Whittle Index Policy for Crawling Ephemeral Content

Konstantin Avrachenkov, Vivek Borkar

Whittle Index Policy for Crawling Ephemeral Content

Konstantin E. Avrachenkov and Vivek S. Borkar

***Abstract*—We consider the task of scheduling a crawler to retrieve from several sites their ephemeral content. This is content, such as news or posts at social network groups, for which a user typically loses interest after some days or hours. Thus development of a timely crawling policy for ephemeral information sources is very important. We first formulate this problem as an optimal control problem with average reward. The reward can be measured in terms of the number of clicks or relevant search requests. The problem in its exact formulation suffers from the curse of dimensionality and quickly becomes intractable even with moderate number of information sources. Fortunately, this problem admits a Whittle index, a celebrated heuristics which leads to problem decomposition and to a very simple and efficient crawling policy. We derive the Whittle index for a simple deterministic model and provide its theoretical justification. We also outline an extension to a fully stochastic model.**

I. Introduction

Nowadays an overwhelming majority of people find new information on the web at news sites, blogs, forums and social networking groups. Moreover, most information consumed is ephemeral in nature, that is, people tend to lose their interest in the content in several days or hours. The interest in a content can be measured in terms of clicks or number of relevant search requests. It has been demonstrated that the interest decreases exponentially over time [11], [16], [20].

In a series of works (see e.g., [8], [10], [9], [23] and references therein) the authors address the problem of refreshing documents in a database. However, these works do not consider the ephemeral nature of the information. Motivated by this challenge, the authors of [16] suggest a procedure for optimal crawling of ephemeral content. Specifically, they formulate an optimization problem for finding optimal frequencies of crawling for various information sources.

The approach presented in [16] is based on static optimization. In particular, it does not depend on any 'state variable(s)' evolving with time. With a dynamic policy, for instance, if there is not much new material on the principal information sources, the crawler could spend some time to crawl the sources with less popular content but which nevertheless bring noticeable rate of clicks or increase information diversity. In fact, in the numerical results section we demonstrate the superior performance of the Whittle index policy in comparison with some commonly used static policies. Therefore, in the present work we investigate a dynamic formulation of the problem as an optimal control problem with average reward. The direct application of dynamic programming quickly becomes intractable even with moderate number of information sources, due to the so-called curse of dimensionality. Fortunately, the problem admits a Whittle index, which leads to problem decomposition and to a very simple and efficient crawling policy. We derive the Whittle index and provide its theoretical justification first for the deterministic model. We also propose an online algorithm for the relaxed control, which can be used in the case of soft constraints and can provide bounds on the performance of the Whittle index policy. Finally, we study the Whittle index approach in the fully stochastic case when the index does not have an explicit form but can be estimated online.

In [5], [17], [29] the authors study the interaction between the crawler and the indexing engine by means of optimization and control theoretic approaches. One of interesting future research directions is to take into account the indexing engine dynamics in the present context.

The general concept of the Whittle index was introduced by P. Whittle in [32]. This has been a very successful heuristic for restless bandits, which, while suboptimal in general, is provably optimal in an asymptotic sense [30], [31] and has good empirical performance. It and its variants have been used extensively in logistical and engineering applications, some recent instances of the latter in communications and control being for sensor scheduling [21], multi-UAV coordination [22], congestion control [3], [4], [14], channel allocation in wireless networks [15], cognitive radio [18] and real-time wireless multicast [26]. Book length treatments of indexable restless bandits appear in [13], [27].

This is a significantly extended version of the conference paper [2]. We have supplied the detailed proofs omitted in *ibid.* and filled in a gap in the Whittle idexability argument. We also take this opportunity to correct the flawed treatment of the stochastic case in *ibid.* (see new section VIII below). That section has a large part of new material on the Whitttle index in the fully stochastic case, in particular, on its online estimation. We have also added a completely new section on the relaxed control problem (section VII). We have also significantly reworked the numerical results section adding comparison with the other heuristic policies.

II. Model

There are N sources of ephemeral content. A content at source $i \in \{1,...,N\}$ is published with an initial utility modelled by a nonnegative random variable ξ_i and decreasing exponentially over time with a deterministic rate μ_i. This model of exponential evolution of the content value was

This work was partially supported by Indo-French CEFIPRA Collaboration Grants No.5100-IT1 "Monte Carlo and Learning Schemes for Network Analytics" and IFC/DST-Inria-2016-01/448 "Machine Learning for Network Analytics". This is the author version of the paper published at *IEEE Trans. on Control of Network Systems*, v.5(1), pp.446-455, 2018.

K.E. Avrachenkov is with Inria Sophia Antipolis, 2004 Route des Lucioles, 06902, Sophia Antipolis, France k.avrachenkov@inria.fr

V.S. Borkar is with the Department of Electrical Engineering, IIT Bombay, Powai, Mumbai 400076, India borkar.vs@gmail.com

introduced and justified by measurements in [16]. In practical terms, the value of content can be measured in the number of clicks or in the number of relevant queries. Clearly, this sort of information is available to search engines. We further assume that the new content arrives at source $i \in \{1, ..., N\}$ according to a time-homogeneous Poisson process with rate Λ_i. Thus, if source i's content is crawled τ time units after its creation, its utility is given by $\xi_i \exp(-\mu_i \tau)$. The base utility ξ_i is assumed independent identically distributed across contents at a given source, with a finite mean $\bar{\xi}_i$. It is also assumed independent across sources. We assume that the crawler crawls periodically at multiples of time $T > 0$ and has to choose at each such instant which sources to crawl, subject to a constraint we shall soon specify. When the crawler crawls a content source, we assume that the crawling is done in an exhaustive manner. In such a case, the crawler obtains the following expected reward from crawling the content of source i:

$$u_i = \Lambda_i E[\xi_i \exp(-\mu_i \tau)] = \frac{\Lambda_i \bar{\xi}_i}{\mu_i}(1 - \exp(-\mu_i T)). \quad (1)$$

Set $\alpha_i = \exp(-\mu_i T)$. Let us define the state of source i at time t as the total expected utility of its content, denoted by $X_i(t)$. Then, if we do not crawl source i at epoch t (formally, the control is $v_i(t) = 0$ - we say the source is 'passive'), we obtain zero reward $r_i(X_i(t))v_i(t) = 0$ and the state evolves as follows:

$$X_i(t+1) = \alpha_i X_i(t) + u_i. \quad (2)$$

On the other hand, if we crawl source i (formally, $v_i(t) = 1$ - we say the source is 'active'), we obtain the expected reward $r_i(X_i(t))v_i(t) = X_i(t)$ and the next state of the source is given by

$$X_i(t+1) = u_i. \quad (3)$$

This is *de facto* a deterministic model, even though it captures via equation (1) some effects of stochasticity of the underlying Poisson process. An analogous stochastic model which fully takes into account the randomness coming from the underlying arrival process and varying content value will be discussed in detail in Section VIII.

Our aim is to maximize the long run average reward

$$\limsup_{t \uparrow \infty} \sum_{i=1}^{N} \frac{1}{t} \sum_{m=0}^{t} r(X_i(m))v_i(m) \quad (4)$$

subject to the constraint

$$\limsup_{t \uparrow \infty} \sum_{i=1}^{N} \frac{1}{t} \sum_{m=0}^{t} C_i v_i(m) = M \quad (5)$$

for a prescribed $M > 0$. If $C_i = 1, i = 1, ..., N$, this case can be interpreted as a constraint on the average number of crawled sites per crawling period T and corresponds to the original Whittle framework for restless bandits [32], where this framework arises as a relaxation of a hard constraint of 'at most M' crawled sites in each time slot (we discuss this later in detail.) The case $C_i \neq 1$ is slightly more general and can represent the situation when various sites have different limits on the crawling rates (typically specified in the file 'robots.txt').

This is a constrained average reward control problem [1], [24]. We address this problem in the framework of restless bandits and derive a simple index policy for the problem, which may be viewed as a variant of the celebrated Whittle index. In the next section, we recall the theory of Whittle index.

III. Whittle index

The original formulation of restless bandits is for discrete state space Markov chains, but we consider here Markov chains with closed domains (i.e., closure of an open set) $S_i \subset \mathcal{R}^d, d \geq 1$, as state space. The original motivation for the index policy remains valid nevertheless as long as we justify the associated dynamic programming equation, which we do. A deterministic dynamics such as ours is a special case, albeit degenerate. The fully stochastic case can be handled similarly and is detailed in a separate Section VIII. While in this section we introduce the broader framework in a general set up, we use the same notation as above to highlight the correspondences. This should not cause any confusion.

Thus consider resp. S_i-valued processes $X_i(t), t \geq 0, 1 \leq i \leq N$, each with two possible dynamics, dubbed active and passive, wherein they are governed by transition kernels $p_i(dy|x), q_i(dy|x)$, resp. These are assumed to be continous as maps $x \in S_i \mapsto \mathcal{P}(S_i)$. (:= the space of probability measures on S_i with Prohorov topology). The control at time t is an $A := \{0,1\}^N$-valued vector $v(t) = [v_1(t), \cdots, v_N(t)] \in A$, with the understanding that $v_i(t) = 1 \iff X_i(t)$ is active. In the original restless bandit problem, exactly $N' < N$ processes are active at any given time. The $v_i(t)$ are assumed to be adapted to the history, i.e., the σ-field $\sigma(X_i(s), s \leq t; v_i(s), s < t; 1 \leq i \leq N)$. Let $r_i : S \mapsto \mathcal{R}^+, 1 \leq i \leq N$, be reward functions so that a reward of $r_i(X_i(t))$ is accrued if process i is active at time t. The objective then is to maximize the long run average reward

$$\beta := \limsup_{t \uparrow \infty} \sum_{i=1}^{N} \frac{1}{t} \sum_{s=0}^{t} E[r_i(X_i(s))v_i(s)].$$

This problem has state space $\times_{i=1}^{N} S_i$. Whittle's heuristic among other things reduces the problem to separate control problems on S_i. The idea is to relax the constraint of 'exactly N' are active' to 'on average, N' are active', i.e., to

$$\limsup_{t \uparrow \infty} \frac{1}{t} \sum_{s=0}^{t} E[\sum_{i=1}^{N} v_i(s)] = N'.$$

This makes it a constrained average reward control problem [1], [24] which permits a relaxation to an unconstrained average reward problem by replacing the above reward by

$$\limsup_{t \uparrow \infty} \sum_{i=1}^{N} \frac{1}{t} \sum_{s=0}^{t} E[r_i(X_i(s))v_i(s) + \lambda(N'/N - v_i(s))],$$

where $\lambda \in \mathcal{R}$ is the Lagrange multiplier. Motivated by this, Whittle introduced a 'subsidy' λ for passivity, i.e., a virtual reward for a process in passive mode. Replace the above control problem by N control problems with the ith problem for process $X_i(\cdot)$ seeking to maximize over admissible $v_i(t), t \geq 0$, the reward

$$\limsup_{t\uparrow\infty} \frac{1}{t}\sum_{s=0}^{t} E[r_i(X_i(s))v_i(s) + \lambda(1 - v_i(s))]. \quad (6)$$

The dynamic programming equation for this average reward problem is

$$V_i(x) + \beta = \max\left(\lambda + \int q_i(dy|x)V_i(y),\ r_i(x) + \int p_i(dy|x)V_i(y)\right), \quad (7)$$

where $V_i(x)$ is the value (or bias) function. If this can be rigorously justified (which is not always easy), one defines $B(\lambda)$ as the set of passive states, i.e.,

$$B(\lambda) := \left\{x : \lambda + \int q_i(dy|x)V_i(y) \geq r_i(x) + \int p_i(dy|x)V_i(y)\right\}.$$

If $B(\lambda)$ increases monotonically from ϕ to S_i as λ increases from $-\infty$ to ∞, the problem is said to be *Whittle indexable*. The Whittle index for the ith process in state x_i is then defined as the value of λ for which the active and passive modes are equally desirable, i.e.,

$$\gamma_i(x_i) := \{\lambda' : \lambda' + \int q_i(dy|x_i)V(y) = r_i(x_i) + \int p_i(dy|x_i)V(y)\}.$$

The Lagrange multiplier multiplier λ can be interpreted as a subsidy for passivity. Then, the Whittle index has an interpretation of such subsidy that makes active and passive actions equally attractive at the current state. The so-called '*Whittle index policy*' [32] then is to set $v_i(t) = 1$ for the i with the top N' indices and $v_j(t) = 0$ for the rest.

IV. Dynamic programming equation

In view of the above, the first step is to justify the counterpart of (7) in our context. For this, we first note that $r_i(x) = x, 1 \leq i \leq N$. Further, let $u_i^* := \frac{u_i}{1-\alpha_i} > u_i$. We argue that without loss of generality, we may take $S_i = [u_i, u_i^*]$. To see this, let $X_i(0) = x_0$. If $x_0 \leq u_i$ and is never crawled, it is easy to see that

$$X_i(t) \ = \ \alpha_i^t x_0 + (1-\alpha_i^t)u_i^* \ \uparrow \ u_i^*.$$

On the other hand, if $x_0 > u_i^*$, then

$$X_i(t) = \alpha_i^t x_0 + (1-\alpha_i^t)u_i^* \downarrow u_i^* \text{ as } t \uparrow \infty$$

if never crawled. (Throughout the paper we use $\uparrow$ and $\downarrow$ to denote the limits from below and above, respectively.) It is reset to u_i in either case if there is even a single crawl. Combining these observations and recalling that we consider the long-run average criterion, we conclude that any $x_0 \notin [u_i, u_i^*]$ is a transient state and can be ignored. Thus we set $S_i = [u_i, u_i^*]$.

Henceforth we focus on the average reward problem for source i. We do not delve into the justification for Lagrange multiplier formulation for constrained average reward problem on a general state space, as this is well understood [1], [6], [24]. (In fact, it follows from standard Lagrange multiplier theory applied to the 'occupation measure' formulation of average reward problem which casts it as an abstract linear program. See section 4.2 of [6] which carries out this program for discrete state space and section 3.2 of *ibid.* which describes how to extend the same to general compact Polish state spaces as long as the controlled transition probability kernel is continuous in the initial state and control.) For notational simplicity we drop the index i for the time being. We approach the problem by the standard 'vanishing discount' argument. Thus let $0 < \delta < 1$ be a discount factor and for $k(x, v) := xv + C\lambda(1 - v)$, consider the infinite horizon discounted reward

$$\sum_{m=0}^{\infty} \delta^t k(X(t), v(t)).$$

(Here $C > 0$ replaces the C_i in (5).) Denote the associated value function by

$$V_\delta(\lambda, x) := \sup_{\{v(t)\}, X(0)=x} \left[\sum_{m=0}^{\infty} \delta^t k(X(t), v(t))\right].$$

We drop the argument λ when it is understood from the context. Then V_δ satisfies the discounted reward dynamic programming equation

$$V_\delta(\lambda, x) = \max\left(C\lambda + \delta V_\delta(\lambda, \alpha x + u),\ x + \delta V_\delta(\lambda, u)\right). \quad (8)$$

Lemma 1 The solution of equation (8) has the following properties:
(1) Equation (8) has a unique bounded continuous solution V_δ;
(2) V_δ is Lipschitz uniformly in $\delta \in (0, 1)$;
(3) V_δ is monotone increasing and convex.

Proof: Claim (1) is standard (See Theorem 4.2.3 and bullet 1 in 'Notes on §4.2', Section 4.2, [12]). For (2), take $x, x' \in S$ with $x' > x$. Consider processes $X(t), t \geq 0$, and $X'(t), t \geq 0$, with initial conditions x, x' resp., both controlled by control sequence $v(t), t \geq 0$, that is optimal for the former. Then

$$\begin{aligned} V_\delta(x') - V_\delta(x) \ &\leq \ \sum_{t=0}^{\infty} \delta^m (k(X'(t), v(t)) - k(X(t), v(t))) \\ &= \ \left(\frac{1 - (\alpha\delta)^\tau}{1 - \alpha}\right)(x' - x), \end{aligned}$$

where $\tau :=$ the time of first crawl ($= \infty$ if never crawled). Interchanging the roles of x', x we get a symmetric inequality, whence it follows that

$$|V_\delta(x') - V_\delta(x)| \ \leq \ \left(\frac{1 - (\alpha\delta)^\tau}{1 - \alpha}\right)|x' - x|.$$

For the first part of (3), take $x' > x$ as above and let $X'(t), X(t), t \geq 0$, be processes generated by a common admissible control sequence $\{v(t)\}$ with initial conditions x', x resp. Then it is easy to check that $X'(t) \geq X(t)$ for all t. Therefore

$$\sum_{t=0}^{\infty} \delta^t k(X'(t), v(t)) \geq \sum_{t=0}^{\infty} \delta^t k(X'(t), v(t)). \quad (9)$$

Taking supremum over all admissible controls on both sides, monotonicity of V_δ follows. For convexity, define the finite horizon discounted value function

$$V_n(x) = \sup_{\{v(t)\}, X(0)=x} \sum_{t=0}^{n} \delta^t k(X(t), v(t)).$$

Then it satisfies the dynamic programming equation

$$V_n(x) = \max\left(C\lambda + \delta V_{n-1}(\alpha x + u),\ x + \delta V_{n-1}(u)\right)$$

for $n \geq 1$ with $V_0(x) = x$. The convexity of V_n for each n then follows by a simple induction. Since $V_\delta(x) = \lim_{n\uparrow\infty} V_n(x)$, V_δ is also convex. □

Define $\bar{V}_\delta(x) = V_\delta(x) - V_\delta(u),\ x \in S$. Then by the above lemma, $\bar{V}_\delta$ is bounded Lipschitz, monotone and convex with $\bar{V}_\delta(u) = 0$. Also, $(1-\delta)V_\delta(u)$ is bounded. Using Arzela-Ascoli and Bolzano-Weirstrass theorems, we may pick a subsequence such that $(\bar{V}_\delta, (1-\delta)V_\delta(u))$ converge in $C(S)\times\mathcal{R}$ to (say) (V, β). From (8), we have

$$\bar{V}_\delta(x) + (1-\delta)V_\delta(u) = \max\left(C\lambda + \delta\bar{V}_\delta(\alpha x + u),\ x\right).$$

Passing to the limit along an appropriate subsequence as $\delta \uparrow 1$, we have

$$\begin{aligned} V(x) + \beta &= \max\left(C\lambda + V(\alpha x + u),\ x\right) && (10)\\ &= \max_{v\in\{0,1\}} \Big(vx + (1-v)(C\lambda + V(\alpha x + u))\Big). && (11)\end{aligned}$$

Then (10) is the desired dynamic programming equation for average reward. We study important structural properties of the value function V in the next section.

V. Properties of the Value Function

We begin with the following result.

Lemma 2 The following statements hold:
(1) V is monotone increasing, convex with $V(u) = 0$;
(2) The maximizer on the right hand side of (11) is the optimal control choice at state x and β is the optimal reward.

Proof: Since monotonicity and convexity are preserved in pointwise limits, the first claim is immediate.

For the second point, let $v^*(x)$ denote the maximizer on the r.h.s. of (11), any tie being settled arbitrarily. Then under $\{v(t) = v^*(X(t)),\ t \geq 0\}$,

$$V(X(t)) + \beta = k(X(t), v(t)) + V(X(t+1)). \quad (12)$$

Summing (12) over $t = 1, 2, \cdots, T$, and dividing by T on both sides, then letting $T \uparrow \infty$, we see that β is equal to the average reward under this control policy. On the other hand, for any other control sequence, the equality in (12) will be replaced by $\geq$, leading to the conclusion that β is greater than or equal to the corresponding average reward by an argument similar to the above. This implies the second claim. □

Now define

$$\begin{aligned} B &:= \{x \in S : C\lambda + V(\alpha x + u) > x\},\\ B^c &:= \{x \in S : C\lambda + V(\alpha x + u) \leq x\}.\end{aligned}$$

These are respectively the sets of passive and active states under subsidy λ.

Recall the stopping time $\tau :=$ the time of first crawl. Suppose $\tau < \infty$. (The case $\tau = \infty$ corresponds to '*never crawl*' which we consider separately below.) Under optimal policy, iterating equation (10) τ times leads to

$$V(x) = (C\lambda - \beta)\tau + \left[\alpha^\tau x + \left(\frac{1-\alpha^\tau}{1-\alpha}\right)u - \beta\right].$$

Under any other policy, we would likewise obtain

$$V(x) \geq (C\lambda - \beta)\tau + \left[\alpha^\tau x + \left(\frac{1-\alpha^\tau}{1-\alpha}\right)u - \beta\right].$$

Thus we have the explicit representation for V given by

$$V(x) = \max\left[(C\lambda - \beta)\tau + \left[\alpha^\tau x + \left(\frac{1-\alpha^\tau}{1-\alpha}\right)u - \beta\right]\right],$$

where the maximum is over all admissible control sequences. In particular, this implies:

Lemma 3 Equation (10) has a unique solution.

We shall now eliminate some irrelevant situations.

1) If $u^* \in B$, i.e., the optimal action at u^* is 0, then u^* is a fixed point of the optimally controlled dynamics and the corresponding reward is $C\lambda$. Then $\beta = C\lambda$ and it is optimal to be passive at all states, i.e., $B = [u, u^*]$, $B^c = \phi$, and

$$\lambda \geq \lambda_M := \max_{x\in[u,u^*]} (x - V(\alpha x + u))/C. \quad (13)$$

2) If $u \in B^c$, then from (10), $0 + \beta = u$, i.e., $\beta = u$ and it is optimal to crawl when at u. Then u is a fixed point of the controlled dynamics and it is optimal to be active at all states, i.e., $B^c = [u, u^*]$, $B = \phi$, and

$$\lambda \leq \lambda_m := \min_{x\in[u,u^*]} (x - V(\alpha x + u))/C. \quad (14)$$

Note that since constant policies $v(t) \equiv 0$ and $v(t) \equiv 1$ lead to costs $C\lambda$ and u resp., $\beta \geq (C\lambda) \vee u$ always and $\beta > (C\lambda) \vee u$ for $\lambda \in (\lambda_m, \lambda_M)$. For each λ in (λ_m, λ_M), both B, B^c are non-empty and there exists an $a \in (u, u^*)$ for which the choice of being active or passive is equally desirable.

Lemma 4 The sets B, B^c are of the form $[u, a), [a, u^*]$ for some $a \in [u, u^*]$.

Proof: Since V is convex, one of the following two must hold:

1) For some $a_2 > a_1$, $B = [u, a_1) \cap (a_2, u^*]$ and $B^c = [a_1, a_2]$, or,
2) for some a, $B = [u, a), B^c = [a, u^*]$.

However, since at u^* the optimal action is to crawl, we conclude that $u^* \in B^c$ and only the second possibility can occur. □

Corollary 1 The map $x \mapsto x - V(\alpha x + u)$ is monotone non-decreasing on $[u, u^*]$.

Thus the optimal policy is a threshold policy, viz., it is passive for states below a certain threshold and allows the dynamics $x_{k+1} = \alpha x_k + u$ to evolve uninterrupted till the threshold is crossed, then gets active and resets the state to u. In particular, it is periodic after the first reset and x_k takes only finitely many distinct values thereafter.

Lemma 5 The above problem is Whittle indexable.
Proof: We shall write the value function as $V(\lambda, x)$ to render its λ-dependence explicit. It suffices to show that the optimal threshold $x^*(\lambda)$ increases with λ for $\lambda \in (\lambda_m, \lambda_M)$. Since the optimal policy is threshold, we may restrict attention to threshold policies with a threshold in (u, u^*). In particular, this implies a periodic trajectory with a finite range after the first reset, therefore the 'occupation measure' ν on $[u, u^*]$ given by

$$\nu(A) := \lim_{n \uparrow \infty} \frac{\sum_{k=0}^{n} I\{x_k \in A\}}{n}$$

for A Borel in $[u, u^*]$ is well defined as a probability measure on $[u, u^*]$. Then the corresponding average reward is

$$C\lambda\nu(B) + \int_{B^c} x\nu(dx), \quad (15)$$

an affine function of λ. Moreover, its slope is $\nu(B) <$ some $\Delta < 1$, otherwise the optimal reward $\beta(\lambda)$ would be equal to $C\lambda$, a situation we have excluded from consideration. Since $\beta(\lambda)$ is the supremum of (15) over all threshold policies, it is convex increasing and Lipschitz, hence absolutely continuous, with $\beta'(\lambda) := \frac{d}{d\lambda}\beta(\lambda) \leq C\Delta$ a.e. For $x \in (x^*(\lambda), u^*]$, x is active. Then for $x' := \alpha x + u$,

$$\begin{aligned} x \leq u^* &= \frac{u}{1-\alpha} \\ \Longrightarrow x' &= \alpha x + u \\ &\geq \alpha x + (1-\alpha)x \\ \Longrightarrow x' &\geq x \\ \Longrightarrow x' &\in B^c \quad \text{(by Lemma 4)} \\ \Longrightarrow V(x') &= x' - \beta. \end{aligned}$$

Combining this with (10), we have the following expression for the threshold:

$$x^*(\lambda) = \frac{C\lambda + u - \beta(\lambda)}{1-\alpha}. \quad (16)$$

Since $C > C\Delta \geq \beta'(\lambda)$, this is increasing in λ. Thus the passive set is monotonically increasing, establishing Whittle indexability. □

VI. Derivation of Whittle index

Consider the situation when $\lambda = \gamma(x)$ for a prescribed $x \in (u, u^*)$. It is clear that after the first crawl when the process is reset to u, the optimal $X(t)$ becomes periodic: not crawling and increasing till it hits B^c and then crawling - thereby being reset to u - to repeat the process. Since finite initial patches do not affect the long run average reward, we may then take $X(0) = u$. Define $\eta(x) = \min\{t : X(t) \in B^c\}$. Then

$$\begin{aligned} X(\eta(x)) &= (1 - \alpha^{\eta(x)})u^* & (17) \\ \Longrightarrow \eta(x) &= \left\lceil \log_\alpha^+ \left(1 - \frac{x}{u^*}\right) \right\rceil, & (18) \end{aligned}$$

where $\log_\alpha^+ x = (\log_\alpha x) I\{x > 0\}$. Since by the renewal reward argument the long run average reward is equal to the average reward over one period, we can write

$$\beta = \frac{C\lambda(\eta(x) - 1) + X(\eta(x))}{\eta(x)}, \quad (19)$$

where $\eta(x)$ is given by (18) and $X(\eta(x))$ is given by (17).

We now revert to using the index i to identify the source being referred to. In particular, β_i, λ_i will refer to the optimal reward, resp. Lagrange multiplier, for the ith decoupled problem. Our main result is:

Theorem 1 The Whittle index for our problem is given by

$$\gamma_i(x) := \frac{1}{C_i} \left[\eta_i(x)((1-\alpha_i)x - u_i) + \left(\frac{1 - \alpha_i^{\eta_i(x)}}{1 - \alpha_i} \right) u_i \right],$$

where

$$\eta_i(x) := \left\lceil \log_{\alpha_i}^+ \left(\frac{u_i - (1-\alpha_i)x}{u_i} \right) \right\rceil.$$

Therefore the index policy is to crawl at time t $(= mT$ for some $m \geq 0)$ the top M sources according to decreasing values of $\gamma_i(X_i(t))$, or alternatively, choose a number of top sources for the constraint to be reached.

Remark: Note that if an arm (say, ith) is crawled even once, the corresponding state process $\{X_i(t)\}$ takes only discrete values thereafter. These depend on α_i and u_i alone. In fact this is also true for an arm that is never crawled, except that the discrete values taken will also depend on the initial condition. Therefore we need restrict attention to only these values of x for the argument of $\gamma_i(\cdot)$. This results in a further simplification of the index formula, to

$$\gamma_i(x) = \frac{1}{C_i} \left(\eta_i(x)((1-\alpha_i)x - u_i) + x \right),$$

where $\eta_i(x)$ is as before, but without the $\lceil\ \rceil$ and the argument x of both γ_i and η_i is now restricted to the aforementioned discrete set.

Proof: We drop the subscript i for notational convenience. For $x \in B^c$, (10) leads to $V(x) = x - \beta$. Recall from the proof of Lemma 5 that

$$x' := \alpha x + u \implies V(x') = x' - \beta.$$

Combining this with (10) and the definition of Whittle index implies that for our problem, it is

$$\gamma_i(x) = \frac{(1-\alpha_i)x - u_i + \tilde{\beta}_i(x)}{C_i}, \quad (20)$$

where by virtue of (19), $\tilde{\beta}_i(x) :=$ the optimal cost if one were to set $\lambda_i = \gamma_i(x)$. The latter is given by:

$$\tilde{\beta}_i(x) := \frac{1}{\eta_i(x)} \left\{ C_i \gamma_i(x)(\eta_i(x) - 1) + \left(1 - \alpha_i^{\eta_i(x)}\right) u_i^* \right\},$$

where

$$\eta_i(x) := \left\lceil \log_{\alpha_i}^+ \left(\frac{u_i - (1-\alpha_i)x}{u_i} \right) \right\rceil.$$

Substituting this back into (20), one gets a linear equation for $\gamma_i(x)$ that can be solved to evaluate $\gamma_i(x)$ as

$$\gamma_i(x) := \frac{1}{C_i}\left[\eta_i(x)((1-\alpha_i)x - u_i) + \left(\frac{1-\alpha_i^{\eta_i(x)}}{1-\alpha_i}\right)u_i\right].$$

This completes the proof. □

VII. Online algorithm for the relaxed control problem

If one is satisfied with the time average constraint on the number of crawled sources rather than a hard integer constraint, we can propose an online algorithm based on dual descent. This relaxation of the control problem is also useful to establish an upper bound on the optimal value, therefore also on the Whittle index policy.

In order to apply the dual descent, we first need to derive an expression for the optimal threshold for a stand alone information source. Let us substitute (17) into (19) and differentiate the resulting expression for β with respect to η. Then, equating the derivative to zero, we obtain the following equation with respect to η

$$\alpha^\eta(1-\log(\alpha)\eta) = 1 - C\lambda/u^*.$$

The solution of this equation can be expressed in terms of the Lambert W function ($W(z)\exp(W(z)) = z$) as follows:

$$\eta = \frac{1}{\log(\alpha)}\left(1 + W(e^{-1}(C\lambda/u^* - 1))\right).$$

The optimal threshold value is then given by

$$\bar{x}(\lambda) = u^*(1 - \exp(1 + W(e^{-1}(C\lambda/u^* - 1)))). \quad (21)$$

This allows us to learn the optimal Lagrange multiplier by dual stochastic descent given by

$$\lambda(k+1) = \lambda(n) - b(k)(M/N - \sum_i C_i v_i(k)),\ k \geq 0, \quad (22)$$

where stepsizes $\{b(k)\}$ satisfy the usual conditions for stochastic approximation stepsize:

$$\sum_n b(n) = \infty,\ \sum_n b(n)^2 < \infty.$$

The fact that (22) is indeed a stochastic gradient descent for the Lagrange multiplier follows from the generalized envelope theorem [19] as in [28], Lemma 2, for which it is essential that at each time k, the $v_i(k)$'s be chosen according to the optimal threshold policy corresponding to $\lambda = \lambda(k)$ as derived above. Being a stochastic gradient descent for a convex function, its convergence to the minimizer, i.e., the actual Lagrange multiplier, is easy to establish by standard arguments (see, e.g., [7], chapter 10). Note that the optimal scheme for the relaxed problem is precisely to crawl those sites for which the Whittle index is above the Lagrange multiplier. In online version above, one simply uses the estimate $\lambda(k)$ at time k in lieu of the true Lagrange multiplier.

VIII. Stochastic case

We now consider the fully stochastic situation when the content generation at each source is observed as a random process. In fact, one could also consider mixed situations when some sources are observed and others are not. As we shall see, the development closely mimics the foregoing.

The stochastic system dynamics can be described as follows: Let $\{\tau_n^i\}$ denote the successive arrival times of content at source i, with utilities $\{\xi_n^i\}$, resp. We assume these arrival processes to be Poisson with rate possibly depending on the site i, and $\{\xi_n^i\}$ to be independent and independent of the process of arrival epochs. Further, we assume that they are identically distributed for each fixed i. The net utility added to source i during k-th epoch will be

$$U_i(k) := \sum_{\tau_n^i\ :\ (k-1)T \leq \tau_n^i < kT} \xi_n^i e^{-\mu_i(kT - \tau_n^i)}.$$

Under our hypotheses, the system state at time $(k+1)T$ is then

$$X_i(k+1) = \begin{cases} \alpha_i X_i(k) + U_i(k+1) & \text{if no crawl,} \\ U_i(k+1) & \text{if crawled.} \end{cases} \quad (23)$$

We define the average reward as

$$\limsup_{t\uparrow\infty} \sum_{i=1}^N \frac{1}{t}\sum_{m=0}^t E[r(X_i(t))v_i(t)],$$

which we seek to maximize subject to the constraint

$$\limsup_{t\uparrow\infty} \frac{1}{t}\sum_{i=1}^N C_i E[v_i(t)] = M.$$

Henceforth we drop the subscript i for notational economy, till such time as it is requited. The discounted value function

$$V_\delta(x) := \sup_{\{v(t)\}, X(0)=x} E\left[\sum_{t=0}^\infty \delta^t k(X(t), v(t))\right]$$

then satisfies the dynamic programming equation

$$V_\delta(x) = \quad (24)$$

$$\max\left(C\lambda + \delta\int V_\delta(\alpha x + u)\varphi(du),\ x + \delta\int V_\delta(u)\varphi(du)\right),$$

where φ is the law of $U(t),\ \forall t$.

Lemma 6 The conclusions of Lemma 1 continue to hold.

Proof: The first claim follows as before from the cited results of [12]. For the second, let $X(t), X'(t)$ be defined with identical arrival and control processes with the control process $v(\cdot)$ being optimal for $X(\cdot)$, differing only in their initial conditions resp., x, x'. Then

$$\begin{aligned} V_\delta(x') - V_\delta(x) &\leq E[\sum_{t=0}^\infty \delta^m(k(X'(t), v(t)) \\ &\quad -k(X(t), v(t)))] \\ &\leq E\left[\frac{1-(\alpha\delta)^\tau}{1-\alpha\delta}\right](x' - x), \end{aligned}$$

where $\tau := \min\{t \geq 0 : X(t) = X'(t)\}$ is the coupling time, which is also the time of the first crawl. The Lipschitz property follows as before. Next let $X(t), X'(t)$ be as in the proof of Lemma 1 (3). Taking expectations in (9) followed by a supremum over all admissible controls proves monotonicity. Convexity also follows as in the deterministic case. □

The 'vanishing discount' argument of Section IV can now be used to establish the average cost dynamic programming equation

$$V(x) + \beta = \max(C\lambda + \int V(\alpha x + u)\varphi(du),\ x), \qquad (25)$$

where we have rendered V unique by setting $\int V(u)\varphi(du) = 0$. Monotonicity and convexity properties of V follow as in Lemma 2, using the additional fact that convexity of a function F also implies that of $x \mapsto \int F(ax+y)\nu(dy)$ for any scalar a and probability measure ν on $\mathcal{R}$. Equation (12) gets modified to

$$E[V(X(t))] + \beta = E[k(X(t), v(t))] + E[V(X(t+1))],$$

from which the optimality of

$$v^*(x) \in \text{Argmax}_v \left(vx + (1-v)(\lambda + \int V(\alpha x + u)\varphi(du)) \right),$$

with $x \in S$, follows by arguments analogous to those of Lemma 2. Furthermore, iterating the above, V can be shown to be the unique solution of (25) by establishing the explicit representation

$$\begin{aligned} V(x) &= \max E_x \left[(C\lambda - \beta(\lambda))\sigma + \alpha^\sigma x + \sum_{t=1}^{\sigma} \alpha^{\sigma - t} U(t) \right] \\ &\quad - \beta(\lambda) \\ &= \max E_x \left[(C\lambda - \beta(\lambda))\sigma + \alpha^\sigma x + (1-\alpha^\sigma)u^* \right] \\ &\quad - \beta(\lambda) \end{aligned} \qquad (26)$$

for passive x and

$$V(x) = x - \beta(\lambda) \qquad (27)$$

for active x. Here:

- the maximum is over all admissible control sequences $\{U(t)\}$,
- σ is the first time to cross the threshold, i.e., first time to crawl,
- $u := E[U(m)], u^* := \frac{u}{1-\alpha}$,

and the second equality in (26) follows from the optional sampling theorem. As before, we assume that $\beta(\lambda) > C\lambda \wedge u$ so as to eliminate uninteresting cases. Then the right hand side of (26) will be $< x - \beta(\lambda)$ for $x > u^*$, implying in particular that it is less than the right hand side of (27), that is, x could not have been passive. Then by the same arguments as for the deterministic case, we can restrict the state space to $[u, u^*]$.

The definitions of B, B^c now change to

$$\begin{aligned} B &:= \{x \in S : C\lambda + \int V(\alpha x + u)\varphi(du) > x\}, \\ B^c &:= \{x \in S : C\lambda + \int V(\alpha x + u)\varphi(du) \le x\}. \end{aligned}$$

Making similar arguments as those in Lemma 4, we establish that the optimal control policy is of threshold type.

Lemma 7 The problem is Whittle indexable.

Proof: Given that the maximum in (26) is over all admissible control policies and that the optimal policy is a threshold policy, it is also the maximum over all threshold policies. Fix a threshold, an initial condition $x \in [u, u^*]$ and consider a process corresponding to this threshold policy. Then the random variable σ is independent of λ. Since $\beta'(\lambda) < C$, it then follows that the expectation on the r.h.s. of (26) is monotone increasing in λ. Then the same is true for its maximum over all threshold policies. Letting $x^*(\lambda)$ denote the optimal threshold under λ, by definition both (26), (27) hold for $x = x^*(\lambda)$. Let $G(\lambda, x) :=$

$$\begin{cases} \max E_x \left[(C\lambda - \beta(\lambda))\sigma + \alpha^\sigma x + (1-\alpha^\sigma)u^* \right], & \forall x \in B, \\ x, & \forall x \in B^c, \end{cases}$$

where the maximum is over all threshold policies. Since V is convex increasing, so is $G(\lambda, \cdot)$. Also, G is increasing in λ. Then $x^*(\lambda)$ is a fixed point of $G(\lambda, \cdot)$. Since it is optimal to be passive at $x = u$ (because $\beta(\lambda) > u$), $G(\lambda, u) > u\ \forall \lambda$. On the other hand, it is optimal to be active at u^* and therefore $G(\lambda, u^*) = u^*$. Thus the convex curve $x \mapsto G(\lambda, x)$ crosses the line $y = x$ exactly once in $[u, u^*)$ and the point of crossing, which by definition is $x^*(\lambda)$, increases with λ (because G does). This proves Whittle indexability. □

Let $\tilde{\eta}$ denote the first hitting time ≥ 1 of B^c when $X(0) = U(0)$, i.e., the time to the next crawl after 0. Then

$$X(\eta) = \sum_{t=1}^{\tilde{\eta}} \alpha^{\eta - t} U(t), \quad m \ge 1.$$

From dynamic programming arguments, we know that it suffices to consider stationary Markov controls. Consider $\lambda = \gamma(x)$, the Whittle index of x. Let

$$\eta(x) := E[\tilde{\eta}]$$

and $\tilde{\beta}(x) :=$ the optimal cost, both corresponding to $\lambda = \gamma(x)$. Then by standard renewal-reward arguments,

$$\tilde{\beta}(x) = \frac{C\lambda(\eta(x) - 1) + E[X(\tilde{\eta}(x))]}{\eta(x)}.$$

Define

$$\begin{aligned} Z_0 &= 0, \\ Z_n &:= \sum_{m=1}^{n} \alpha^{n-m} U(m), n \ge 1, \\ &\qquad \text{(e.g., } Z_2 = U(2) + \alpha U(1)), \end{aligned}$$

From the definition of $\gamma(x)$, it follows that $x \in B^c$, so the optimal decision at x is to crawl. By definition of $\gamma(x)$,

$$x = C\gamma(x) + E[V(\alpha x + U(1))].$$

Hence,

$$\begin{aligned} C\gamma(x) &= x - E\left[V(\alpha x + U(1))\right] \\ &= x - \int \varphi(du) E_{\alpha x+u}[(C\gamma(x) - \tilde{\beta}(x))\tau \\ &\quad + \alpha^\tau x + \sum_{t=0}^{\tau} \alpha^{\tau-t} U(t) - \tilde{\beta}(x)] \\ &= x - (C\gamma(x) - \tilde{\beta}(x)) \int \varphi(du) E_{\alpha x+u}[\tau] \\ &\quad + \int \varphi(du) E_{\alpha x+u}[\alpha^\tau] x \\ &\quad - \tilde{\beta}(x) + \int \varphi(du) E_{\alpha x+u}\left[\sum_{t=0}^{\tau} \alpha^{\tau-t} U(t)\right], \end{aligned}$$

which can be solved for $\gamma(x)$ because all three expectations on the right can be calculated from the problem specifications.

One possible computational approach is as follows. If we fix the threshold to be $\breve{x}$, consider V satisfying

$$V(x) = C\lambda - \beta + \int V(\alpha x + u)\varphi(du), \ x < \breve{x}, \quad (28)$$

$$V(x) = x - \beta, \qquad x \geq \breve{x}, \quad (29)$$

$$V(u) = 0. \quad (30)$$

The Whittle index $\gamma(\breve{x})$ should satisfy $\gamma(\breve{x}) = \lambda$ with

$$C\lambda + \int V(\alpha x + u)\varphi(du) - x = 0. \quad (31)$$

Write $V(x)$, which is the unique solution to (28)-(30), as $V(\lambda, x)$ to make its λ-dependence explicit. We can then 'learn' $\gamma(\breve{x})$ by the stochastic approximation scheme

$$\gamma_{n+1} = \gamma_n - a(n)\left(C\gamma_n + V(\gamma_n, \alpha\breve{x} + \tilde{U}_n) - \breve{x}\right), \quad (32)$$

which incrementally adjusts γ_n in the direction of forcing (31) to hold. Here $\{\tilde{U}_n\}$ are i.i.d. random variables with distribution φ. By standard stochastic approximation theory ([7], Chapter 2), this will converge to a solution of (31) as desired. Note that the infinite linear system (28)-(30) has to be solved at each step. This can be done approximately by solving a discretized version of it by a standard linear system solver. Alternatively, one can concurrently solve this discretized version iteratively but on a faster time scale, using the 'two time scale stochastic approximation' framework of [7], Chapter 6. For example, one can replace (32) with

$$\begin{aligned} \gamma_{n+1} &= \gamma_n - a(n)\Big(C\gamma_n + V_n(\gamma_n, \alpha\breve{x} + \tilde{U}_n) - V_n(u) \\ &\quad - \breve{x}\Big), \end{aligned} \quad (33)$$

and augment it with

$$V_{n+1}(x) = C\gamma_n - V_n(u) + \int V_n(\alpha x + u)\varphi(du), \ x < \breve{x}, \quad (34)$$

$$V_{n+1}(x) = x - V_n(u), \qquad x \geq \breve{x}. \quad (35)$$

This is nothing but the classical relative value iteration algorithm [25] but for the time-dependent γ_n. The algorithm is non-incremental, i.e., on the 'natural' time scale dictated by the clock $n = 0, 1, 2, \cdots$, which is faster than the time scale of (33) dictated by the decreasing step sizes $\{a(n)\}$, i.e., 'algorithmic clock' $t(0) = 0, t(1) = a(1), t(2) = a(1) + a(2), \cdots, t(n) = \sum_{m=1}^{n} a(m), \cdots$. (For example, for $a(n) = \frac{1}{n}, t(n) \approx \log n$.) See ([7], chapter 6) for detailed analysis of two time scale algorithms. We summarize only the main conclusions here. Because $\{\gamma_n\}$ are updated on a slower time scale, (34)-(35) see γ_n as quasi-static, i.e., nearly a constant, whence the pair reduces to exactly the classical relative value iteration that would converge to the solution V of (28)-(29) corresponding to $V(u) = \beta$ for λ 'frozen' at γ_n. What this means is that if $\hat{V}^n$ denotes the solution to (28)-(29)-(30), then

$$V_n - V_n(u) - \hat{V}^n \to 0$$

a.s. Thus asymptotically, (33) mimics

$$\gamma_{n+1} = \gamma_n - a(n)\left(C\gamma_n + \hat{V}^n(\gamma_n, \alpha\breve{x} + \tilde{U}_n) - \breve{x}\right), \quad (36)$$

which converges to the Whittle index a.s. Note that this has to be done separately for each $\breve{x}$. In practice one may do so for a finite family of $\breve{x}$'s and interpolate.

IX. Numerical example

Let us illustrate the theoretical results by a small numerical example. The goal of the numerical example is to obtain further insights into the performance of the Whittle index policy rather than to investigate an implementation of the Whittle index policy in a real crawling system. Real implementation is of course a good future direction for research and development. Let us consider four information sources with parameters given in Table I. Without loss of generality, we take the crawling period $T = 1$. One can see how the user interest decreases over time for each source in Figure 1. The initial interest in the content of sources 1 and 2 is high, whereas the initial interest in the content of sources 3 and 4 is relatively small. The interest in the content of sources 1 and 3 decreases faster than the interest in the content of sources 2 and 4.

Let us first consider the deterministic model. We run the system with parameter specified in Table I for 1000 time steps. In Figure 2 we show an example of the state evolution under the Whittle index policy and with the constraint that the crawler can visit only one site per crawling period T, i.e., $M = 1$. The application of Whittle index results in periodic crawling of sources 1 and 2, crawling each with period two. Sources 3 and 4 should never be crawled. This resulted in the average reward 260.30. We can also apply the relaxed control described in Section VII. This gives the average reward 260.96 and, since the relaxed control provides bounds on the deviation from the optimum, this indicates that the Whittle index policy in this particular case gives solution which is very close to the optimal one. While using the relaxed control, the constraint violation appeared in less than 1% of time steps. In contrast, if one always greedily crawls only source 1, he obtains the average reward 179.8.

In Figure 3 we show the state evolution of the bandits under the constraint that the crawler can visit two information sources per crawling period, i.e., $M = 2$. It is interesting that now the policy becomes much less regular. Source 1 is always crawled. Sources 2 and 3 are crawled in a non-trivial periodic way and sources 4 is crawled periodically with a rather long period.

Let us now consider the stochastic model. Since the computation of the Whittle index for the stochastic model is very cumbersome, even numerically, we propose to use deterministic Whittle index for the stochastic model as well. We compare the performance of the deterministic Whittle index policy with the Round Robin (RR) policy and the Greedy-type policy, which chooses the source maximizing

$$\frac{\Lambda_i \bar{\xi}_i}{\mu_i} \left(1 - \exp(-\mu_i \times t_i)\right),$$

where t_i is the time elapsed since the last crawl of source i. To have good averaging, we increased the time interval of simulations to 10000 steps. It is very interesting to observe that if we have randomness only in the content arrival process (we model the content arrival by the Poisson process), the Whittle index policy produces the average reward 253.1, which is smaller than the reward produced by the Greedy-type policy, 260.2. The RR policy results in just 208.4. However, if we introduce randomness also in the initial value of the content (e.g., we now model it as exponential random variable with the parameter ξ_i given in Table I), the Whittle index policy results in the average reward 283.8 and outperforms both the Greedy-type policy, 259.3, and RR policy, 207.3. Thus, we can recommend using the deterministic Whittle policy heuristic when we have high level of randomness in the system.

As example, in Figure 4 we present the state evolution of the stochastic model with dynamics (23), with Poisson arrivals but deterministic value of the content using the deterministic Whittle index policy. One can see that now in the stochastic setting source 1 is crawled from time to time.

TABLE I
DATA FOR NUMERICAL EXAMPLE

i	1	2	3	4
ξ_i	1.0	0.7	0.2	0.08
μ_i	0.7	0.35	0.7	0.21
Λ_i	250	250	250	250

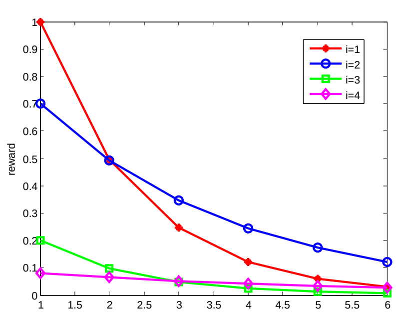

Fig. 1. Content value as a function of time.

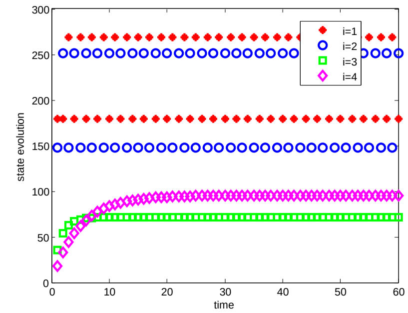

Fig. 2. The case of $M = 1$.

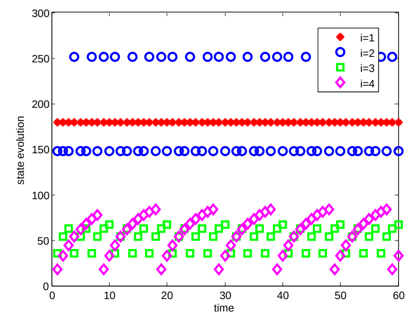

Fig. 3. The case of $M = 2$.

X. Conclusions and Future Work

We have formulated the problem of crawling web sites with ephemeral content as an average reward optimal control problem and have shown that it is indexable. We have found that the Whittle index has a very simple form for a deterministic model and in principle computable for a fully stochastic case. In particular, we have proposed online algorithms for the Whittle index in the fully stochastic case as well as for the relaxed deterministic control. The numerical example demonstrates that the Whittle index policies, unlike the policies suggested in [16], do not generally have a trivial periodic structure. We have compared the Whittle index policy with the Round Robin and Greedy-type policies and have demonstrated that the deterministic Whittle index provides good performance even in the stochastic case with high randomness. The proposed approach can also be used in the cases when some states are observed. In such cases, the Whittle index will act as a self-tuning mechanism. In the future research we plan to elaborate an adaptive version when some parameters (e.g., the rate of new information arrival) need to be estimated online. The other interesting future research directions are addition to the model the dynamics of the indexing engine and practical implementation and testing of the Whittle index approach in real crawling system.

Acknowledgments

The authors gratefully acknowledge the discussions with Liudmila O. Prokhorenkova and Egor Samosvat from Yandex during the preparation of the manuscript.

References

[1] E. Altman, *Constrained Markov Decision Processes*, Chapman and Hall / CRC, London, 1999.

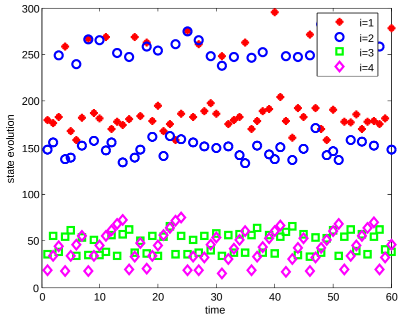

Fig. 4. The case of $M = 2$ (stochastic model).

[2] K. Avrachenkov and V. Borkar, "Whittle Index Policy for Crawling Ephemeral Content", In Proceedings of IEEE CDC 2015.

[3] K. Avrachenkov, U. Ayesta, J. Doncel and P. Jacko, "Congestion Control of TCP Flows in Internet Routers by Means of Index Policy", *Computer Networks*, vol. 57(17), pp. 3463-3478, 2013.

[4] K. Avrachenkov, O. Habachi, A. Piunovskiy and Y. Zhang, "Infinite Horizon Optimal Impulsive Control with Applications to Internet Congestion Control", *International Journal of Control*, vol. 88(4), pp.703-716, 2015.

[5] K. Avrachenkov, A. Dudin, V. Klimenok, P. Nain and O. Semenova, "Optimal Threshold Control by the Robots of Web Search Engines with Obsolescence of Documents", *Computer Networks*, vol. 55(8), pp. 1880-1893, 2011.

[6] V.S. Borkar, "Convex Analytic Methods in Markov Decision Processes", *in 'Handbook of Markov Decision Processes'*, (A. Shwartz and E. Feinberg, eds.), Kluwer Academic, New York, 2002, pp. 347-375.

[7] V.S. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint*, Hindustan Publishing Agency, New Delhi, and Cambridge Uni. Press, Cambridge, UK, 2008.

[8] J. Cho and H. Garcia-Molina, "Synchronizing a Database to Improve Freshness", In Proceedings of ACM SIGMOD 2000, vol. 29(2), pp. 117-128.

[9] J. Cho and H. Garcia-Molina, "Effective Page Refresh Policies for Web Crawlers", *ACM Transactions on Database Systems (TODS)*, vol. 28(4), pp. 390-426.

[10] J. Cho and A. Ntoulas, "Effective Change Detection Using Sampling", In Proceedings of VLDB 2002, pp. 514-525.

[11] A. Goyal, F. Bonchi and L.V. Lakshmanan, "Learning Influence Probabilities in Social Networks", In Proceedings of ACM WSDM 2010, pp. 241-250, 2010.

[12] O. Hernández-Lerma and J.-B. Lasserre, *Discrete Time Markov Control Processes: Basic Optimality Criteria*, Springer Verlag, New York, 1996.

[13] P. Jacko, *Dynamic Priority Allocation in Restless Bandit Models*, Lambert Academic Publishing, 2010.

[14] P. Jacko and B. Sanso, "Congestion Avoidance with Future-Path Information", in Proceedings of EuroFGI Workshop on IP QoS and Traffic Control, IST Press, pp. 153-160, 2007.

[15] M. Larranaga, U. Ayesta and I.M. Verloop, "Stochastic and Fluid Index Policies for Resource Allocation Problems", in Proceedings of IEEE INFOCOM 2015, pp. 1-9.

[16] D. Lefortier, L. Ostroumova, E. Samosvat and P. Serdyukov, "Timely Crawling of High-quality Ephemeral New Content", In Proceedings of CIKM 2013, 27 Oct. - 1 Nov., 2013, San Francisco, pp. 745-750.

[17] Z. Liu and P. Nain, "Optimization Issues in Web Search Engines", In *Handbook of Optimization in Telecommunications*, pp. 981-1015, Springer US, 2006.

[18] K. Liu and Q. Zhao, "Indexability of Restless Bandit Problems and Optimality of Whittle Index for Dynamic Multichannel Access", *IEEE Trans. Info. Theory*, vol. 56(11), 2010, pp. 5547-5567.

[19] P. Milgrom and I. Segal, "Envelope Theorems for Arbitrary Choice Sets", *Econometrica*, 70, 2002, pp. 583-603.

[20] T. Moon, W. Chu, L. Li, Z. Zheng and Y. Chang, "Refining Recency Search Results with User Click Feedback". ArXiv preprint arXiv:1103.3735, 2011.

[21] J. Nino-Mora and S.S. Villar, "Sensor Scheduling for Hunting Elusive Hiding Targets via Whittle's Restless Bandit Index Policy", in Proceedings of NetGCoop 2011, 12-14 Oct., pp. 1-8.

[22] J.L. Ny, M. Dahleh and E. Feron, "Multi-UAV Dynamic Routing with Partial Observations Using Restless Bandit Allocation Indices", in Proceedings of American Control Conf. (ACC 2008), 11-13 June 2008, Seattle, pp. 4220-4225.

[23] C. Olston and M. Najork, "Web Crawling", In *Foundations and Trends in Information Retrieval*, vol. 4(3), pp. 175-246, 2010.

[24] A.B. Piunovskiy, *Optimal Control of Random Sequences in Problems with Constraints*, Springer, 348 pages, 1997.

[25] M.I. Puterman, *Markov Decision Processes*, John Wiley and Sons, Hoboken, NJ, 1994.

[26] V. Raghunathan, V.S. Borkar, M. Cao and P.R. Kumar, "Index Policies for Real-time Multicast Scheduling for Wireless Bradcast Systems", in Proceedings of IEEE INFOCOM 2008, 13-18 April 2008, Phoenix, pp. 2243-2251.

[27] D. Ruiz-Hernandez, *Indexable Restless Bandits*, VDM Verlag, 2008.

[28] N. Salodkar, A. Karandikar and V.S. Borkar, "A Stable Online Algorithm for energy-efficient multi-user scheduling", *IEEE Transactions on Mobile Computing* vol. 9(10), 2010, pp.1391-1406.

[29] J. Talim, Z. Liu, P. Nain and E.G. Coffman, Jr., "Controlling the Robots of Web Search Engines", *Performance Evaluation Review*, vol. 29(1), pp. 236-244, 2001.

[30] I.M. Verloop, "Asymptotically Optimal Priority Policies for Indexable and Non-indexable Restless Bandits", *Annals of Applied Probability*, vol. 26(4), pp. 1947-1995, 2016.

[31] R.R. Weber and G. Weiss, "On an Index Policy for Restless Bandits", *J. Appl. Prob.*, vol. 27, pp. 637-648, 1990.

[32] P. Whittle, "Restless Bandits: Activity Allocation in a Changing World", *J. Appl. Prob.*, vol. 25, pp. 287-298, 1988.

Konstantin E. Avrachenkov received Master degree in Control Theory from St. Petersburg State Polytechnic University (1996), Ph.D. degree in Mathematics from University of South Australia (2000) and Habilitation (Doctor of Science) from University of Nice Sophia Antipolis (2010). Currently, he is a Director of Research at Inria Sophia Antipolis, France. He is an associate editor of International Journal of Performance Evaluation and ACM TOMPECS. His main research interests are Markov processes, singular perturbation theory, queueing theory, mathematical programming, game theory and performance evaluation of communication networks.

Vivek S. Borkar obtained his B.Tech. (Elec. Eng.) from IIT Bombay ('76), M.S. (Systems and Control Eng.) from Case Western Reserve Uni. ('77), and Ph.D. (EECS) from Uni. of California, Berkeley ('80). He has held positions at TIFR Centre for Applicable Math. and Indian Institute of Science in Bangalore and TIFR and IIT Bombay in Mumbai. He is currently an Institute Chair Professor of Electrical Eng. in the latter. He has held visiting positions at Uni. of Twente, MIT, Uni. of Maryland at College Park and Uni. of California at Berkeley. He is a Fellow of IEEE, AMS, TWAS and the science and engineering academies in India. His research interests are stochastic optimization – theory, applications and algorithms, in which he has published widely.