

On the Power of False Negative Awareness in Indicator-based Caching Systems

*Original*

On the Power of False Negative Awareness in Indicator-based Caching Systems / Cohen, Itamar; Einziger, Gil; Scalosub, Gabriel. - ELETTRONICO. - (2021), pp. 46-56. ( INTERNATIONAL CONFERENCE ON DISTRIBUTED COMPUTING SYSTEMS DC, USA 07-10 July 2021) [10.1109/icdcs51616.2021.00014].

*Availability:*

This version is available at: 11583/2978438 since: 2023-05-10T16:32:10Z

*Publisher:*

IEEE COMPUTER SOC

*Published*

DOI:10.1109/icdcs51616.2021.00014

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

IEEE postprint/Author's Accepted Manuscript

©2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

# On the Power of False Negative Awareness in Indicator-based Caching Systems

Itamar Cohen\*, Gil Einziger†, and Gabriel Scalosub†

\*Politecnico di Torino, Italy. Email: itamar.cohen@polito.it

†Ben-Gurion University of the Negev, Beer Sheva, Israel. Email: {gilein,sgabriel}@bgu.ac.il

**Abstract**—Distributed caching systems such as content distribution networks often advertise their content via lightweight approximate indicators (e.g., Bloom filters) to efficiently inform clients where each datum is likely cached. While false-positive indications are necessary and well understood, most existing works assume no false-negative indications. Our work illustrates practical scenarios where false-negatives are unavoidable and ignoring them has a significant impact on system performance. Specifically, we focus on false-negatives induced by indicator staleness, which arises whenever the system advertises the indicator only periodically, rather than immediately reporting every change in the cache. Such scenarios naturally occur, e.g., in bandwidth-constraint environments or when latency impedes each client’s ability to obtain an updated indicator.

Our work introduces novel false-negative aware access policies that continuously estimate the false-negative ratio and sometimes access caches despite negative indications. We present optimal policies for homogeneous settings and provide approximation guarantees for our algorithms in heterogeneous environments. We further perform an extensive simulation study with multiple real system traces. We show that our false-negative aware algorithms incur a significantly lower access cost than existing approaches or match the cost of these approaches while requiring an order of magnitude fewer resources (e.g., caching capacity or bandwidth).

## I. INTRODUCTION

Caches are extensively used in networking environments such as Content Delivery Networks [1]–[3], Named Data Networks [4], 5G networks [5], and Information Centric Networks [6]. In such networks, accessing caches often incurs some overhead in terms of latency, bandwidth, or energy [3], [7]. On the other hand, fetching a datum without caches usually incurs a larger *miss penalty*, e.g., for retrieving the requested item from a remote server [2].

In large distributed systems, caches often optimize performance by advertising their content [1]–[4], [6], [8]. Such advertisements allow clients to minimize costs by selecting which cache to access for a requested datum. Ideally, the advertisement policy would always accurately reflect the up-to-date content at every cache. However, such a solution requires a prohibitive amount of memory, computation, and bandwidth resources. Hence, systems often compromise some accuracy for efficiency by advertising periodical approximate *indicators*. Indicators are data structures that trade accuracy for space efficiency. Common embodiment of such indicators are Bloom filters [7], [9]–[11], and fingerprint hash tables [12].

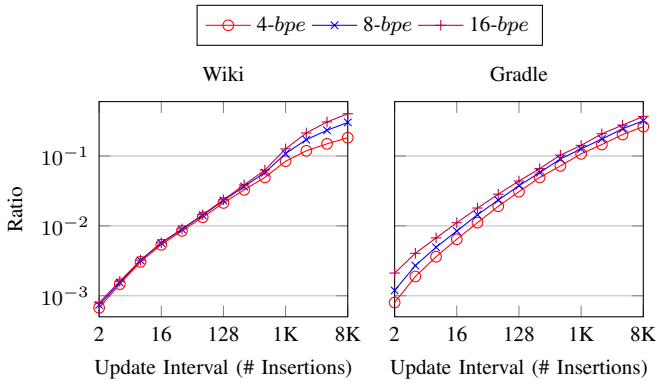
Such approximations commonly introduce the risk of *false-positive* errors, i.e., the indicator sometimes wrongly indicates that a datum is stored in the cache. In such a case, accessing the cache results in an unnecessary cache access, which translates to an excessive cost. Consequently, the problem of advertising space-efficient indicators while keeping a low false-positive rate has attracted a bulk of research effort [7], [9]–[11], [13]. Other works addressed the *cache selection* problem, namely, selecting which cache to access when there exist one (or more) positive indications, where some of them may actually be false-positives [7], [14].

Most previous works [3], [7], [14] assume that there are no *false-negative* indications. Indeed, there exist indicators that theoretically guarantee a false-negative ratio of zero (e.g., simple Bloom filter [9]), or a negligible false-negative ratio (e.g., a Counting Bloom Filter (CBF) [10]). However, to manifest this guarantee in practical distributed environment, every cache should advertise its indicator to all the clients in the network upon every change in the cached content, usually resulting in prohibitive bandwidth consumption. For instance, a leading CDN provider reports using Bloom filters of about 70MB in size in every cache [2]. Insisting on sending an update upon every change in the cached content in such a system may result in having the advertised indicators consume more bandwidth than the cached content itself. Hence, caches commonly advertise their content only periodically.

When using periodical updates, the advertised content gradually becomes *stale*. Namely, it takes time for the indicator available at the clients to reflect changes in the cached content. Unfortunately, such staleness may lead to a significant increase in false-negative indications. To illustrate the problem, consider a cache that advertises a fresh indicator, and later admits a new item  $x$ . When the client tests for  $x$ , the indicator is likely to wrongly indicate that  $x$  is not in the cache (as it wasn’t in the cache when the advertisement was sent), resulting in a false-negative error. Such scenarios are quite common in highly dynamic networks, such as 5G networks [5].

To explore the significance of false-negatives caused due to staleness, consider Fig. 1a. The figure presents the false-negative ratio indications as a function of the time between subsequent advertisements, referred to as the *update interval* (both axes are in logarithmic scale). We measure the update interval by the number of cache changes (insertions of new items). The indicator used is an optimally configured simple Bloom filter [9], where the figure shows distinct indicators with

\* The work was done while this author was with Ben-Gurion University.



(a) Effect of the update interval on the false-negative ratio. Both axes are in log-scale, the cache size is 10K, and policy is LRU. The traces are Wiki and Gradle (described in Sec. VI). The distinct plots for increasing values of bits-per-cached-element (*bpe*), correspond to increasing indicator sizes, with decreasing false-positive ratios, respectively.

	Cache 1	Cache 2	Cache 3
Cost	10	20	1
Indication	$I(x) = 0$	$I(x) = 1$	$I(x) = 0$
True answer	$C(x) = \text{No}$	$C(x) = \text{No}$	$C(x) = \text{Yes}$

(b) The client is looking for item  $x$  and needs to select a cache. Accessing each cache incurs some known access cost. Each cache  $j$  provides an indication w.r.t. item  $x$ , where  $I(x) = 1$  means that the item is likely to be cached, and  $I(x) = 0$  means that it is likely not cached. The true answer is captured by  $C(x)$ , which is “Yes” iff  $x$  is found in the cache. The indications of caches 1, 2 and 3 for item  $x$  are true-negative, false-positive, and false-negative, respectively. Failing to retrieve  $x$  from a cache incurs a miss penalty of 100.

Fig. 1. Motivation for a false-negative aware approach.

varying number of Bits Per cached Element (*bpe*); a higher *bpe* implies a larger indicator, that is guaranteed to provide a lower false-positive ratio [13]. Fig. 1a shows that the false-negative ratio dramatically increases for all indicator sizes. Furthermore, this phenomenon is manifested for various types of workloads, where Fig. 1a shows this for two specific traces, Wiki and Gradle (which represent significantly distinct workloads, as described in Sec. VI-A). For instance, it is not uncommon to have a false-negative ratio as high as 10% when the update interval is above 1K. Most interestingly, using a larger indicator, which guarantees a lower inherent false-positive ratio [13], results in a *higher* false-negative ratio. We discuss and explain this phenomenon with more detail in Sec. VI-C.

Fig. 1b exemplifies the potential benefit of a false-negative aware access strategy. A false-negative oblivious strategy would access only caches with positive indication, i.e., cache 2. However, this indication is a false-positive, thus incurring a miss penalty of 100 for a total cost of 120. In contrast, a false-negative-aware approach may access both caches 2 and 3, implying an access cost of 21 and a hit. Intuitively, as the miss penalty is 100 while the access cost of cache 3 is only 1, it is beneficial to access cache 3 despite a negative indication if the likelihood of a false-negative event is more than 1%.

Designing an access strategy that considers both false-positives and false-negatives is a challenging task. In particular, it is unclear whether, or when, it may be beneficial to access a cache despite a negative indication. Furthermore, it is unclear

even how to estimate the probability of a false-negative event. However, to the best of our knowledge, despite its importance, this problem has never been studied.

We stress that we do not address cache replacement policies, the design of new indicators, or optimizing update intervals. We focus our attention on a false-negative-aware *cache access strategy*, namely, which cache(s) should the client access, being aware to the non-negligible probability of false-negatives, which naturally arises in scenarios where indicators become stale.

#### A. Our Contribution

We consider the problem of accessing a multi-cache system while using indicators that exhibit both false-positive and false-negative indications. We challenge the common practice, which assumes that it is *always* better not to access caches with negative indications. Specifically, we develop a framework that supports *false-negative awareness*, and design policies that actively access caches with negative indications, aiming at minimizing the overall access cost.

After presenting our system model and some preliminaries in Sec. II, we turn to consider Bloom filters in Sec. III, which serve as our main example of indicators. Specifically, we discuss how staleness affects the false-negative and false-positive ratios manifested in such indicators. We then turn to present an algorithm for fully-homogeneous environments, and show that it is optimal in terms of the overall access cost, assuming that the false-negative ratio, the false-positive ratio, and the hit ratio, are all known to the client. These results appear in Sec. IV. Later, in Sec. V, we develop a strategy for realistic environments which are both heterogeneous (i.e., distinct caches have distinct attributes) and dynamic (i.e., the client may not know the precise attributes of each cache). We show how both the cache and the client may cooperate, allowing the client to estimate some of the underlying distributions (which may depend inter alia on the system configuration and the workload being served). We make explicit use of our analysis of Bloom filters, to serve as a concrete example of how to implement our approach. Our suggested false-negative aware (FNA) approach makes deliberate accesses also to caches with negative indications. Furthermore, we show that any approximation guarantee provided by a false-negative oblivious (FNO) access strategy (in our model), can be used by our false-negative aware framework. In particular, we show how to employ known FNO strategies as subroutines, which induce their performance guarantees on our proposed FNA solution.

Finally, in Sec. VI, we present the results of our in-depth simulation study, where we evaluate the performance of our proposed solution in varying system configurations. We show that our FNA strategy implies a significant reduction in access costs in many real-life scenarios, compared to state-of-the-art FNO approaches. Furthermore, our results show that our FNA strategy with minimal resources obtains comparable results to those obtained by FNO strategies that use considerably more resources. For instance, our results indicate that to match the performance of our FNA strategy, an FNO approach might require as much as an order of magnitude more resources (e.g.,

in terms of system caching capacity or the bandwidth required for indicator advertisement).

### B. Related Work

As previously described, indicators are commonly used to periodically advertise the content of caches in an efficient manner. Since indicators are of bounded size, they usually fail to provide a precise representation of the cache content and exhibit false-positive indications [9], [13]. The pioneering work of [7] shows that due to these false-positives, sometimes naively relying on an indicator for accessing even a single cache may do worse than not using an indicator at all. Subsequent works [14], [15] tackle a distributed scenario where multiple caches advertise indicators, and develop access strategies that take into account both the access cost, and the false-positive rate in each cache, to minimize the overall expected cost. However, [7], [14], [15] disregard *false-negative* indications.

The work of [16] studies the problem of false-negatives in practical deployment of Counting Bloom Filters [10]. Other techniques to reduce the false-negative ratio in numerous variants of Bloom filters are surveyed in [11]. These works address false-negatives that stem from *architectural design*, i.e., from concrete data structures used to implement indicators. Consequently, these works focus on developing enhanced data structures that reduce such false-negatives. In contrast, we focus on false-negatives caused by staleness, i.e., false-negatives that follow from the *operational usage* of the system. Such false-negatives may occur in *any* indicator, even if its design is false-negative-free, such as a simple Bloom filter [9], [13]. In this sense our approach is orthogonal to previous work targeting the reduction of false-negatives [11], [16], and these approaches may be seamlessly combined with our solutions.

Since constantly advertising a fresh indicator might be prohibitively costly, in practice caches commonly advertise fresh indicators only periodically [1], [17]–[19], where one usually refers to the period between the advertisements of fresh indicators as the *update interval*. Several works addressed the interplay between the update interval and performance by means of simulations [1], [18]. The work [19] presented an algorithm that dynamically scales the update interval and the indicator size, to comply with a budget constraint. The work [20] reduces the transmission overheads by accurately advertising important information, while allowing less important information to be stale, or less accurate. The work [17] analyzed the impact of stale Bloom filter replicas on the false-positive ratio and the false-negative ratio. However, the framework of [17] implicitly assumes that requests are drawn from a uniform distribution, and that each object is stored in a single cache. This framework conforms primarily with distributed storage systems. However, in many (if not most) real-life distributed caching environments, requests need not be drawn from a uniform distribution, and furthermore objects may be found in either a single cache, multiple caches, or no cache at all [3], [4], [6]. Part of our analysis of such general environments is inspired by ideas introduced in [17] (see, e.g., Sec. V-A1).

TABLE I  
LIST OF SYMBOLS. THE TOP PART CORRESPONDS TO OUR SYSTEM MODEL (SEC. II), THE MIDDLE PART CORRESPONDS TO THE STRUCTURE OF BLOOM FILTERS (SEC. III), AND THE BOTTOM PART CORRESPONDS TO THE FULLY-HOMOGENEOUS CASE (SEC. IV), AND THE HETEROGENEOUS SETTINGS (SEC. V-VI)

Symbol	Meaning
$N$	Set of caches.
$n$	Number of caches: $n =  N $
$x(x_t)$	Item request (issued at time $t$ )
$N_x$	Set of caches with positive indications for requested item $x$
$n_x$	Number of positive indications for requested item $x$
$S_j$	The set of data items in cache $j$
$h_j$	Hit ratio of cache $j$ : $h_j = \Pr(x \in S_j)$
$I_j$	Indicator of cache $j$
$I_j(x)$	Indication of indicator $I_j$ for item $x$
$FP_j$	False-positive ratio for $I_j$
$FN_j$	False-negative ratio for $I_j$
$\pi_j$	Probability of a miss in cache $j$ given a positive indication
$\nu_j$	Probability of a miss in cache $j$ given a negative indication
$q_j$	Probability of a positive indication in indicator $I_j$
$c_j$	Access cost of cache $j$
$M$	Miss penalty
$\phi$	Cost function. See Eq. (4).
$k$	Number of hash functions in the Bloom filter
$C_j$	Size of cache $j$ ( $ S_j  \leq C_j$ )
bpe	Bits per cached element in the Bloom filter
$B_1(t)$	Number of '1' bits in the updated Bloom filter at time $t$
$B_0(t)$	Number of '0' bits in the updated Bloom filter at time $t$
$\Delta_1(t)$	Number of bits that are '1' in the updated Bloom filter, but '0' in the stale Bloom filter at time $t$
$\Delta_0(t)$	Number of bits that are '0' in the updated Bloom filter, but '1' in the stale Bloom filter at time $t$
$r_0$	Number of caches with negative indication accessed
$r_1$	Number of caches with positive indication accessed
$\hat{\phi}$	Cost function for the fully-homogeneous case. See Eq. (7).
$r_0^*$	Optimal choice of $r_0$ . See Eq. (8).
$r_1^*$	Optimal choice of $r_1$ . See Eq. (8).
$\rho_j$	Probability of a miss in cache $j$ given its indication
$\delta$	Smoothness parameter of moving average. See Eq. (9).

The problem of stale indicators relates to other problems of decision making under uncertainty. In particular, our problem is closely related to the concept of the *Age of Information* (AoI). The AoI quantifies the time since the generation of the last successfully received information from a remote system. The AoI paradigm was applied to numerous environments, e.g. vehicular networks, scheduling, and buffer management; a detailed survey can be found in [21]. The AoI was applied also to caches [22], but in the context of the coherency of the cached data, while we focus on the coherency of the indicators.

## II. SYSTEM MODEL AND PRELIMINARIES

This section formally defines our system model and notations, which are summarized in Table I. We consider a set  $N$  of  $n = |N|$  caches, containing possibly overlapping sets of items. We associate each time  $t$  with a unique item request  $x_t$  issued at time  $t$ , and we refer to the entire sequence of requests as  $\sigma$ . Let  $S_{j,t}$  denote the set of items stored at cache  $j$  at time  $t$ , prior to handling request  $x_t$ . For every request  $x_t$  issued at time  $t$ , drawn from some distribution, we let  $h_{j,t}$  denote the probability that  $x_t \in S_{j,t}$ . This probability depends both on the distribution of the requests, as well as on the cache policy. The

average  $h_{j,t}$  over the entire sequence is commonly referred to as the *hit ratio*, i.e., the fraction of requests in  $\sigma$  that were available in cache  $j$ , upon being issued. Similarly to previous works, we assume that the past hit ratio is a reasonable estimate of  $h_{j,t}$  [23], [24]. We refer to this estimation as the probability that the next accessed item  $x_t$  is available in  $S_{j,t}$ .<sup>1</sup>

Each cache  $j$  maintains an *indicator*  $I_{j,t}$ , which approximates the set of items in cache  $j$  at time  $t$ ; given an item  $x$ ,  $I_{j,t}(x) = 1$  is referred to as a *positive indication* while  $I_{j,t}(x) = 0$  is considered a *negative indication*.

The *false-positive ratio* of  $I_{j,t}$  is defined by  $FP_{j,t} = \Pr(I_{j,t}(x) = 1 | x \notin S_{j,t})$ . It captures the probability that given a request  $x$  issued at time  $t$ , the indicator would mistakenly indicate that it is in  $S_{j,t}$ . Similarly, the *false-negative ratio* of  $I_{j,t}$ , defined by  $FN_{j,t} = \Pr(I_{j,t}(x) = 0 | x \in S_{j,t})$ , is the probability that the indicator mistakenly indicates that a request for  $x$ , issued at time  $t$  is not in  $S_{j,t}$ .

For every cache  $j$ , and every time  $t$ , we denote by  $\pi_{j,t} = \Pr(x \notin S_{j,t} | I_{j,t}(x) = 1)$  the *positive exclusion probability*, that is, the probability that a requested item  $x$  is not in the cache, despite a positive indication. Similarly, we let  $\nu_{j,t} = \Pr(x \notin S_{j,t} | I_{j,t}(x) = 0)$  denote the *negative exclusion probability*, that is, the probability that a requested item  $x$  is not in the cache, given a negative indication. We denote by  $q_{j,t}$  the probability of a positive indication for an item requested from cache  $j$ , and refer to  $q_{j,t}$  as the *positive indication ratio*. When clear from the context, we abuse notation and omit the time  $t$ .

Positive indications occur when either  $x \in S_j$  and no false-negative occurs; or  $x \notin S_j$ , and a false-positive occurs. Hence,

$$q_j = \Pr(I_j(x) = 1) = h_j \cdot (1 - FN_j) + (1 - h_j) \cdot FP_j. \quad (1)$$

Using Bayes' theorem it follows that

$$\pi_j = \Pr(x \notin S_j | I_j(x) = 1) = FP_j \cdot (1 - h_j) / q_j \quad (2)$$

$$\begin{aligned} \nu_j &= \Pr(x \notin S_j | I_j(x) = 0) \\ &= (1 - FP_j) \cdot (1 - h_j) / (1 - q_j), \end{aligned} \quad (3)$$

for  $q_j$  as defined in Eq. (1).

We say that a system is *sufficiently-accurate* if for every indicator of cache  $j$ ,  $FP_j + FN_j < 1$ .<sup>2</sup> We note that in most real-life scenarios, both the false-positive ratio  $FP_j$ , and the false-negative ratio  $FN_j$ , are well below 0.5, and therefore such systems are sufficiently-accurate.

The following simple condition characterizes sufficiently-accurate systems. Due to space constraints, some of the proofs are omitted, and available in [26].

**Proposition 1.** *A system is sufficiently-accurate iff for every  $j$  it holds that  $\nu_j > \pi_j$ .*

The proof of Prop. 1 follows from algebraic manipulation. Intuitively, the condition  $\nu_j > \pi_j$  implies that an item is more likely to be in the cache given a positive indication than its likelihood of being in the cache given a negative indication. This condition states that indications (either positive or negative) are more likely to be correct than incorrect.

<sup>1</sup>We provide further details of how to obtain such an estimation in Sec. V-A2.

<sup>2</sup>This definition is inspired by the notions of *accuracy* and *informedness* [25].

For any query  $x$ , let  $N_x$  denote the set of caches with positive indications, i.e.,  $N_x = \{j | I_j(x) = 1\}$ , and  $n_x = |N_x|$ .

A request for datum  $x$  triggers a *data access* which consists of (i) querying for  $x$  in all the  $n$  indicators, (ii) selecting a subset  $D \subseteq N$  of caches, and (iii) accessing all the  $|D|$  selected caches in parallel. Accessing each cache incurs some predefined *access cost*,  $c_j$ . For ease of presentation, we assume without loss of generality that  $\min_j c_j = 1$ . The overall access costs of accessing a set  $D$  of caches is  $c_D = \sum_{j \in D} c_j$ .

A multi-cache data access is considered a *hit* if the item  $x$  is found in at least one of the accessed caches, and a *miss* otherwise. A miss incurs a *miss penalty* of  $M$ , for some  $M \geq 1$ .

In our model, we do not assume any specific sharing policy among the caches. Yet, in the analysis of our system (Sections IV-V) we assume that the exclusion probabilities are mutually independent. Under this assumption, our analysis provides a baseline for understanding the performance of such systems. However, in the evaluation of our algorithms, we consider environments where the exclusion probabilities are not necessarily mutually independent (Sec. VI).

For a subset of caches  $D$ , we define its (expected) *miss cost* for a query  $x$  by  $M \cdot \prod_{I_j(x)=1}^{j \in D} \pi_j \cdot \prod_{I_j(x)=0}^{j \in D} \nu_j$ .

The (expected) *service cost* of a query is the sum of the access cost and the miss cost, namely,

$$\phi_x(D) = \sum_{j \in D} c_j + M \prod_{I_j(x)=1}^{j \in D} \pi_j \cdot \prod_{I_j(x)=0}^{j \in D} \nu_j. \quad (4)$$

The *Cache Selection (CS) problem* is to find a subset of caches  $D \subseteq N$  that minimizes the expected cost  $\phi_x(D)$ .<sup>3</sup>

In what follows we refer to an access to a cache with a positive indication as a *positive access*, and refer to an access to a cache with a negative indication as a *negative access*. In particular, we consider two types of approaches to solving the CS problem: (i) *false-negative oblivious (FNO)* schemes, which only perform positive accesses, and (ii) *false-negative aware (FNA)* schemes, which may also perform negative accesses. While the former may be viewed as the traditional manner in which access strategies are designed, the latter is a more speculative approach, which sometimes accesses a cache even with no positive indication, risking an increased access cost.

### III. BLOOM FILTERS AND STALENESS

In this section, we consider Bloom filters [9], which we use as a primary example of indicators. In particular, we will discuss how staleness affects false-negatives and false-positives in such indicators. We note that our approach described hereafter can also be applied to other types of indicators (e.g., TinyTable [12]).<sup>4</sup>

A Bloom filter is a randomized data structure that approximately represents a set of items. A Bloom filter  $I$  consists of a bit array of size  $|I|$ , as well as  $k$  *independent* hash functions. When adding an item to the filter, each of the  $k$  hash functions

<sup>3</sup>When clear from the context, we will omit the subscript  $x$  from  $\phi_x$ .

<sup>4</sup>In general, the false-negative and false-positive ratio are strongly related to the structure of the indicator, and at times even to its specific implementation.

	$B_0(t)$		$B_1(t)$	
updated BF	0...0	0.....0	1...1	1...1
stale BF	1...1	0.....0	0...0	1...1
	$\Delta_0(t)$	$B_0(t) - \Delta_0(t)$	$\Delta_1(t)$	$B_1(t) - \Delta_1(t)$

Fig. 2. An example of an updated and a stale Bloom filter at time  $t$ .

is applied to the item, and the corresponding bit in the array is set. When testing for an item's existence, we apply the  $k$  hash functions and test the corresponding bits. If all bits are set, the Bloom filter replies with a positive indication. Otherwise, the indication is negative.

In a fresh Bloom filter, which is updated upon every insertion of an item to the set, positive indications may be false due to hash collisions, but negative indications are guaranteed to be correct. However, in a stale Bloom filter, negative indications may also be erroneous. Such false-negatives occur, e.g., when indicators are advertised to the clients only periodically. In such a scenario, when a new item is admitted, but the updated indicator is not yet advertised, the stale indicator available to the client fails to represent this change.

To allow a meaningful analysis of the trade-off between accuracy and memory footprint, it is useful to express the size of Bloom filters using the notion of *Bits Per Element (bpe)*. Intuitively, optimally configured Bloom filters of the same bpe have the same false-positive accuracy, regardless of the size of the set being approximated by the Bloom filter. More formally, given the value of bpe, one can calculate the optimal number of hash functions  $k$ , minimizing the false-positive ratio [13].

In the context of caches, let  $C_j$  denote the *size* of cache  $j$ , i.e., the maximum number of elements that can be stored in cache  $j$ . The actual size of the bloom filter indicator  $I_j$  associated with cache  $j$  is therefore  $|I_j| = \text{bpe} \cdot C_j$ . Each cache manages its own Bloom filter, and occasionally advertises the indicator to the client. At any time  $t$ , we consider the *updated* Bloom filter maintained by the cache, and let  $B_1(t)$  and  $B_0(t)$  denote the number of bits set (i.e., with value 1) and reset (i.e., with value 0), at time  $t$ , respectively, in the Bloom filter approximating the content of the cache. A client uses a replica of the cache's Bloom filter, representing a snapshot of the Bloom filter that the cache advertised at some time  $t' \leq t$ . We refer to this replica as the *stale* Bloom filter. Let  $\Delta_1(t)$  denote the number of bits that are set in the updated Bloom filter but are reset in the stale Bloom filter. Similarly, we let  $\Delta_0(t)$  denote the number of bits that are reset in the updated Bloom filter, but are set in the stale Bloom filter. Fig. 2 illustrates this situation. For clarity of presentation, and WLOG, we group together all the bits contributing to  $\Delta_0(t)$ ; and group together the bits accounted by  $\Delta_1(t)$ .

*The false-negative ratio:* Consider a query for an item  $x$  that is stored in the cache at time  $t$ . Recall that an updated Bloom filter never exhibits false-negatives. Hence we know that all the  $k$  hashes of  $x$  are mapped to the bits that are set in the updated Bloom filter. The query for  $x$  in a stale indicator is a true positive iff all the hashes are mapped to the set of  $B_1(t) - \Delta_1(t)$  bits that are also set in the stale indicator; by the fact that the  $k$  hash functions are independent,

and uniformly distributed over their range, this happens with probability  $\left[\frac{B_1(t) - \Delta_1(t)}{B_1(t)}\right]^k$ . Otherwise, the query for  $x$  is a false-negative. It follows that the false-negative ratio of the cache at time  $t$  can be estimated by

$$\text{FN}_t = 1 - \left[\frac{B_1(t) - \Delta_1(t)}{B_1(t)}\right]^k. \quad (5)$$

*The false-positive ratio:* Consider a query for an item  $y$  that is not stored in the cache at time  $t$ . For uniformly distributed and independent hash functions, the hashes of  $y$  are mapped to arbitrary locations in the Bloom filter. The stale Bloom filter exhibits a false-positive iff all the  $k$  hashes of  $y$  map to bits that are set in the stale Bloom filter. Hence, the probability of a false-positive in the stale indicator at time  $t$  can be estimated by:

$$\text{FP}_t = \left[\frac{B_1(t) - \Delta_1(t) + \Delta_0(t)}{|I_j|}\right]^k. \quad (6)$$

We note that Eqs. (5) and (6) are only estimations, while the exact miss probabilities strongly depend upon the workload, and the cache policy. For instance, consider the case where immediately after cache  $j$  sends an update, it caches an item  $x$ . Then, any subsequent requests for  $x$ , until the cache advertises the next update, are false-negatives. However, until the cache advertises the next updated indicator,  $x$  may be accessed many times, or not accessed at all, according to the concrete workload.

To conclude this section, we note that the above analysis may be applied to more complex filters. One such example, which we use in Sec. VI, is a compressed *Counting Bloom Filter (CBF)* [10], which behaves similarly to a simple BF, while supporting also the removal of objects.

#### IV. THE FULLY HOMOGENEOUS CASE

In this section, we focus on a simplified fully-homogeneous case. In such settings, the access cost of all caches is the same, and is normalized to one ( $c = 1$ ). The per-cache hit ratio, false-positive ratio, and false-negative ratio, are identical for all caches. I.e., for each cache  $j$ ,  $h_j = h$ ,  $\text{FP}_j = \text{FP}$  and  $\text{FN}_j = \text{FN}$  for some constants  $h, \text{FP}, \text{FN} \in [0, 1]$ .

We use this homogeneous setting to explore the challenges and potential benefits arising from developing a false-negative aware cache selection strategy. We first describe the aspects specific to such homogeneous settings, and then describe and analyze our false-negative aware *Homogeneous Cache Selection* policy,  $\text{HoCS}_{\text{FNA}}$ . Our analysis shows that  $\text{HoCS}_{\text{FNA}}$  minimizes the service cost in the fully-homogeneous case. Later on, we use  $\text{HoCS}_{\text{FNA}}$  to derive insights as to when it is beneficial to access a cache despite a negative indication.

**Preliminaries:** In fully-homogeneous settings, the task of selecting a subset of the caches  $D \subseteq N$  that minimizes the service cost is reduced to selecting two integers:  $0 \leq r_1 \leq n_x$ , the number of caches with positive indication to access; and  $0 \leq r_0 \leq n - n_x$ , the number of caches with negative indication to access. The objective function  $\phi$  (Eq. (4)) is reduced to

$$\hat{\phi}(r_0, r_1) = r_0 + r_1 + M \cdot \nu^{r_0} \cdot \pi^{r_1}. \quad (7)$$

---

**Algorithm 1** HoCS<sub>FNA</sub>

---

- 1:  $r_0^* = 0; r_1^* = \arg \min_{r_1 \in [0, n_x]} [r_1 + M \cdot \pi^{r_1}]$
  - 2: **if**  $M \cdot \pi^{r_1^*} > 1$  **then**
  - 3:      $r_0^* = \arg \min_{r_0 \in [0, n - n_x]} [r_0 + M \cdot \pi^{r_1^*} \nu^{r_0}]$
  - 4: **return**  $r_0^*, r_1^*$
- 

We let  $r_0^*$  and  $r_1^*$  denote the values of  $r_0$  and  $r_1$  that minimize the service cost, namely

$$\hat{\phi}(r_0^*, r_1^*) = \min_{\substack{0 \leq r_0 \leq n_x, \\ 0 \leq r_1 \leq n - n_x}} \hat{\phi}(r_0, r_1) \quad (8)$$

In our analysis of  $\hat{\phi}$ , we use the extension of  $\hat{\phi}$  over the reals, which we hereafter denote by  $\tilde{\phi}$ . Observe that for any fixed constants  $a, b$ , the functions  $\tilde{\phi}(a, r_1)$ , and  $\tilde{\phi}(r_0, b)$  are strictly convex. Throughout our analysis we will be using the following proposition.

**Proposition 2.** *Let  $\tilde{f} : \mathbb{R} \rightarrow \mathbb{R}$  be a strictly convex function, and let  $\hat{f}$  be its restriction over the integers. If  $\hat{f}$  obtains its minimum value at some integer  $\hat{a} \geq 1$ , then  $\tilde{f}(1) < \hat{f}(0)$ .*

Intuitively, Prop. 2 stems from the fact that a strictly convex function obtains its minimum in a single point. When applied to integers, we can determine whether it is beneficial to access some  $\hat{a} \geq 1$  caches by only checking whether  $\tilde{f}(1) < \hat{f}(0)$ .

We recall that the function  $\tilde{\phi}(r_0, r_1^*)$  is strictly convex. By applying Prop. 2, we therefore obtain the following corollary.

**Corollary 3.** *If  $r_0^* \geq 1$ , then  $\hat{\phi}(1, r_1^*) < \hat{\phi}(0, r_1^*)$ .*

**An Optimal Strategy:** Our algorithm for the fully-homogeneous settings, HoCS<sub>FNA</sub>, is formally defined in Algorithm 1. The algorithm first calculates the number of caches with positive indication to access,  $r_1^*$ , assuming no cache with negative indication is accessed (line 1). Next, if the expected miss cost is still higher than accessing an additional cache (the condition in line 2), the algorithm also considers caches with negative indications (line 3). The following theorem shows that HoCS<sub>FNA</sub> is optimal in the fully-homogeneous case.

**Theorem 4.** *If a fully-homogeneous system is sufficiently-accurate, then HoCS<sub>FNA</sub> minimizes the service cost.*

*Proof.* By Prop. 1, we have  $\nu > \pi$ . In addition, of  $\hat{\phi}$  (Eq. (7)), the objective function  $\hat{\phi}(r_0, r_1)$  is symmetric in  $r_0, r_1$ . It follows that assigning  $r_0 > 0$  may reduce cost only if  $r_1$  is maximized, namely, if  $r_1 = n_x$ . Hence, line 1 indeed calculates the optimal value of  $r_1^*$ .

If  $M \cdot \pi^{r_1^*} > 1$ , then the algorithm sets  $r_0^*$  to a value which minimizes the total cost (line 3). Else (namely, if  $M \cdot \pi^{r_1^*} \leq 1$ ), accessing cache(s) with negative indications can only increase the total cost, as it increases the aggregate access cost by at least 1, which is at least as high as the potential marginal decrease in the miss cost, which is at most  $M \cdot \pi^{r_1^*} \leq 1$ . Hence, when the if-condition (line 2) is not met, it is best to keep the default value  $r_0^* = 0$ .  $\square$

The following proposition characterizes cases where it is better to access a cache in spite of a negative indication.

**Proposition 5.** *Assume  $\nu, \pi \in (0, 1)$ . (i) If  $n_x = 0$ , then accessing at least one cache with negative indication strictly reduces the service cost iff  $\nu < 1 - \frac{1}{M}$ . (ii) If  $n_x > 0$ , then accessing at least one cache with negative indication strictly reduces the service cost iff  $\nu < 1 - \frac{1}{M\pi^{n_x}}$  and  $M \cdot \pi^{n_x - 1}(1 - \pi) > 1$ .*

*Proof sketch.* Note that  $\nu$  is a proxy to the true negative rate. Hence, the conditions in cases (i) and (ii) reflect the fact that if the expected miss cost after accessing all the caches with a positive indication is high, and the true-negative rate is low, then it might be beneficial to access a cache with a negative indication. The proof follows by using Theorem 4, Prop. 2, and Cor. 3, while considering the properties of the convex function  $\tilde{\phi}(0, r_1)$ , and applying some algebraic manipulation.  $\square$

The following proposition characterizes cases where it is better to access no cache, even if there exist positive indications.

**Proposition 6.** *If positive indications exist but  $(1 - h)FP \geq h(1 - FN)(M - 1)$ , then the best policy is to access no cache.*

*Proof sketch.* Observe that in a system where  $(1 - h)FP$  is large, a positive indication is very likely to be false-positive. Hence, it may be beneficial not to access a cache despite a positive indication. This is true especially if the miss penalty is small, as reflected by the right-hand-side of the condition. The proof follows from combining Theorem 4 and Prop. 2, and applying some algebraic manipulation.  $\square$

## V. DYNAMIC AND HETEROGENEOUS SETTINGS

In Sec. IV we assumed that all caches have the same false positive/negative ratios and hit ratios. In this section, we consider more realistic settings where cache attributes are *dynamic*, implying that they might not be accurately available at the client, and *heterogeneous*, i.e., caches may have distinct attributes. Such an extension poses two main challenges for solving the CS problem. The first challenge is dealing with incomplete information, which requires estimating the exclusion probabilities  $\pi_j$  and  $\nu_j$ . This is presented in Sec. V-A. The second challenge is dealing with heterogeneity, which calls for efficient algorithms for choosing the set of caches to access. This is presented in Sec. V-B.

The algorithm executed by the client, which we dub CS<sub>FNA</sub>, is formally defined in Algorithm 2. In a nutshell, the algorithm periodically obtains information from the caches (lines 1-2), and then uses this information in order to estimate the exclusion probabilities for each cache (lines 3-10). Using these values, CS<sub>FNA</sub> runs a cache selection algorithm (e.g., one of the algorithms in [14]). In what follows, we provide a detailed description of our algorithm's inner workings and analysis.

### A. Estimating the Exclusion Probabilities

We now show how one can collect recent statistics of the various parameters governing system behavior, which allow the client to obtain good estimates of the current caches'

---

**Algorithm 2**  $CS_{\text{FNA}}(N, \vec{c}, M, \text{Alg})$ 

---

```
1: periodically obtain updated  $FP_j, FN_j$  from each cache  $j$ 
2: periodically obtain updated indicator  $I_j$  from each cache  $j$ 
3: estimate  $q_j$  for each cache  $j$ 
4: for every request for datum  $x$  do
5:   for  $j = 1, \dots, n$  do
6:      $h_j = \frac{q_j - FN_j}{1 - FP_j - FN_j}$   $\triangleright$  Eq. (1)
7:     if  $I_j(x) == 1$  then  $\triangleright$  should calculate  $\pi_j$ 
8:        $\rho_j = FP_j \cdot (1 - h_j) / q_j$   $\triangleright$  Eq. (2)
9:     else  $\triangleright$  should calculate  $\nu_j$ 
10:       $\rho_j = (1 - FP_j) \cdot (1 - h_j) / (1 - q_j)$   $\triangleright$  Eq. (3)
11:    $D = \text{Alg}(N, \vec{c}, \vec{\rho}, M)$   $\triangleright$  reduction of Theorem 7
12:   access  $D$ 
```

---

attributes. Our solutions use the insights presented in Sec. III. In particular, we will use Eqs. (5) and (6) for estimating the false-positive ratios and the false-negative ratios of the distinct caches, enabling us to compute  $\pi_j$ , and  $\nu_j$ .

1) *Cache-side Algorithm:* The cache maintains both the stale Bloom filter (i.e., the most recently advertised Bloom filter, which is also available at the client) and the updated Bloom filter. Along a sequence of requests  $\sigma$ , each cache  $j$  estimates the false-negative ratio and the false-positive ratio, according to Eqs. (5) and (6), by comparing the stale and updated Bloom filters. We note that these estimations can be done periodically to reduce the computational overhead of comparing the stale and updated bloom filters. These estimates are sent (periodically) to the client (line 1 of  $CS_{\text{FNA}}$ ). Each cache further (periodically) sends an updated indicator to the client (line 2 of  $CS_{\text{FNA}}$ ). We note that these updates are sent in an arbitrary asynchronous manner to the client.

2) *Client-side Algorithm:* We now show how the client may estimate the exclusion probabilities  $\pi_j$  and  $\nu_j$  for every cache  $j$ , given the estimations of  $FP_j$  and  $FN_j$  which are periodically provided by the cache.

For evaluating  $q_j$ , the client periodically estimates the probability  $\Pr(I_j(x) = 1)$  *empirically*, using a weighted exponential moving average. Formally, consider a sequence of requests  $\sigma$ , and consider epochs of  $T$  requests. Let  $a_j(s, t)$  denote the number of positive indications of indicator  $I_j$  for requests  $s + 1, \dots, t$  made by the client. For any  $t \leq T$  we let the estimated positive indication ratio after handling request  $t$  be  $q_{j,t} = \frac{a_j(0,t)}{t}$ . For every  $i = 1, 2, \dots$  and every  $iT < t < (i+1)T$ , we let  $q_{j,t}$  be the most recent estimate over epochs of  $T$  requests, i.e.,  $q_{j,t} = q_{j, \lfloor t/T \rfloor \cdot T}$ , and the estimate is updated at  $t = (i+1)T$  such that

$$q_{j,(i+1)T} = \delta \cdot \frac{a_j(iT, (i+1)T)}{T} + (1 - \delta) \cdot q_{j,iT}, \quad (9)$$

where  $\delta \in (0, 1)$  is some constant governing the dynamics of the estimate change. We note that only the client can perform such an estimation since it requires knowing all the requests in  $\sigma$ , and not only requests for which the cache has been accessed.

Given the current values for  $FP_j$ ,  $FN_j$ , and  $q_j$ , for every item being requested in the sequence  $\sigma$ , the client estimates the

hit ratio  $h_j$  (line 6), and the exclusion probabilities  $\pi_j$  (line 8) and  $\nu_j$  (line 10) using Eqs. (1), (2), and (3), respectively.

### B. Choosing the Caches to Access

This section shows how to use the estimations of the exclusion probability of each cache  $j$  to develop a false-negative-aware access strategy. In particular, we show how to extend any false-negative-oblivious access strategy (e.g., those in [14]), to consider a non-zero false-negative ratio.

For any set of caches  $D$ , the client's estimations of the exclusion probabilities essentially determine the expected miss cost. We let  $\rho_j$  denote the probability of a miss while accessing cache  $j$ , given its indication for the requested item. Formally,  $\rho_j = \pi_j$  if  $I_j(x) = 1$ , and  $\rho_j = \nu_j$  if  $I_j(x) = 0$ . Then, the expected miss cost can be expressed by  $M \cdot \prod_{j \in D} \rho_j$ , and the objective function defined in Eq. (4) translates to finding the set of caches  $D$  minimizing

$$\phi_x(D) = \sum_{j \in D} c_j + M \cdot \prod_{j \in D} \rho_j. \quad (10)$$

The problem of finding a set of caches  $D$  (out of those with a positive indication) minimizing an objective of the form depicted in Eq. (10) has been studied in [14], where they present several approximation algorithms for the problem. The problem studied in [14] is essentially equivalent to assuming that there are no false-negative indications, and therefore it sufficed to consider only caches for which  $I_j(x) = 1$ . We refer to this special case as the *restricted CS problem*.

When considering the restricted CS problem within our model, the framework of [14] can be viewed as assuming that *all* caches have a positive indication, and  $\rho_j$  represents the positive exclusion probability of cache  $j$ . Equivalently, the model of [14] essentially assumed that  $\nu_j = 1$  for all  $j$ , which is fundamentally not the case in the CS problem.

Our proposed algorithm  $CS_{\text{FNA}}$  selects the set of caches to access as follows: (i)  $CS_{\text{FNA}}$  gets as input an algorithm  $\text{Alg}$  for solving the restricted CS problem (assuming all caches have a positive indication), (ii) generates the appropriate input for this algorithm (as described above) in lines 7-10, and (iii) accesses the set of caches prescribed by algorithm  $\text{Alg}$ .

The following theorem serves to analyze the worst-case performance guarantees of  $CS_{\text{FNA}}$ .

**Theorem 7.** *If there exists an algorithm  $\text{Alg}$  that is an  $\alpha$ -approximation algorithm for the restricted CS problem, then there exists an  $\alpha$ -approximation algorithm for the general CS problem (with arbitrary values of  $\nu_j$ ).*

*Proof sketch.* Assume an input to the CS problem such that every cache  $j$  has its indicator  $I_j$ , and its positive and negative exclusion probabilities  $\pi_j$  and  $\nu_j$ , respectively. In what follows we abuse notation, and refer to  $\phi_{x, \vec{\pi}, \vec{\nu}, \vec{I}}$  as the expected service cost for an input  $x$ , given these system parameters.

Let  $\text{Alg}$  be an  $\alpha$ -approximation algorithm for the restricted CS problem. Assume each cache  $j$  has some arbitrary negative exclusion probability,  $\nu_j$ . For every cache  $j$ , we let  $\pi_j^* = \pi_j$  if  $I_j(x) = 1$ , and let  $\pi_j^* = \nu_j$  if  $I_j(x) = 0$ . Furthermore, we let  $\nu_j^* = 1$  for all  $j$ . Finally, for every cache  $j$  we define indicator

$I_j^*$  such that  $I_j^*(x) = 1$ , implying that the set of caches with a positive indication according to  $\vec{I}^*$  is the set of all caches,  $N$ .

We define algorithm  $\text{Alg}^*$  such that  $\text{Alg}^*$  returns the output of  $\text{Alg}$  for the inputs of  $\vec{\pi}^*$  (for the positive exclusion probabilities),  $\vec{\nu}^*$  (for the negative exclusion probabilities), and the set of all caches with a positive indication according to  $\vec{I}^*$  (i.e.,  $N$ ). We now show that the solution returned by  $\text{Alg}^*$  is an  $\alpha$ -approximate solution for the CS problem with negative exclusion probabilities  $\nu_j$ .

By the assumption on  $\text{Alg}$ , its output  $D$  satisfies

$$\phi_{x, \vec{\pi}^*, \vec{\nu}^*, \vec{I}^*}(D) \leq \alpha \cdot \phi_{x, \vec{\pi}^*, \vec{\nu}^*, \vec{I}^*}(D^*) \quad (11)$$

where  $D^*$  is the optimal solution to the CS problem with  $\vec{\pi}^*$ ,  $\vec{\nu}^*$ , and the set of caches induced by  $\vec{I}^*$  as inputs. Since by the definition of  $\vec{\pi}^*$ ,  $\vec{\nu}^*$ , and  $\vec{I}^*$  it follows that

$$\phi_{x, \vec{\pi}^*, \vec{\nu}^*, \vec{I}^*}(\tilde{D}) = \phi_{x, \vec{\pi}, \vec{\nu}, \vec{I}}(\tilde{D}) \quad (12)$$

for every set of caches  $\tilde{D}$ , we are guaranteed to have that  $D^*$  is also the optimal solution to the CS problem with  $\vec{\pi}$ ,  $\vec{\nu}$ , and  $\vec{I}$  as inputs. The theorem follows.  $\square$

The proof of Theorem 7 implies the following corollary.

**Corollary 8.** *If the estimations of  $\pi_j$  and  $\nu_j$  produced by  $\text{CS}_{\text{FNA}}$  are precise, and  $\text{Alg}$  used by  $\text{CS}_{\text{FNA}}$  is an  $\alpha$ -approximation algorithm for the restricted CS problem, then  $\text{CS}_{\text{FNA}}$  produces an  $\alpha$ -approximate solution to the CS problem.*

Combining Corollary 8 with the results of [14], we obtain a myriad of tradeoffs and possible approximation guarantees for  $\text{CS}_{\text{FNA}}$ . In particular, in Sec. VI we consider the performance of one specific realization of  $\text{CS}_{\text{FNA}}$ , which uses algorithm  $\text{DS}_{\text{PGM}}$  for the restricted CS problem presented in [14].

## VI. SIMULATION STUDY

In this section, we evaluate the performance and trade-offs of our proposed false-negative aware algorithm,  $\text{CS}_{\text{FNA}}$ , in a variety of scenarios, using traces of real-life workloads. Our evaluation shows that false-negative awareness improves the oblivious approach across the board. In some cases, one needs an order of magnitude more bandwidth or more cache entries to match our false-negative aware approach's service cost. The effect is consistent for diverse cache sizes, workloads, and when increasing the number of caches. The difference is especially significant when the miss penalty and update interval are large. We begin by describing our evaluation settings and parameters.

### A. Simulation Settings

*Traces:* We use the first 1M requests from each of these real workload traces. (i) *Wiki:* Read requests to Wikipedia pages [27]. (ii) *Gradle:* Gradle is a build tool for caching compiled libraries in large projects. The traces was provided by [28]. (iii) *Scarab:* A trace from Scarab Research, a personalized recommendation system for e-commerce sites [28]. (iv) *F2:* Traces from a financial transaction processing system [29].

*Caches:* We consider a system-wide request distribution where a missed item is placed in a single cache that is chosen by the controller. Such an approach is common in large distributed systems, such as Memcached [30], and Kademlia [31] for load balancing, and for maximizing the cached content.

Each cache applies the *Least Recently Used (LRU)* eviction policy which is arguably the most commonly used policy.

*Indicators:* Each cache  $j$  of size  $C_j$  periodically advertises an indicator  $I_j$  of size  $\text{bpe} \cdot C_j$ . For computing the indicator, cache  $j$  maintains a *Counting Bloom Filter (CBF)* [10] with 3-bit counters, where the number of counters is  $\text{bpe} \cdot C_j$ . The advantage of the CBF over a simple Bloom filter [9] is that the CBF supports also removal of items. Thus, we add an item to the CBF upon admission to the cache and remove an item from the CBF upon eviction. The cache constructs the advertised indicator by compressing the CBF to a simple (1 bit-counter) Bloom filter where a bit is set iff the respective counter in the CBF is strictly positive. We pick the number of hash functions that minimizes the false-positive probability [11].

*Access Strategy Algorithms compared:* Recall that  $\text{CS}_{\text{FNA}}$  makes use of an algorithm for solving the CS problem for the case where indicators exhibit no false-negatives. In our evaluation, we make use of the  $\text{DS}_{\text{PGM}}$  algorithm from [14]. This strategy was shown to produce a  $(\log M)$ -approximation for the CS problem with no false-negatives. By Theorem 7 this guarantee also applies to the general CS problem. Furthermore, this algorithm exhibits close-to-optimal results in practice, when tested on real-world workloads [14].

We consider two benchmarks for evaluating the performance of  $\text{CS}_{\text{FNA}}$ : (i) applying the vanilla  $\text{DS}_{\text{PGM}}$  algorithm ( $\text{CS}_{\text{FNO}}$ ), which only considers accessing caches with a positive indication (albeit stale), using only the estimates of  $\pi_j$  for every cache  $j$ , and using  $\nu_j = 1$  for all  $j$ , and (ii) the ideal strategy that uses *perfect information (PI)*, i.e., a strategy that always has access to the precise cache content, which accesses the cheapest cache containing an item if such a cache exists, and doesn't access any cache otherwise.

Throughout our evaluation, both  $\text{CS}_{\text{FNA}}$  and  $\text{CS}_{\text{FNO}}$  evaluate  $q_j$  with a time horizon of  $T = 100$  requests and using  $\delta = 0.25$  for the weighting of the moving average. Furthermore, each cache  $j$  re-estimates the false-positive ratio  $\text{FP}_j$  and the false-negative ratio  $\text{FN}_j$  once every 50 insertions to the cache.

*Evaluation metric:* We consider the mean service cost per request over the entire input. We also consider the *normalized cost* where we divide each algorithm's mean cost by that of the PI strategy. While infeasible, it is instructive to use it as a lower bound on the cost of *any* policy for solving the CS problem.

*Baseline scenario:* Unless stated otherwise, our evaluation considers three 10K elements caches whose access costs are 1, 2, and 3, and a miss penalty of 100 (i.e., 50 times the average cache access cost). Similar cache sizes were considered by existing works in the field [23], [24], and can further be motivated, e.g., by Trivago's Memcached [32] that utilizes a distributed system of caches, each of size 4GB, containing items with a typical size of about 1MB.

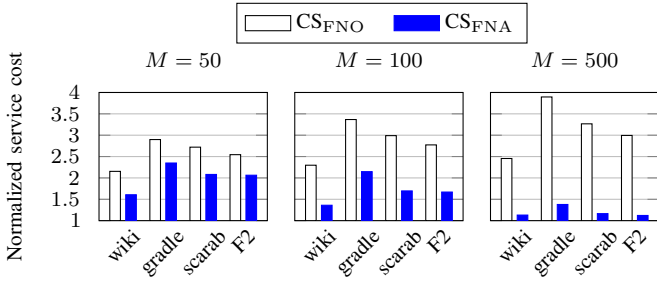


Fig. 3. Normalized cost of the heterogeneous 3-caches baseline scenario for varying traces and miss penalty values.

The update interval is measured by the number of insertions. In our baseline scenario,  $0.1 \cdot C_j$  insertions are performed between subsequent indicator advertisements. This translates to an advertisement once every 1K insertions for the default 10K-items cache. This is in accordance with previous work evaluating such systems [1]. Note that periodically advertising the indicator is sometimes done once in every fixed time interval (e.g., by Squid [33]). However, the *optimal* time interval length strongly depends on the workload being served. Our approach removes this dependency on the characteristics of the workload, and allows for a clearer evaluation of the effect the various system parameters have on performance, in scenarios where indicators become stale.

The advertised indicator of each cache  $j$  uses  $bpe = 14$ , which implies an indicator size of  $14 \cdot C_j$ , where the number of hash functions is optimized to minimize the false-positive ratio. In particular, in our baseline scenario, this translates to a designed false-positive ratio of 0.1% [13]. In each evaluation, we explore the impact of varying one of the system’s parameters, where the remaining parameters are set according to our baseline scenario. Our Python code is available in [34].

### B. Impact of Miss Penalty and Workload Diversity

We first compare the performance of  $CS_{FNO}$  and  $CS_{FNA}$  when varying the miss penalty values  $M$  in the range  $\{50, 100, 500\}$ . The results in Fig. 3 show that while the performance of the false-negative oblivious policy  $CS_{FNO}$  degrades as the miss penalty increases, the performance of our proposed false-negative aware algorithm  $CS_{FNA}$  improves significantly. Furthermore, the performance of  $CS_{FNA}$  tends to the *optimal* performance as the miss penalty increases. This behavior follows from the fact that a higher miss penalty accentuates the impact of false-negative events. In particular, ignoring negative indications (as is done by  $CS_{FNO}$ ) is severely penalized by an increased expected miss cost in cases where the miss penalty is large.

Fig. 3 also demonstrates significant differences across distinct workloads.  $CS_{FNO}$ ’s worst performance is exhibited for the Gradle trace, whereas its best performance is obtained for the Wiki trace. To understand this phenomenon, we observe that Gradle exhibits a high recency-bias, where items are requested shortly after their first appearance. As false-negatives occur when the indicator does not reflect the insertion of new items,  $CS_{FNO}$ , which never accesses caches with a negative indication,

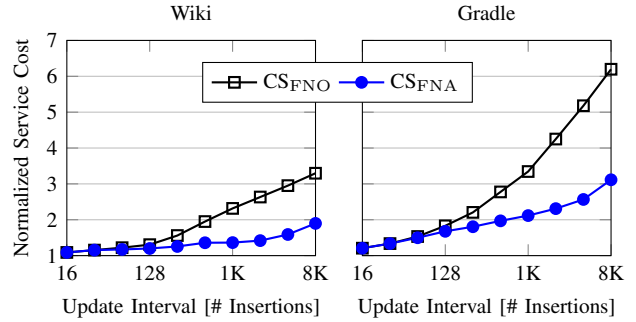


Fig. 4. Normalized cost of the heterogeneous 3-caches baseline scenario for varying update intervals. Update intervals are measured by the number of cache insertions between subsequent updates.

fails to take advantage of this recency-bias. In contrast, the Wiki trace is more frequency-biased, which implies that popular items do not rapidly change over time and that the impact of false-negatives is less pronounced. We continue with the Wiki and Gradle traces which are more sensitive to false negatives.

### C. Impact of Advertisement Policy and Indicator Parameters

1) *Update interval*: We now turn to study the effect of staleness on the performance of our algorithm. To this end, we let the update interval, which is the number of insertions between indicator advertisement, vary between 16 and 8K (8192), and consider the normalized cost of both  $CS_{FNA}$  and  $CS_{FNO}$ . These results are presented in Fig. 4, where we consider the performance for the Gradle and Wiki workloads.

Our results show that both algorithms’ performance degrades as the update interval increases. When updates are relatively frequent (i.e., up to 128), the performance of  $CS_{FNA}$  and  $CS_{FNO}$  is similar. However, a significant gap emerges between the performance of both algorithms for larger update intervals. In particular, the performance of  $CS_{FNO}$ , which ignores negative indications, quickly degrades, whereas  $CS_{FNA}$  shows a considerably milder degradation. This phenomenon is directly related to the fact that when the update interval is large, the false-negative ratio increases significantly (as demonstrated in Fig. 1a). Under such regimes,  $CS_{FNO}$  fails to access a cache even when the item is available at the cache, whereas  $CS_{FNA}$  relies on its false-negative awareness to make accesses even in cases of negative indications, taking into account the false-negative ratio estimation provided by the caches. Our results imply that  $CS_{FNA}$  matches the performance of  $CS_{FNO}$  while using a significantly lower bandwidth overhead for cache advertisements. For instance, for the Wiki workload  $CS_{FNA}$  matches the service cost as  $CS_{FNO}$  while using 16x less bandwidth for indicator advertisements. To see this, notice that  $CS_{FNA}$ ’s cost using an update interval of 8K is on par with that of  $CS_{FNO}$  with an update interval of 512.

2) *Indicator size*: Fig. 5 illustrates our results for varying the size of the indicator being used and advertised by the cache. We vary the number of indicator bits per cached element (bpe) and study the impact of the indicator’s size on the service cost. In our evaluation, we compare the performance of  $CS_{FNO}$  and  $CS_{FNA}$  with update intervals of 256 and 1024.

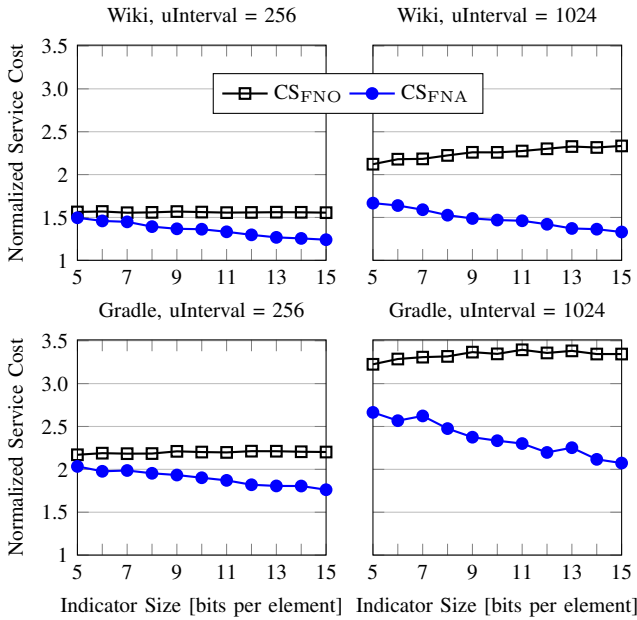


Fig. 5. Normalized cost of the heterogeneous 3-caches baseline scenario for varying indicator sizes, measured by bits-per-cached-element (bpe).

As expected,  $CS_{FNA}$ 's performance improves when increasing the indicator size as larger indicators exhibit fewer false-positive errors. Interestingly, and somewhat counter-intuitively, there exist cases where the performance of  $CS_{FNO}$  does not improve when increasing the indicator size, and in some cases, performance actually *degrades*. To explain this anomaly, let us understand the impact of false-positive and false-negative indications and their interplay. First, note that the false-positive rate is often inversely proportional to the false-negative rate. I.e., a constant decrease in the false-positive ratio is usually associated with an increase in the false-negative ratio. An extreme case occurs when all indications are negative, thus exhibiting a false-positive ratio of 0 and a sizeable false-negative ratio. Next, note that a false-positive event typically translates to unnecessary cache access, resulting in a relatively small penalty (e.g., an access cost of 1, 2, or 3 in our evaluation). However, a false-negative event typically translates to a “non-compulsory” miss, translating to a high miss penalty (e.g., 100, in our evaluation). It follows that even a mild decrease in the false-positive ratio may result in a non-negligible increase in the false-negative ratio that may nullify its benefits. Such effects are especially significant when the miss penalty is high, which is common as misses often result in accessing memories whose access time may be orders of magnitude higher than that of the cache [23], [35]. Still, our proposed false-negative aware algorithm  $CS_{FNA}$  handles such scenarios seamlessly and benefits from the reduced false-positive ratio without adverse performance impact.

#### D. Impact of Caching Capacity

We now study the effect of having a larger or more diverse caching capacity on system performance. For such

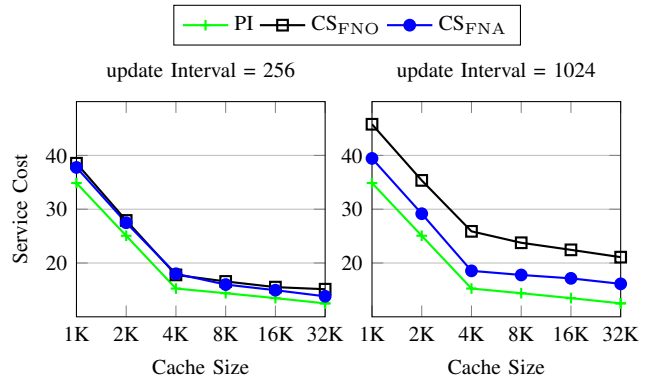


Fig. 6. Cost of the heterogeneous 3-caches baseline for varying cache size.

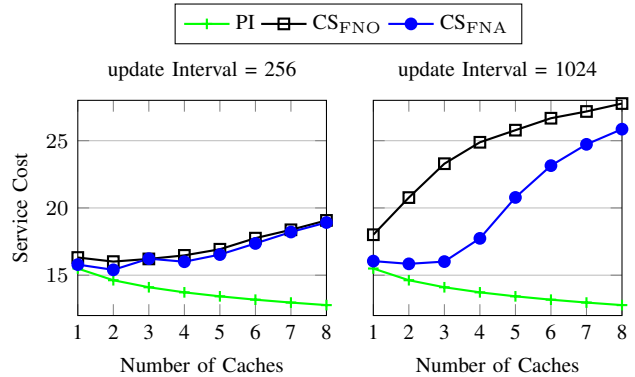


Fig. 7. Effect of varying the number of caches on the service cost. The access cost to each cache is 2, and the miss penalty is 100.

an evaluation, we use a 4.3M requests from the Wiki trace (instead of 1M). Further, we now consider the *actual* mean cost per request (and not the normalized service cost) as the cost of PI decreases when increasing the caching capacity.

1) *Scaling the cache size:* We study the impact of the cache size on the performance of  $CS_{FNO}$  and  $CS_{FNA}$  with update interval of 256 and 1024. The results in Fig. 6 show that, as could be expected, for every given setting, scaling-up the caches' capacities decreases the service cost due to the improved hit ratio. Our results show that when updates are relatively frequent (e.g., the case where the update interval is 256), the performance of  $CS_{FNO}$  is comparable to that of  $CS_{FNA}$  and they both exhibit a performance close to that of the ideal PI strategy. However, once the updates are less frequent (e.g., the case where the update interval is 1024),  $CS_{FNO}$  exhibits a significant degradation in performance.  $CS_{FNA}$ , on the other hand, is far less affected by the increase in the update interval, and is still quite comparable to PI. In general,  $CS_{FNA}$  shows up to 25% reduction in cost compared to  $CS_{FNO}$ . The differences between  $CS_{FNO}$  and  $CS_{FNA}$  become more accentuated when one considers the cache size required to maintain a certain level of cost;  $CS_{FNA}$  performs better with 4K items caches, than  $CS_{FNO}$  with caches of size 32K.

2) *Scaling the number of caches:* We now vary the number of caches in homogeneous settings. All caches have an access cost of 2, ensuring that the average access cost is the same as

in other scenarios examined in our evaluation. Fig. 7 shows the results for update intervals of 256 and 1024. Notice that  $CS_{FNA}$  consistently outperforms  $CS_{FNO}$ , and the difference is more significant for large update intervals. The results also imply that having more caches may hinder the performance of  $CS_{FNA}$  and  $CS_{FNO}$ . Intuitively, in such a case there are more false positives, and it is harder to guarantee that we access true positive items. Similarly, there are more negative indications which makes it harder for  $CS_{FNA}$  to identify a false-negative.

## VII. CONCLUSIONS

This work studies the cache selection problem while using approximate indicators exhibiting both false-positive and false-negative errors. The client in such a system selects a subset of the caches to minimize the expected service cost. While there is extensive work in this field, all previous access strategies do not access caches with negative indications. While reasonable at first glance, our work shows that such an omission severely hinders the system performance. We argue that caches that periodically advertise their content indicators inherently introduce false-negative indications, and the rate of such indications is non-negligible. In particular, we show that it is sometimes advisable to access caches with a negative indication, as it may reduce the overall system cost.

We devise false-negative-aware access strategies in two main scenarios: (i) fully-homogeneous settings, where we show a policy that attains the optimal (minimal) access cost, and (ii) general heterogeneous environments, where we present a strategy for which we can bound its approximation guarantee compared to the optimal solution. We complete our study through an extensive evaluation based on real system traces. Our results show that our proposed methods perform significantly better than the state-of-the-art in diverse settings. Furthermore, our false-negative aware solutions can match the cost of competitive false-negative oblivious approaches while requiring an order of magnitude fewer resources (e.g., caching capacity or bandwidth required for indicators advertisement).

Our results demonstrate the potential benefits of embracing false-negative awareness into the algorithmic design space. We expect our work to further induce both analytic and experimental research on the role of false-negatives in large distributed systems, including dealing with non-homogeneous object size, adhering to bandwidth constraints, and studying correlated distributed caching schemes.

## REFERENCES

- [1] L. Fan, P. Cao, J. Almeida, and A. Z. Broder, "Summary cache: a scalable wide-area web cache sharing protocol," *IEEE/ACM Trans. Netw.*, vol. 8, no. 3, pp. 281–293, 2000.
- [2] B. M. Maggs and R. K. Sitaraman, "Algorithmic nuggets in content delivery," *ACM SIGCOMM CCR*, vol. 45, no. 3, pp. 52–66, 2015.
- [3] X. Guo, T. Wang, and S. Wang, "Joint optimization of caching and routing strategies in content delivery networks: A big data case," in *IEEE ICC*, 2019.
- [4] R. Hou *et al.*, "Bloom-filter-based request node collaboration caching for named data networking," *Clust. Comput.*, vol. 22, no. 3, pp. 6681–6692, 2019.
- [5] X. Wang *et al.*, "Cache in the air: exploiting content caching and delivery techniques for 5G systems," *IEEE Comm. Mag.*, vol. 52, no. 2, pp. 131–139, 2014.
- [6] M. Zhang, H. Luo, and H. Zhang, "A survey of caching mechanisms in information-centric networking," *IEEE Comm. Surv. & Tut.*, vol. 17, no. 3, pp. 1473–1499, 2015.
- [7] O. Rottenstreich and I. Keslassy, "The bloom paradox: When not to use a bloom filter," *IEEE/ACM Trans. Netw.*, vol. 23, no. 3, pp. 703–716, 2015.
- [8] T. Le, Y. Lu, and M. Gerla, "Social caching and content retrieval in disruption tolerant networks (dtns)," in *IEEE ICNC*, 2015, pp. 905–910.
- [9] B. H. Bloom, "Space/time trade-offs in hash coding with allowable errors," *Commun. ACM*, vol. 13, no. 7, pp. 422–426, 1970.
- [10] F. Bonomi *et al.*, "An improved construction for counting bloom filters," in *ESA*, 2006, pp. 684–695.
- [11] L. Luo *et al.*, "Optimizing bloom filter: Challenges, solutions, and comparisons," *IEEE Comm. Surv. & Tut.*, vol. 21, no. 2, pp. 1912–1949, 2018.
- [12] G. Einziger and R. Friedman, "Counting with tinytable: Every bit counts!" *IEEE Access*, vol. 7, pp. 166 292–166 309, 2019.
- [13] S. Tarkoma, C. E. Rothenberg, and E. Lagerspetz, "Theory and practice of bloom filters for distributed systems," *IEEE Comm. Surv. & Tut.*, vol. 14, no. 1, pp. 131–155, 2012.
- [14] I. Cohen, G. Einziger, R. Friedman, and G. Scalosub, "Access strategies for network caching," *IEEE/ACM Trans. Netw.*, vol. 29, pp. 609–622, 2021.
- [15] L. Chen, A. Giovanidis, W. Wang, and L. Shan, "Sequential resource access: Theory and algorithm," in *IEEE INFOCOM*, 2021. [Online]. Available: <https://arxiv.org/pdf/2012.03734.pdf>
- [16] D. Guo, Y. Liu, X. Li, and P. Yang, "False negative problem of counting bloom filter," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 5, pp. 651–664, 2010.
- [17] Y. Zhu and H. Jiang, "False rate analysis of bloom filter replicas in distributed systems," in *ICPP*, 2006, pp. 255–262.
- [18] I.-W. Ting and Y.-K. Chang, "Improved group-based cooperative caching scheme for mobile ad hoc networks," *J. Parallel. and Distrib. Comp.*, vol. 73, no. 5, pp. 595–607, 2013.
- [19] I. Cohen, G. Einziger, and G. Scalosub, "Self-adjusting advertisement of cache indicators with bandwidth constraints," in *IEEE INFOCOM*, 2021.
- [20] S. Z. Kiss *et al.*, "Bloom filter with a false positive free zone," in *IEEE INFOCOM*, 2018, pp. 1412–1420.
- [21] A. Kosta, N. Pappas, and V. Angelakis, "Age of information: A new concept, metric, and tool," *Foundat. and Trend. in Netw.*, vol. 12, no. 3, pp. 162–259, 2017.
- [22] R. D. Yates, P. Ciblat, A. Yener, and M. Wigger, "Age-optimal constrained cache updating," in *IEEE ISIT*, 2017, pp. 141–145.
- [23] G. Einziger, O. Eytan, R. Friedman, and B. Manes, "Adaptive software cache management," in *ACM Middleware*, 2018, pp. 94–106.
- [24] G. Einziger, R. Friedman, and B. Manes, "Tinylfu: A highly efficient cache admission policy," *TOS*, vol. 13, no. 4, pp. 35:1–35:31, 2017.
- [25] D. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation," *J. Mach. Learn. Tech.*, vol. 2, no. 1, pp. 37–63, 2011.
- [26] I. Cohen, G. Einziger, and G. Scalosub, "On the power of false negative awareness in indicator-based caching systems." [Online]. Available: <https://arxiv.org/abs/2102.01724>
- [27] G. Urdaneta, G. Pierre, and M. van Steen, "Wikipedia workload analysis for decentralized hosting," *Comp. Netw.*, vol. 53, no. 11, pp. 1830–1845, 2009.
- [28] "Caffeine's simulator cache traces." [Online]. Available: <https://github.com/ben-manes/caffeine/tree/master/simulator/src/main/resources/com/github/benmanes/caffeine/cache/simulator/parser>
- [29] M. Liberatore and P. Shenoy, "Umass trace repository," 2016. [Online]. Available: <http://traces.cs.umass.edu/>
- [30] "Memcached." [Online]. Available: <https://memcached.org>
- [31] G. Einziger, R. Friedman, and E. Kibbar, "Kaleidoscope: Adding colors to kademia," in *IEEE P2P*, 2013.
- [32] I. Petrovic and P. Pokalyukhina, "How trivago reduced memcached memory usage by 50%," 2017. [Online]. Available: <https://tech.trivago.com/2017/12/19/how-trivago-reduced-memcached-memory-usage-by-50/>
- [33] "squid-cache cache digests." [Online]. Available: <https://wiki.squid-cache.org/SquidFaq/CacheDigests>
- [34] "Access strategies with staleness simulations." [Online]. Available: [https://github.com/ofanan/Access\\_strategies\\_with\\_staleness](https://github.com/ofanan/Access_strategies_with_staleness)
- [35] C. Scott, "Latency numbers every programmer should know," 2020. [Online]. Available: [https://colin-scott.github.io/personal\\_website/research/interactive\\_latency.html](https://colin-scott.github.io/personal_website/research/interactive_latency.html)