Scheduling Group Tests over Time

Akhil Bhimaraju
University of Illinois Urbana-Champaign
akhilb3@illinois.edu

Lav R. Varshney
University of Illinois Urbana-Champaign
varshney@illinois.edu

2022 IEEE International Symposium on Information Theory (ISIT) | 978-1-6654-2159-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISIT50566.2022.9834434

***Abstract*—Group testing has been successful in minimizing the cost of testing a large batch of samples by pooling them together. In this work, we study the setting where samples arrive over time. Since not all samples are available at the same time, we incur a waiting cost in letting many samples accumulate. However, testing too soon leads to a large testing cost by missing the benefits of pooling a larger number of samples. We consider the problem of minimizing the combined objective of average wait time plus testing cost, and develop online algorithms that are provably competitive for a broad range of testing-cost functions. We also give a lower bound on the competitive ratio that no online algorithm can beat.**

***Index Terms*—Online algorithm, competitive ratio, scheduling**

I. Introduction

Group testing [1] is a means to test a large number of samples for their infection status by pooling the samples intelligently to reduce the number of tests required: if a pool tests negative, then every sample included in the pool is negative, and if the pool tests positive, at least one sample in the pool is positive. Group testing can lead to dramatic reductions in the number of tests required compared to naively testing every sample for the infection [2]–[4], and information-theoretically optimal algorithms for group testing have been developed for multiple regimes of infection prevalence [5]–[8]. Please see [9] for a comprehensive survey. More recently, there has also been work on modeling connections between individuals (and thus the samples provided by the individuals) with graphical structures and using this information to further reduce the number of tests required [10]–[13].

Much of this prior theoretical work focuses on proving the optimality of various group-testing algorithms in the asymptotic sense as the number of available samples goes to infinity. However, not all samples might be available at the same time, and waiting for more samples to accumulate increases the turnaround time for samples that arrived earlier. Further, since pandemic-control measures typically involve quarantining individuals suspected of being infected, waiting too long before testing also imposes societal costs in terms of unnecessarily quarantining uninfected individuals. The tradeoff between quarantining and testing costs has been empirically investigated in [14] for a fixed set of individuals with a community structure. To the best of our knowledge, there has been no work on provably competitive algorithms that inform when to perform the tests as new samples arrive over time.

To tackle this problem, we frame it as a minimization of the average wait time of the samples plus the average testing cost. We assume we have no information about samples that may arrive in the future, and hence the proposed algorithm must be online (see Sec. II) and continuously update the known information as new samples arrive. The specific (adaptive or non-adaptive) group-testing approach within the algorithm being used can depend on various factors such as the current infection prevalence, testing capacity, and tolerance for errors. Further, even for the same group-testing scheme, the cost of testing might be different depending on local conditions governing the costs of chemical reagents, maintaining a lab, and desired accuracy for the performed tests. In order to not restrict ourselves to any particular deployment, we consider a very broad class of testing-cost functions, including those that correspond to group tests with various kinds of noise.

As a performance metric for online algorithms, we consider the competitive ratio, which is the worst-case ratio between the cost of running the online algorithm and the cost of the optimal testing schedule. So if an online algorithm (i.e., an algorithm which uses no information about the arrival of future samples in determining which samples to test at a certain time) has a certain competitive ratio, then the cost of running the algorithm is never worse than this ratio times the optimal cost (computed in hindsight with full information about the arrival times of all samples). This also means that our performance guarantee holds regardless of the potentially time-varying statistics of the sample-arrival process, and our algorithm makes no assumptions on these statistics. Moreover, our guarantee does not depend on the total number of samples in the system and holds for every possible instance of the problem.

A related line of work is on batch-service queuing [15]–[20], where a (typically Poisson) distribution is assumed for the sample-arrival process with the testing done in groups once a large enough number of samples have accumulated. Much work has focused on finding the optimal batch size, or the number of accumulated samples, to be tested. In contrast, our model and algorithm make no assumptions on the statistics of the sample-arrival process. Our performance guarantee in terms of the competitive ratio with respect to the offline optimal solution holds for any (static or time-varying) sample-arrival process. Another related line of work is on accurately estimating the state of an ongoing epidemic in a Susceptible-

This work was supported in part by NSF grant ECCS-2033900, and the Center for Pathogen Diagnostics through the ZJU-UIUC Dynamic Engineering Science Interdisciplinary Research Enterprise (DESIRE).

Infected-Recovered (SIR) model using group testing to inform quarantining measures and keep the total number of infections at manageable levels [21], [22].

Our problem also bears similarity to job scheduling on processors [23], [24], where the objective is to minimize the weighted sum of the completion times of the jobs. More specifically, "speed scaling," where the objective is to minimize the average wait time until completion plus the energy consumed [25]–[27] is quite similar to our formulation of wait time plus testing cost. However, group testing suggests a concave testing cost (see Claim 2), which is incurred at discrete instants of time when the tests happen. This makes the convex (or linear) programs and primal-dual analysis such as used in [24], [27] inapplicable for our case. It is not even clear if we can write an integer program here which could be relaxed into a continuous concave optimization program.

The remainder of this paper is organized as follows. Sec. II describes the system model, objective, and assumptions on the testing-cost function. Sec. III then presents our online algorithm which balances the waiting and testing costs to achieve the performance guarantee (Theorem 1) in terms of its competitive ratio. Sec. IV gives a lower bound on the competitive ratio that any online algorithm might achieve (Theorem 2), and Sec. V concludes the paper with a discussion on directions for future work.

II. Model

Let the arrival times of the samples to be tested be denoted by $a_1, a_2, \ldots, a_n$, subject to $a_1 \leq a_2 \leq \cdots \leq a_n$. Using an algorithm ALG, the samples are tested in m^{ALG} batches $\mathcal{B}_1^{\text{ALG}}, \mathcal{B}_2^{\text{ALG}}, \ldots, \mathcal{B}_{m^{\text{ALG}}}^{\text{ALG}}$ at times $s_1^{\text{ALG}}, s_2^{\text{ALG}}, \ldots, s_{m^{\text{ALG}}}^{\text{ALG}}$, where $\mathcal{B}_j^{\text{ALG}}$ is the set of samples tested together at time s_j^{ALG}, with $\mathcal{B}_i^{\text{ALG}} \cap \mathcal{B}_j^{\text{ALG}} = \emptyset$ for $i \neq j$. For each sample i, define d_i^{ALG} as the time when it gets tested: $d_i^{\text{ALG}} = s_j^{\text{ALG}}$ for $i \in \mathcal{B}_j^{\text{ALG}}$. Note that we need $d_i^{\text{ALG}} \geq a_i$ for all $1 \leq i \leq n$ for ALG to be valid. Let the function $f : \mathbb{N} \cup \{0\} \mapsto \mathbb{R}^+ \cup \{0\}$ be such that $f(x)$ denotes the cost incurred when we test x samples together. Further, based on the nature of all standard group tests (noisy/noiseless, adaptive/non-adaptive), we make the following assumptions on $f(\cdot)$:

(i) $f(0) = 0$;
(ii) $f(\cdot)$ is a non-decreasing function, i.e., $f(x) \geq f(y)$ if $x \geq y$;
(iii) the increase in the testing cost per sample decreases with the number of samples: $f(x+1) - f(x) \leq f(x) - f(x-1)$ for all $x \in \mathbb{N}$.

We define the cost of the algorithm ALG, J^{ALG}, on this instance as the per-sample average of the total time the samples are kept waiting[1] plus the testing cost:

$$J^{\text{ALG}} = \frac{1}{n}\left(\sum_{i=1}^{n}(d_i - a_i) + \sum_{j=1}^{m^{\text{ALG}}} f\left(|\mathcal{B}_j^{\text{ALG}}|\right)\right). \qquad (1)$$

[1]Note that we only consider the time the sample is kept waiting until the testing process starts. Any "waiting time" during the testing process can be included in the testing-cost function $f(\cdot)$.

Let OPT denote the algorithm that selects the batches $\mathcal{B}$ to minimize the value of J: $J^{\text{OPT}} = \min_{\text{ALG}} J^{\text{ALG}}$.

An algorithm ALG is *online* if in determining to test the batch $\mathcal{B}_j^{\text{ALG}}$ at time s_j^{ALG}, ALG makes no use of information about any sample in $\{i : a_i > s_j^{\text{ALG}}\}$. An online algorithm ALG is said to be *ρ-competitive* for $\rho \geq 1$ if the supremum of $\frac{J^{\text{ALG}}}{J^{\text{OPT}}}$ over all possible problem instances $\{a_1, a_2, \ldots . a_n\}$ for all n is less than or equal to ρ, and ρ is a *competitive ratio* for ALG. Our objective in this paper is twofold: design online algorithms for this problem that admit a low (and constant) competitive ratio; and establish lower bounds on the value of ρ that online algorithms can achieve.

III. 4-Competitive Algorithm

In this section, we present our algorithm for the queued group-testing problem from Sec. II and prove it is 4-competitive. Let us first introduce some useful notation. For a given problem instance, let W^{ALG} denote the average wait time and F^{ALG} denote the average testing cost:

$$W^{\text{ALG}} = \frac{1}{n}\sum_{i=1}^{n}(d_i^{\text{ALG}} - a_i) \text{ and } F^{\text{ALG}} = \frac{1}{n}\sum_{j=1}^{m^{\text{ALG}}} f\left(|\mathcal{B}_j^{\text{ALG}}|\right);$$
$$J^{\text{ALG}} = W^{\text{ALG}} + F^{\text{ALG}}.$$

For an algorithm ALG, let u_t^{ALG} denote the number of samples that have arrived by time t but not yet been tested by time t: $u_t^{\text{ALG}} = |\{i : a_i \leq t < d_i^{\text{ALG}}\}|$. Observe that the average wait time can be written as

$$W^{\text{ALG}} = \frac{1}{n}\int_0^{\infty} u_\tau^{\text{ALG}} d\tau. \qquad (2)$$

The main tradeoff in this problem is the following. If we wait for a long time before testing, we accumulate a lot of samples, thus decreasing the per-sample testing cost (via group testing), but increase the waiting time. However, if we test too aggressively soon after a few samples have arrived, we might lose out on accumulating enough samples before testing, and thus incur a high per-sample testing cost even though the waiting time would be lower.

Algorithm 1 WaitTillEqual (WTE)

Initialize: $b_{\text{prev}} \leftarrow 0$
At each time t:
if $\int_{b_{\text{prev}}}^{t} u_\tau d\tau = f(|\{i : a_i \in [0, t] \text{ and } d_i \notin [0, t)\}|)$ **then**[2]
  Test all the available samples together at time t
  $b_{\text{prev}} \leftarrow t$
end if

We propose the WaitTillEqual (WTE) algorithm which balances these two components. We first compute the cumulative waiting time of arrived samples that are yet to be tested and compare this with the cost of testing them all together. Initially, the cumulative wait time would be small, but as time progresses, it would start getting larger. Once its value

[2]For $b_{\text{prev}} \neq 0$, the right side can be written as $f\left(|\{i : a_i \in (b_{\text{prev}}, t]\}|\right)$.

equals the cost of testing all these samples together, we test them. We state this formally as Alg. 1, and the 4-competitive performance guarantee as Theorem 1.

Theorem 1. *The online algorithm* WAITTILLEQUAL *(Alg. 1) admits a competitive ratio of* 4*, i.e.,* $J^{\text{WTE}} \leq 4J^{\text{OPT}}$ *for any problem instance.*

It is straightforward to see that WTE is online since computing the set u_t only requires us to know the samples that have already arrived, but have not yet been tested. Before proving Theorem 1, we first state a key result as Claim 1.

Claim 1. *The following is true for the* WTE *algorithm:*[3]

$$\frac{1}{n}\int_0^\infty \left(u_\tau^{\text{WTE}} - u_\tau^{\text{OPT}}\right)^+ d\tau \leq 2F^{\text{OPT}},$$

where $(x)^+ = \max\{0, x\}$.

While we defer the proof of Claim 1 to Sec. III-A, let us now see how this claim gives us Theorem 1.

Proof of Theorem 1: Observe that for any algorithm ALG, we can write

$$u_\tau^{\text{ALG}} \leq u_\tau^{\text{OPT}} + \left(u_\tau^{\text{ALG}} - u_\tau^{\text{OPT}}\right)^+ \text{ for all } \tau \geq 0.$$

Writing this for WTE and integrating over time gives us

$$\int_0^\infty u_\tau^{\text{WTE}} d\tau \leq \int_0^\infty u_\tau^{\text{OPT}} d\tau + \int_0^\infty \left(u_\tau^{\text{WTE}} - u_\tau^{\text{OPT}}\right)^+ d\tau.$$

Using (2) and Claim 1, we get

$$W^{\text{WTE}} \leq W^{\text{OPT}} + 2F^{\text{OPT}} = J^{\text{OPT}} + F^{\text{OPT}} \leq 2J^{\text{OPT}}. \quad (3)$$

Since WTE tests samples at the time instant when their testing cost is equal to their cumulative wait time, we have

$$W^{\text{WTE}} = F^{\text{WTE}},$$

which gives $J^{\text{WTE}} = 2W^{\text{WTE}}$. Since $W^{\text{WTE}} \leq 2J^{\text{OPT}}$ from (3), this gives us $J^{\text{WTE}} = 2W^{\text{WTE}} \leq 4J^{\text{OPT}}$, which concludes the proof. ■

A. Proof of Claim 1

Let us first make some useful observations about the cost function $f(\cdot)$, which we state as Claim 2.

Claim 2. *Under the assumptions on* $f(\cdot)$ *in Sec. II, the following are true:*

(i) $f(x+y) \leq f(x) + f(y)$ *for all* $x, y \in \mathbb{N}$*;*
(ii) $\theta f(x) + (1-\theta) f(y) \leq f\left(\lceil \theta x + (1-\theta) y \rceil\right)$ *for* $\theta \in [0,1]$*, where* $\lceil\cdot\rceil$ *denotes the integer ceiling function.*

Proof: For proving part (i), we use induction on y. Using assumptions on $f(\cdot)$, we get

$$\begin{aligned} f(x+1) - f(x) &\leq f(x) - f(x-1) \\ &\leq f(x-1) - f(x-2) \\ &\leq \cdots \leq f(1) - f(0) = f(1). \end{aligned}$$

[3] Note that if there are multiple optimal testing schedules, all of them should have the same J^{OPT}, but they might have different values for W^{OPT} and F^{OPT}. For the purpose of proving Theorem 1, it is sufficient if Claim 1 holds for any one possible optimal schedule.

This gives us $f(x+1) \leq f(x) + f(1)$, which is the base case for $y = 1$. Now assume that part (i) of Claim 2 is true for $y = k-1$, and consider

$$\begin{aligned} f(x+k) - f(x+k-1) &\leq f(x+k-1) - f(x+k-2) \\ &\leq f(x+k-2) - f(x+k-3) \\ &\leq \cdots \leq f(k) - f(k-1). \end{aligned}$$

This gives us $f(x+k) \leq f(x+k-1) + f(k) - f(k-1)$, but since part (i) is true for $y = k-1$, using $f(x+k-1) \leq f(x) + f(k-1)$, we get

$$f(x+k) \leq f(x) + f(k).$$

This completes the induction and so the assertion holds for all $y \in \mathbb{N}$. Since we have made no assumptions on x, the statement is thus true for all $x, y \in \mathbb{N}$.

For proving part (ii) of Claim 2, it will be helpful to define a real extension of $f(\cdot)$ using linear interpolation. Let $\tilde{f} : \mathbb{R}^+ \cup \{0\} \mapsto \mathbb{R}^+ \cup \{0\}$ be defined as follows:

$$\tilde{f}(x) = \begin{cases} f(x), & \text{if } x \in \mathbb{N} \cup \{0\}, \\ (x - \lfloor x \rfloor) f(\lceil x \rceil) + (\lceil x \rceil - x) f(\lfloor x \rfloor), & \text{else}, \end{cases}$$

where $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$ are the integer ceiling and floor functions respectively. It follows directly that $\tilde{f}$ is continuous and non-decreasing. Further, since the slope at $x \notin \mathbb{N} \cup \{0\}$ is $f(\lceil x \rceil) - f(\lfloor x \rfloor)$ which can only decrease with x from our assumptions on $f(\cdot)$, $\tilde{f}(\cdot)$ is a concave function in its domain. This gives us

$$\begin{aligned} \theta f(x) + (1-\theta) f(y) &= \theta \tilde{f}(x) + (1-\theta)\tilde{f}(y) \\ &\leq \tilde{f}(\theta x + (1-\theta) y) \\ &\leq \tilde{f}\left(\lceil \theta x + (1-\theta) y \rceil\right) \\ &= f\left(\lceil \theta x + (1-\theta) y \rceil\right), \end{aligned}$$

which completes the proof of part (ii) of Claim 2. ■

Let us now state a useful property of the optimal schedule OPT as Claim 3.

Claim 3. *There is an optimal testing schedule such that if a test is performed at* $t = s_j^{\text{OPT}}$ *for some* j*, then all the available untested samples at* t *are tested together, i.e.,* $u_t^{\text{OPT}} = 0$*.*

Proof: We prove this claim by showing that any (optimal) schedule can be converted, without increasing the cost, into one where the following property is true: if there is a test at some $t = s_j^{\text{OPT}}$, then $u_t^{\text{OPT}} = 0$.

In an optimal schedule OPT, assume only n_1 samples are tested at some time t, while n_2 samples, already available at t, are tested in a different batch along with n_{new} samples (all released at times after t) at some time $t + \Delta t$. We prove this claim by showing that at least one of the following does not increase the cost:

(i) test the $n_1 + n_2$ samples together at time t, or
(ii) test all the $n_1 + n_2 + n_{\text{new}}$ samples together at time $t + \Delta t$.

Let us assume the contrary (which will lead us to a contradiction): both the options above strictly increase the cost. Note that the waiting time of the n_{new} samples is the same in all cases. Further, the waiting time and testing costs of samples that are not in these $n_1+n_2+n_{\text{new}}$ samples is the same in all the cases. Considering only the terms that are different, the above two cases imply

$$f(n_1)+f(n_2+n_{\text{new}})+n_2\Delta t < f(n_1+n_2)+f(n_{\text{new}}) \text{ and}$$
$$f(n_1)+f(n_2+n_{\text{new}})+n_2\Delta t < f(n_1+n_2+n_{\text{new}}) + (n_1+n_2)\Delta t.$$

These inequalities give us

$$\Delta t < \frac{f(n_1+n_2)+f(n_{\text{new}})-f(n_1)-f(n_2+n_{\text{new}})}{n_2},$$
$$\Delta t > \frac{f(n_1)+f(n_2+n_{\text{new}})-f(n_1+n_2+n_{\text{new}})}{n_1}.$$

Putting the two together, we get

$$\frac{f(n_1)+f(n_2+n_{\text{new}})-f(n_1+n_2+n_{\text{new}})}{n_1} < \frac{f(n_1+n_2)+f(n_{\text{new}})-f(n_1)-f(n_2+n_{\text{new}})}{n_2}.$$

Rearranging the terms gives us

$$f(n_1)+f(n_2+n_{\text{new}}) < \tfrac{n_1}{n_1+n_2}f(n_1+n_2) + \tfrac{n_1}{n_1+n_2}f(n_{\text{new}}) + \tfrac{n_2}{n_1+n_2}f(n_1+n_2+n_{\text{new}}). \quad (4)$$

Using Claim 2(ii) with $\theta = \frac{n_1}{n_1+n_1}$, $x = n_1+n_2$, and $y=0$ gives

$$\frac{n_1}{n_1+n_2}f(n_1+n_2) \le f(\lceil n_1 \rceil) = f(n_1). \quad (5)$$

With $\theta = \frac{n_1}{n_1+n_2}$, $x = n_{\text{new}}$, and $y = n_1+n_2+n_{\text{new}}$ we get

$$\frac{n_1}{n_1+n_2}f(n_{\text{new}}) + \frac{n_2}{n_1+n_2}f(n_1+n_2+n_{\text{new}}) \le f(\lceil n_2+n_{\text{new}} \rceil) = f(n_2+n_{\text{new}}). \quad (6)$$

Substituting (5) and (6) into (4) gives

$$f(n_1)+f(n_2+n_{\text{new}}) < f(n_1)+f(n_2+n_{\text{new}})$$

which is a contradiction. ■

We are now ready to prove Claim 1.

Proof of Claim 1: Let OPT be a schedule with optimal cost that satisfies Claim 3. For proving Claim 1, we divide the integral $\int_0^\infty (u_\tau^{\text{WTE}} - u_\tau^{\text{OPT}})^+ d\tau$ into segments:

$$\int_0^\infty (u_\tau^{\text{WTE}} - u_\tau^{\text{OPT}})^+ d\tau = \sum_{j=1}^{m^{\text{WTE}}} \int_{s_{j-1}^{\text{WTE}}}^{s_j^{\text{WTE}}} (u_\tau^{\text{WTE}} - u_\tau^{\text{OPT}})^+ d\tau, \quad (7)$$

where we have defined s_0^{WTE} as 0. Note that for $\tau > s_{m^{\text{WTE}}}^{\text{WTE}}$, the integrand is 0 since all the samples have been tested by WTE, and so $u_\tau^{\text{WTE}} = 0$.

The set of testing batches in OPT can be partitioned into sets $\mathcal{S}_j$ based on the $[s_{j-1}^{\text{WTE}}, s_j^{\text{WTE}})$ segment in which OPT tests them:

$$\mathcal{S}_j = \left\{k : s_k^{\text{OPT}} \in [s_{j-1}^{\text{WTE}}, s_j^{\text{WTE}})\right\} \text{ for } j \in \{1,2,\ldots,m^{\text{WTE}}\},$$

and let $\mathcal{L}$ denote the leftover samples that OPT might test at a time $t \ge s_{m^{\text{WTE}}}^{\text{WTE}}$.

Observe that at $\tau = s_{j-1}^{\text{WTE}}$ for all j, $u_\tau^{\text{WTE}} = 0$ since WTE tests all the samples available at τ. So $(u_\tau^{\text{WTE}} - u_\tau^{\text{OPT}})^+ = 0$ at $\tau = s_{j-1}^{\text{WTE}}$. Further, if $\mathcal{S}_j = \emptyset$, then OPT does not test any samples before WTE tests all the available samples again, and so $(u_\tau^{\text{WTE}} - u_\tau^{\text{OPT}})^+ = 0$ for all $\tau \in [s_{j-1}^{\text{WTE}}, s_j^{\text{WTE}})$. This gives

$$\int_{s_{j-1}^{\text{WTE}}}^{s_j^{\text{WTE}}} \left(u_\tau^{\text{WTE}} - u_\tau^{\text{OPT}}\right)^+ d\tau = 0 \text{ if } \mathcal{S}_j = \emptyset. \quad (8)$$

For each $j \in \{1,2,\ldots,\max\{j : \mathcal{S}_j \ne \emptyset\}-1\}$, define $\mathcal{K}_j^{\text{NE}}$ to be the first non-empty $\mathcal{S}_k$ after j:

$$\mathcal{K}_j^{\text{NE}} = \mathcal{S}_{\min\{k \,:\, k>j \text{ and } \mathcal{S}_k \ne \emptyset\}}.$$

If $\mathcal{S}_j \ne \emptyset$, we have[4]

$$\int_{s_{j-1}^{\text{WTE}}}^{s_j^{\text{WTE}}} \left(u_\tau^{\text{WTE}} - u_\tau^{\text{OPT}}\right)^+ d\tau \le \int_{s_{j-1}^{\text{WTE}}}^{s_j^{\text{WTE}}} u_\tau^{\text{WTE}} d\tau = f\left(|\{i : s_{j-1}^{\text{WTE}} < a_i \le s_j^{\text{WTE}}\}|\right), \quad (9)$$

where the equality follows directly from the definition of the WTE algorithm. For all but the last non-empty set $\mathcal{S}_j$, the set $\left\{i : s_{j-1}^{\text{WTE}} < a_i \le s_j^{\text{WTE}}\right\}$ is a subset of the union of all the samples that OPT tests in this interval and the next non-empty interval. This is because any additional samples in $\left\{i : s_{j-1}^{\text{WTE}} < a_i \le s_j^{\text{WTE}}\right\}$ not tested by OPT in this interval have to be tested in the next testing batch from Claim 3. This gives us

$$\left\{i : s_{j-1}^{\text{WTE}} < a_i \le s_j^{\text{WTE}}\right\} \subseteq \bigcup_{k \in \mathcal{S}_j \cup \mathcal{K}_j^{\text{NE}}} \mathcal{B}_k^{\text{OPT}}$$

for $j \le \max\{k : \mathcal{S}_k \ne \emptyset\} - 1$. Using Claim 2(i) and (9), we get

$$\int_{s_{j-1}^{\text{WTE}}}^{s_j^{\text{WTE}}} \left(u_\tau^{\text{WTE}} - u_\tau^{\text{OPT}}\right)^+ d\tau \le \sum_{k \in \mathcal{S}_j \cup \mathcal{K}_j^{\text{NE}}} f\left(|\mathcal{B}_k^{\text{OPT}}|\right). \quad (10)$$

Similarly, for $j = \max\{k : \mathcal{S}_k \ne \emptyset\}$, we get

$$\int_{s_{j-1}^{\text{WTE}}}^{s_j^{\text{WTE}}} \left(u_\tau^{\text{WTE}} - u_\tau^{\text{OPT}}\right)^+ d\tau \le \sum_{k \in \mathcal{S}_j} f\left(\mathcal{B}_k^{\text{OPT}}\right) + f\left(|\mathcal{L}|\right).$$

Since $\mathcal{K}_j^{\text{NE}}$ is also one of $\{\mathcal{S}_k\}$ and unique for all j, adding these inequalities together with (8) gives us

$$\sum_{j=1}^{m^{\text{WTE}}} \int_{s_{j-1}^{\text{WTE}}}^{s_j^{\text{WTE}}} \left(u_\tau^{\text{WTE}} - u_\tau^{\text{OPT}}\right)^+ d\tau \le 2\sum_{j=1}^{m^{\text{WTE}}} \sum_{k \in \mathcal{S}_j} f\left(|\mathcal{B}_k^{\text{OPT}}|\right) + f\left(|\mathcal{L}|\right).$$

However, using Claim 2(i), OPT cannot test the leftover $\mathcal{L}$ samples using a testing cost lower than $f\left(|\mathcal{L}|\right)$. This gives us

$$\sum_{j=1}^{m^{\text{WTE}}} \int_{s_{j-1}^{\text{WTE}}}^{s_j^{\text{WTE}}} \left(u_\tau^{\text{WTE}} - u_\tau^{\text{OPT}}\right)^+ d\tau \le 2nF^{\text{OPT}}.$$

[4] Note that for $j=1$, we have $s_{j-1}^{\text{WTE}} = 0$, and for this case, the appropriate set is $\{s_{j-1}^{\text{WTE}} \le a_i \le s_j^{\text{WTE}}\}$.

Together with (7), this concludes the proof of Claim 1. ■

Remark 1. *We can view* (10) *as a variant of the amortized local-competitiveness argument [28], [29]. If the summation on the right side was only over $\mathcal{S}_j$ (with an additional factor of* 2*), it would be local competitiveness. However, we amortize it over $\mathcal{S}_j$ and the next non-empty $\mathcal{S}_k$, $\mathcal{K}_j^{\mathrm{NE}}$.*

IV. Lower Bound

Here we give a lower bound on the competitive ratio that any online algorithm can achieve, formalized as Theorem 2.

Theorem 2. *Let* ALG *be an online algorithm and $f(\cdot)$ be the testing-cost function for the queued group-testing problem that achieves a competitive ratio of ρ. Then,*

$$\rho \geq 1 + \frac{1}{2}\left(\sqrt{\Gamma^2 - 4\Gamma + 8} - \Gamma\right),$$

where $\Gamma = \min_{x\in\mathbb{N}} \frac{f(2x)}{f(x)}$.

Proof: Consider a problem instance where n samples arrive at time $t = 0$. Assume the algorithm ALG tests them at some time $t = \Delta t$ (testing them together is better from Claim 3, so if ALG tests them separately, the competitive ratio will only be worse). From this, the competitive ratio is lower-bounded by

$$\rho \geq \frac{f(n) + n\Delta t}{f(n)}, \tag{11}$$

since the optimal is to test them all at $t = 0$.

If an additional n samples arrive at time $t = \Delta t + \epsilon$ for some $\epsilon > 0$, ALG would still test the first n samples at $t = \Delta t$ because ALG is online and the known information in both cases (when the additional samples do and do not arrive) is identical until time $t = \Delta t$. Since ALG tests n of them at Δt, its total cost can be no better than $2f(n) + n\Delta t$.[5] However, in this case, it is better to wait for all samples to arrive and test them all together at time $t = \Delta t + \epsilon$. Since the optimal algorithm can only do better, the competitive ratio is lower bounded by

$$\rho \geq \frac{2f(n) + n\Delta t}{f(2n) + n(\Delta t + \epsilon)}. \tag{12}$$

Since ϵ can be arbitrarily small, using $\epsilon \to 0$, (11) and (12) give us

$$\begin{aligned}\rho &\geq \max\left\{\frac{f(n) + n\Delta t}{f(n)}, \frac{2f(n) + n\Delta t}{f(2n) + n\Delta t}\right\} \\ &= 1 + \max\left\{\frac{n\Delta t}{f(n)}, \frac{2f(n) - f(2n)}{f(2n) + n\Delta t}\right\}.\end{aligned}$$

Since this is true for any Δt that an online algorithm might use, any online algorithm must satisfy

$$\rho \geq 1 + \min_{\Delta t}\max\left\{\frac{n\Delta t}{f(n)}, \frac{2f(n) - f(2n)}{f(2n) + n\Delta t}\right\}.$$

[5] This is assuming ALG tests the second batch of n samples as soon as they arrive at Δt. However, if an online algorithm waits for Δt time before testing the first batch, it will probably wait for Δt before testing the second batch as well, unless some form of arrival statistics are being used. In any case, we have $J^{\mathrm{ALG}} \geq 2f(n) + \Delta t$.

Observe that at $\Delta t = 0$, the second term in the above equation is greater. As Δt increases, the second term gets increasingly smaller and the first term (starting from $\frac{n\Delta t}{f(n)} = 0$ at $\Delta t = 0$), get increasingly larger. So the minima over Δt would occur when the two terms are equal, or when

$$\frac{n\Delta t}{f(n)} = \frac{2f(n) - f(2n)}{f(2n) + n\Delta t}.$$

This is a quadratic equation in Δt with only one positive root, solving which gives us

$$\rho \geq 1 + \frac{1}{2}\left(-\frac{f(2n)}{f(n)} + \sqrt{\frac{f^2(2n)}{f^2(n)} - 4\frac{f(2n)}{f(n)} + 8}\right). \tag{13}$$

This expression holds for all n, and so we choose the n which gives the largest expression to get the tightest lower bound. Claim 2(i) gives us $f(2n) \leq 2f(n)$, and in the range $[1, 2]$, (13) is strictly decreasing in $\frac{f(2n)}{f(n)}$, which gives us

$$\rho \geq 1 + \frac{1}{2}\left(\sqrt{\Gamma^2 - 4\Gamma + 8} - \Gamma\right),$$

where $\Gamma = \min_{x\in\mathbb{N}} \frac{f(2x)}{f(x)}$. ■

V. Conclusion & Future Steps

In this work, we give a competitive algorithm for group testing of samples that arrive over time. We consider the average wait time of the samples plus the average per-sample testing cost as the objective to minimize. We also give a lower bound on the competitive ratio that no online algorithm can beat.

However, there is a gap between the competitive ratio we guarantee for our algorithm and the lower bound. One avenue for future research is to close this gap: we could either (i) show that WTE has a better competitive ratio using more precise analysis; (ii) improve the lower bound by showing it is impossible for an online algorithm to achieve even some competitive ratios higher than guaranteed by Theorem 2; or (iii) develop new algorithms better than WTE and show an improved competitive ratio. Developing random algorithms that achieve a lower competitive ratio *in expectation* than the lower bound of Theorem 2 might also be pursued. We believe all these have much potential for future research.

Another direction is to study the offline problem: at time $t = 0$, assume we know when samples would arrive in the future. It is not implausible that the testing center might know when the samples might arrive if people have booked testing slots in advance, etc. While we can find the optimal schedule in this case using brute force by just computing every possible schedule and its corresponding cost, it remains to be seen if we can do this using a time complexity that is polynomial in the number of samples. If the problem indeed turns out to be NP-hard, it is of interest to know if we can develop approximation algorithms that are close to the optimal. While WTE can certainly be implemented in polynomial time in the offline scenario, it is to be seen if the knowledge of when future samples arrive can be used for developing algorithms with an even better approximation ratio than WTE.

REFERENCES

[1] R. Dorfman, "The detection of defective members of large populations," *Ann. Math. Stat.*, vol. 14, no. 4, pp. 436–440, Dec. 1943.

[2] C. L. Chan, S. Jaggi, V. Saligrama, and S. Agnihotri, "Non-adaptive group testing: Explicit bounds and novel algorithms," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 3019–3035, Mar. 2014.

[3] M. Aldridge, L. Baldassini, and O. Johnson, "Group testing algorithms: Bounds and simulations," *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3671–3687, Mar. 2014.

[4] J. Scarlett and O. Johnson, "Noisy non-adaptive group testing: A (near-) definite defectives approach," *IEEE Trans. Inf. Theory*, vol. 66, no. 6, pp. 3775–3797, Jan. 2020.

[5] M. Ruszinkó, "On the upper bound of the size of the r-cover-free families," *J. Comb. Theory, Ser. A*, vol. 66, no. 2, pp. 302–310, May 1994.

[6] L. Baldassini, O. Johnson, and M. Aldridge, "The capacity of adaptive group testing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2013, pp. 2676–2680.

[7] M. Aldridge, "Individual testing is optimal for nonadaptive group testing in the linear regime," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2058–2061, Oct. 2018.

[8] V. Gandikota, E. Grigorescu, S. Jaggi, and S. Zhou, "Nearly optimal sparse group testing," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 2760–2773, Jan. 2019.

[9] M. Aldridge, O. Johnson, and J. Scarlett, "Group testing: An information theory perspective," *Found. Trends Commun. Inf. Theory*, vol. 15, no. 3-4, pp. 196–392, Dec. 2019.

[10] B. Arasli and S. Ulukus, "Group testing with a graph infection spread model," arXiv:2101.05792 [cs.IT], Jan. 2021.

[11] P. Nikolopoulos, S. R. Srinivasavaradhan, T. Guo, C. Fragouli, and S. Diggavi, "Group testing for connected communities," in *Proc. Int. Conf. Artif. Intell. Stat. (AISTATS)*, Mar. 2021, pp. 2341–2349.

[12] ——, "Group testing for overlapping communities," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2021, pp. 1–7.

[13] S. Ahn, W. Chen, and A. Özgür, "Adaptive group testing on networks with community structure," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 1242–1247.

[14] M. Doger and S. Ulukus, "Dynamical Dorfman testing with quarantine," in *Proc. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2022.

[15] R. K. Deb and R. F. Serfozo, "Optimal control of batch service queues," *Adv. Appl. Probab.*, vol. 5, no. 2, pp. 340–361, Aug. 1973.

[16] M. F. Neuts and Y. Chandramouli, "Statistical group testing with queueing involved," *Queueing Syst.*, vol. 2, no. 1, pp. 19–39, Mar. 1987.

[17] S. K. Bar-Lev, M. Parlar, D. Perry, W. Stadje, and F. A. Van der Duyn Schouten, "Applications of bulk queues to group testing models with incomplete identification," *Eur. J. Oper. Res.*, vol. 183, no. 1, pp. 226–237, Nov. 2007.

[18] D. Claeys, J. Walraevens, K. Laevens, and H. Bruneel, "A queueing model for general group screening policies and dynamic item arrivals," *Eur. J. Oper. Res.*, vol. 207, no. 2, pp. 827–835, Dec. 2010.

[19] D. Claeys, B. Steyaert, J. Walraevens, K. Laevens, and H. Bruneel, "Tail probabilities of the delay in a batch-service queueing model with batch-size dependent service times and a timer mechanism," *Comput. Oper. Res.*, vol. 40, no. 5, pp. 1497–1505, May 2013.

[20] S. R. Chakravarthy, Shruti, and A. Rumyantsev, "Analysis of a queueing model with batch markovian arrival process and general distribution for group clearance," *Methodol. Comput. Appl. Probab.*, vol. 23, no. 4, pp. 1551–1579, Dec. 2021.

[21] S. R. Srinivasavaradhan, P. Nikolopoulos, C. Fragouli, and S. Diggavi, "An entropy reduction approach to continual testing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 611–616.

[22] D. Acemoglu, A. Fallah, A. Giometto, D. Huttenlocher, A. Ozdaglar, F. Parise, and S. Pattathil, "Optimal adaptive testing for epidemic control: combining molecular and serology tests," arXiv:2101.00773 [eess.SY], Jan. 2021.

[23] R. L. Graham, "Bounds for certain multiprocessing anomalies," *Bell Syst. Tech. J.*, vol. 45, no. 9, pp. 1563–1581, Nov. 1966.

[24] L. A. Hall, A. S. Schulz, D. B. Shmoys, and J. Wein, "Scheduling to minimize average completion time: Off-line and on-line approximation algorithms," *Math. Oper. Res.*, vol. 22, no. 3, pp. 513–544, Aug. 1997.

[25] S. Albers and H. Fujiwara, "Energy-efficient algorithms for flow time minimization," *ACM Trans. Algorithms*, vol. 3, no. 4, pp. 49–es, Nov. 2007.

[26] N. Bansal, K. Pruhs, and C. Stein, "Speed scaling for weighted flow time," *SIAM J. Comput.*, vol. 39, no. 4, pp. 1294–1308, Oct. 2010.

[27] N. R. Devanur and Z. Huang, "Primal dual gives almost optimal energy-efficient online algorithms," *ACM Trans. Algorithms*, vol. 14, no. 1, pp. 1–30, Jan. 2018.

[28] J. Edmonds, "Scheduling in the dark," *Theor. Comput. Sci.*, vol. 235, no. 1, pp. 109–141, Mar. 2000.

[29] K. Pruhs, "Competitive online scheduling for server systems," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 34, no. 4, pp. 52–58, Mar. 2007.