} dt.$$

Third, consider the case where the aggregator has one response at a time u in the range $[0, 2]$. We compare rewards if the adversary waits until time 2 and if it returns immediately. If it returns immediately, the reward is e^{-u} . If it waits, the reward is

$$\left(\int_{t=u}^2 \frac{10}{10-u} e^{-t} dt \right) + \frac{8}{10-u} e^{-2}.$$

Graphing the two results numerically shows that it is better to wait for times in the range $[0.407, 2]$ but return for times in the range $[0.405, 2]$.