AdaGT: An Adaptive Group Testing Method for Improving Efficiency and Sensitivity of Large-Scale Screening Against COVID-19

Xiaofang Xia, *Member, IEEE*, Yang Liu, Yang Xiao, *Fellow, IEEE*, Jiangtao Cui, *Member, IEEE*, Bo Yang, *Member, IEEE*, and Yanguo Peng, *Member, IEEE*

***Abstract*— The ongoing coronavirus disease 2019 (COVID-19) is a pandemic causing millions of deaths, devastating social and economic disruptions. Testing individuals for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the pathogen of COVID-19, is critical for mitigating and containing COVID-19. Many countries are implementing group testing strategies against COVID-19 to improve testing capacity and efficiency while saving required workloads and consumables. A group of individuals' nasopharyngeal/oropharyngeal (NP/OP) swab samples is mixed to conduct one test. However, existing group testing methods neglect the fact that mixing samples usually leads to substantial dilution of viral ribonucleic acid (RNA) of SARS-CoV-2, which seriously impacts the sensitivity of tests. In this paper, we aim to screen individuals infected with COVID-19 with as few tests as possible, under the premise that the sensitivity of tests is high enough. To achieve this goal, we propose an Adaptive Group Testing (AdaGT) method. By collecting information on the number of positive and negative samples that have been identified during the screening process, the AdaGT method can estimate the ratio of positive samples in real-time. Based on this ratio, the AdaGT algorithm adjusts its testing strategy adaptively between an individual testing strategy and a group testing strategy. The group size of the group testing strategy is carefully selected to guarantee that the sensitivity of each test is higher than a predetermined threshold and that this group contains at most one positive sample on average. Theoretical performance analysis on the AdaGT algorithm is provided and then validated in experiments. Experimental results also show that the AdaGT algorithm outperforms existing methods in terms of efficiency and sensitivity.**

***Note to Practitioners*—Real-time reverse transcription-polymerase chain reaction (rRT-PCR) tests provide scope for automation and are one of the most widely used laboratory methods for detecting the SARS-CoV-2 virus. This paper is motivated by the following challenges: (1) Many countries are experiencing an acute shortage of professionals and consumables for conducting rRT-PCR tests; (2) Group sizes of existing group testing methods against COVID-19 may not be optimal, which adversely impacts the efficiency of the screening of the SARS-CoV-2 virus; (3) Existing group testing methods do not consider the fact that the sensitivity of rRT-PCR tests usually decreases with the group size. The objective of this paper is to improve the efficiency and sensitivity of large-scale screening against COVID-19. For achieving this goal, we propose an Adaptive Group Testing (AdaGT) algorithm, which has the following advantages: (1) It can improve the efficiency for screening the SARS-CoV-2 virus, mainly by adaptively adjusting its testing strategy between an individual testing strategy and a group testing strategy based upon an estimated ratio of positive samples during the screening process; (2) It can guarantee a high sensitivity of the rRT-PCR tests by determining the group sizes of the group testing strategy based upon some constraints; (3) We derive an appropriate threshold for the estimated ratio of positive samples such that the AdaGT algorithm can achieve a minimum average number of rRT-PCR tests and can be directly employed in practical applications.**

***Index Terms*—Searching algorithms, tree, binary tree searching, COVID-19, SARS-CoV-2, group testing, screening, sensitivity, efficiency.**

I. Introduction

THE coronavirus disease 2019 (COVID-19) is an infectious disease caused by a newly discovered virus named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1]. It was first detected in China and later characterized as a pandemic by World Health Organization (WHO) [2], [3]. As of 15 February 2021, COVID-19 is affecting more than 200 countries and territories globally [4], and the total number of confirmed cases amounts to nearly 110 million, including almost 2.4 million deaths [5]. Since half of the world's population is required by their governments to

Manuscript received June 26, 2021; revised October 3, 2021; accepted November 9, 2021. This article was recommended for publication by Associate Editor Prof. Jingshan Li and Editor C. Si upon evaluation of the reviewers' comments. The work of Xiaofang Xia, Yang Liu, Jiangtao Cui, Bo Yang, and Yanguo Peng was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61902299, Grant 61976168, and Grant 62172314; in part by the China Post-Doctoral Science Foundation under Grant 2019TQ0239 and Grant 2019M663636; in part by the Key Research and Development Plan of Shaanxi Province under Grant 2019ZDLGY13-09; in part by the Natural Science Basic Research Program of Shaanxi Province under Grant 2019CGXNG-023; in part by the S&T Program of Hebei under Grant 20310102D; and in part by the International Post-Doctoral Exchange Fellowship Program. *(Corresponding authors: Yang Xiao; Jiangtao Cui.)*

Xiaofang Xia, Yang Liu, Jiangtao Cui, and Yanguo Peng are with the School of Computer Science and Technology, Xidian University, Xi'an 710071, China (e-mail: xiaofangxia89@gmail.com; liuyang_work2012@163.com; cuijt@xidian.edu.cn; ygpeng@xidian.edu.cn).

Yang Xiao is with the Department of Computer Science, The University of Alabama, Tuscaloosa, AL 35487 USA (e-mail: yangxiao@ieee.org; yangxiao@cs.ua.edu).

Bo Yang is with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: yangboac@gmail.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TASE.2021.3131585.

Digital Object Identifier 10.1109/TASE.2021.3131585

stay at home to prevent the further spread of the deadly virus, the pandemic has caused devastating social and economic disruptions [6], [7]. It is reported that millions of enterprises face an existential threat, and nearly half of the world's 3.3 billion global workforces are at the risk of losing their livelihoods [7]. The pandemic is projected to cause around 9 trillion dollars of economic losses in 2020 and 2021 and has caused the worst global economic crisis since the Great Depression in 1929 [8].

Among all efforts to contain and mitigate the pandemic (such as social distancing and contact tracing), testing individuals for SARS-CoV-2 plays a critical role in helping isolate or hospitalize infected people and release uninfected people from quarantine. Specifically, individuals with positive results are isolated immediately to prevent further spread of the COVID-19 and provided with medical treatment at once to avoid exacerbation of the disease [9]. Furthermore, since their close contacts have a high risk of getting infected with the COVID-19, they are also tested quickly and treated appropriately. Besides, individuals with negative tests can resume normal daily activities such as working and schooling, which can maximally alleviate the economic disruption and inconvenience to people's life. Moreover, testing individuals for SARS-CoV-2 can also help investigators characterize the disease's prevalence, spread, and contagiousness [10].

Currently, real-time reverse transcription-polymerase chain reaction (rRT-PCR) is one of the most widely used laboratory methods for detecting the SARS-CoV-2 virus (we will explain how to conduct the rRT-PCR tests clinically later). Since asymptomatic and presymptomatic COVID-19 patients are common and are likely a major source of COVID-19 transmission, experts have recommended large-scale screening for COVID-19 infections using rRT-PCR tests [11]. However, many countries are experiencing an acute shortage of laboratories, trained personal, testing kits, and other consumables like tips, reagents, and bins for conducting rRT-PCR tests, limiting the testing capacity. For example, in the United States of America, in the week of December 9-16, 2020, 68.8% of laboratories on average suffer from the shortage of COVID-19 testing supplies [12]. In the Maharashtra state of India, due to the shortage of rRT-PCR kits, people have to wait at least two to three days for getting rRT-PCR tests [13].

To improve testing capacity and efficiency and save required workloads and consumables, many countries are implementing group testing strategies against COVID-19 [14], by which a group of individual nasopharyngeal/oropharyngeal (NP/OP) swab samples are grouped to get an rRT-PCR test. A negative result implies that all individuals in the group are clear. A positive result indicates that there is at least one infected person among this group of individuals. Existing group testing methods for screening the SARS-CoV-2 virus can be roughly categorized into hierarchical and non-hierarchical methods. The popular hierarchical methods consist of a certain number of stages, and in each stage, samples are tested once in non-overlapping groups. In contrast, non-hierarchical methods also involve testing over stages, but samples may be tested more than once per stage via overlapping pools [15]. Existing methods suffer from the following limitations:

1) When determining stage numbers, group sizes, and the set of samples to be tested, both categories of methods do not explain the rationality. Since little information of prior test results is leveraged, these parameters are usually not optimal, adversely impacting the screening efficiency.
2) Both categories of methods do not consider that mixing samples usually leads to substantial dilution of viral RNA in the grouped samples to impact the sensitivity of rRT-PCR tests seriously [16].

To address the above limitations, in this paper, we aim to screen the SARS-CoV-2 virus with as few rRT-PCR tests as possible, under the premise that the sensitivity of tests is high enough. To achieve this goal, we propose an Adaptive Group Testing (AdaGT) method. By collecting information on the number of both positive and negative samples that have been determined during the screening process, the AdaGT method can estimate the ratio of positive samples in real-time. Based on this ratio, the AdaGT method adaptively adjusts its testing strategy between an individual testing strategy and a group testing strategy. Specifically, if the estimated ratio is larger than a predetermined threshold, the AdaGT method applies an individual testing strategy by which the NP/OP swab samples are tested one by one. Otherwise, the AdaGT method applies a group testing strategy. The group size of the group testing strategy is carefully selected to guarantee that the sensitivity of each rRT-PCR test is higher than a predetermined threshold and that, on average, there is at most one positive sample among these samples. If an rRT-PCR test on a group of NP/OP swab samples gets a positive result, the AdaGT algorithm adopts a binary testing strategy to conduct more rRT-PCR tests on this group of samples until one positive sample is identified. The contributions of this paper are highlighted as follows:

- To the best of our knowledge, this is the first work of group testing methods to consider the sensitivity of rRT-PCR tests during the screening process against COVID-19;
- We propose an AdaGT algorithm which can improve the efficiency for screening the SARS-CoV-2 virus, mainly by adaptively adjusting its testing strategy between an individual testing strategy and a group testing strategy based upon an estimated ratio of positive samples during the screening process;
- The AdaGT algorithm can guarantee a high sensitivity of the rRT-PCR tests, mainly by choosing an appropriate group size for the group testing strategy;
- We provide performance analysis on the AdaGT algorithm, mainly including the minimum upper bound of the number of rRT-PCR tests and the selection of threshold parameter for the estimated ratio of positive samples during the screening process;
- Experiments are conducted to evaluate the performance of the AdaGT algorithm. Experimental results show that the AdaGT algorithm outperforms existing group testing methods against COVID-19 efficiency and sensitivity.

We organize the rest of this paper as follows. We introduce related works in Section II. We present the problem statement in Section III. We demonstrate the working strategy of the proposed AdaGT algorithm in Section IV. The performance analysis on the AdaGT algorithm is provided in Section V. The simulation results are reported in Section VI. We conclude this paper in Section VII.

II. Related Works

Group testing originates from World War II for syphilis screening [17]. Since then, it has been used to screen other infectious viruses such as hepatitis B/C and human immunodeficiency virus (HIV) [18]. Group testing methods have also been used for many other applications such as compressive sensing [19], multi-access channel management [20], electricity theft detection [21], [22], etc. Since the outbreak of the COVID-19 pandemic, researchers have done some works on developing efficient group testing methods to screen the SARS-CoV-2 virus, which, as aforementioned, can be roughly categorized into hierarchical and non-hierarchical methods [15].

Most of the existing methods belong to hierarchical methods, in which groups of samples tested in one stage are non-overlapping. For example, the authors in [23] present a binary testing protocol by which if the samples in a group are tested to be positive, half of the samples in this group are tested further. This process iterates until all individuals' statuses ("infected" or "not infected") are determined [23]. The authors in [24] propose a multi-stage group testing scheme, by which the initial group size is a power function whose exponent is equal to a predetermined total number of stages. If the rRT-PCR test on a group of samples gets a positive result, then this group is divided into non-overlapping smaller groups to get more rRT-PCR tests [24]. In [25], the authors employ a group testing strategy that switches the group size from eight samples to five samples when the prevalence rate increases from 0.5% to 6%. If an rRT-PCR test on a group of samples gets a positive result, then the samples in this group are retested individually [25]. The array testing algorithm in [15] is a typical non-hierarchical method, by which samples arranged in a two-dimensional grid are first tested by rows and by columns, respectively. Afterward, samples at intersections of positive rows and columns are retested separately [15].

However, in the above group testing methods, the (initial) group sizes or the total number of stages are decided in advance. As pointed out in [26], it is important to determine the group size before implementing the group testing strategies. However, the authors in [15], [23], [24] do not explain why it is reasonable to set them as these numbers. Also, when determining which samples to be tested in the next rRT-PCR test, these group testing methods consider the local information (e.g., results of current rRT-PCR tests), instead of the global information (e.g., results of all past rRT-PCR tests). This may result in determining which samples to test in the next rRT-PCR test may not be optimal. The above limitations adversely impact the efficiency of screening against COVID-19.

Furthermore, these group testing methods against COVID-19 do not consider that mixing samples usually leads to substantial dilution of viral RNA in the grouped samples. This seriously impacts the sensitivity of the rRT-PCR tests [16], where sensitivity is defined as the ratio of the number of positive samples that are correctly identified as being positive to the total number of positive samples. For example, the authors in [27] observe that when a single positive sample is mixed with 15 negative samples, the sensitivity is approximately 96%. When it is mixed with 31 negative samples, the sensitivity reduces to about 90%. Clinically, the cycle threshold is an important parameter for determining whether an NP/OP swab sample contains the SARS-CoV-2 virus or not (the definition is given out later). As indicated in [28], with the cutoff value of cycle threshold of a single sample and grouped samples being set as 35 and 40, respectively, the grouped positive samples have 100% sensitivity in group sizes 2, 4, and 6 and 97% $\sim$ 99% sensitivity in group sizes 8, 10, and 16. In [29], with the group size being set as 10, the authors have the following observations: (1) if the original sample has a high viral load, sample grouping does not affect the sensitivity of the assay; (2) however, for samples with a low initial viral load, the false-negative rate is about 13.3%, which means that the sensitivity (i.e., the true positive rate) reduces to about 86.7%.

In this paper, to address the above limitations, we propose the AdaGT method, which adaptively adjusts its testing strategy between an individual and a group testing strategy during the screening process. When the group testing strategy is applied, the group size is carefully selected such that the sensitivity of rRT-PCR tests is higher than a predetermined threshold and, on average, there is at most one positive sample among these samples for improving the efficiency.

III. Problem Statement

Real-time reverse transcription-polymerase chain reaction (rRT-PCR) tests are one of the most widely used laboratory methods for detecting the SARS-CoV-2 virus, mainly due to the following advantages: (1) easy to perform (2) have high sensitivity, (3) have more specificity, and (4) provide scope for automation. Clinically, the rRT-PCR tests are performed with the following steps: (1) Step 1: ribonucleic acid (RNA) extraction, in which several chemical solutions are used to extract the RNA present in the clinical nasopharyngeal/oropharyngeal (NP/OP) swab samples; (2) Step 2: reverse transcription, in which the extracted RNA is reversely transcribed to deoxyribonucleic acid (DNA) using a specific enzyme; (3) Step 3: polymerase chain reaction (PCR), in which a series of repeated temperature changes, called thermal cycles, are used to trigger specific chemical reactions on the products of Step 2 to amplify target sections of viral DNA [30].

For monitoring the progress of the PCR reaction in real-time, dyes emitting fluorescent signals are attached to new copies of the viral DNA sections [30]. With each successive cycle of amplification, the products of Step 3 double, and hence the fluorescence signal of DNA binding dyes increases [30]. Let C_t denote the cycle threshold, defined as the number of thermal cycles of chemical reactions required for the fluorescent signal of DNA binding dyes to exceed that of

the background [31]. Generally, a smaller C_t value indicates a higher viral load in the starting NP/OP swab samples, which indicates a more severe viral infection [32]. In applications, we usually set a cutoff value for C_t to help judge whether the NP/OP swab samples contain the SARS-CoV-2 virus. Specifically, if C_t does not exceed the cutoff value, the corresponding NP/OP swab samples are considered positive samples (i.e., containing the SARS-CoV-2 virus); otherwise, the corresponding NP/OP swab samples are considered to be negative samples [32].

In this paper, we have the following assumptions:

- There are a total number of n NP/OP swab samples, denoted by setting $N = \{x_1, x_2, \cdots, x_n\}$, where x_i denotes the i-th sample. We assume that among the n samples, m samples containing SARS-CoV-2 virus, where $m \in \{0, 1, \cdots, n\}$.
- We apply the rRT-PCR tests for detecting the SARS-CoV-2 virus. If an individual's NP/OP sample is tested to be negative, then this individual is considered to be not infected with COVID-19; otherwise, if an individual's NP/OP sample is tested to be positive, then this individual is considered to be infected with COVID-19.
- An rRT-PCR test that probes a group of samples containing/not containing the SARS-CoV-2 virus is likely to return a false negative/positive result mistakenly. As indicated in [33], false-negative results mainly occur through sample deficiency, concurrent respiratory infection and test inhibitors; and false-positive results mainly occur in erroneous testing and cross-reactions. More details about reasons causing false negative/positive results are analyzed in Section IV-B.

This paper aims to screen out the m positive samples with as few rRT-PCR tests as possible, under the premise that the sensitivity of the rRT-PCR tests is high enough.

IV. An Adaptive Group Testing Method

A. Overview

This section explains the working strategy of the proposed adaptive group testing (AdaGT) method for screening the SARS-CoV-2 virus. The basic idea of the AdaGT algorithm is to estimate the ratio of positive samples during the screening process. According to the estimated positive sample ratio, it then adaptively adjusts the testing strategy between an individual testing strategy and a group testing strategy. Particularly, when the group testing method is applied, we first analyze the maximum group size under some constraints (demonstrated later in subsection IV-B). Afterwards, the group size is determined to be of some special forms (demonstrated later in subsection IV-C). Then, we further apply different group testing strategies (demonstrated later in subsection IV-D) to screening samples. For better understanding, we depict a flow chart demonstrating the above working strategies of the AdaGT algorithm in Fig. 1.

We first demonstrate how to estimate the ratio of positive samples during the screening process. Let M and H denote the set of samples whose statuses have been determined as being "positive" and "negative" during the screening process,

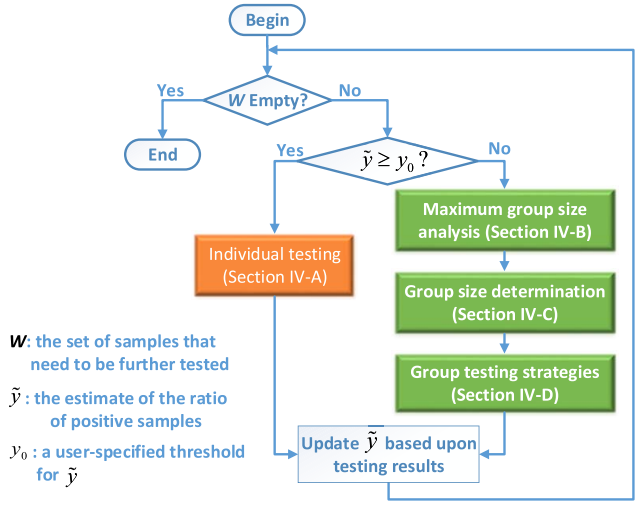


Fig. 1. A flow chart demonstrating the working strategies of the AdaGT algorithm.

respectively. Let W denote the set of samples to be further tested. Then, we have: $W = N - M - H$. Let $y (0 \leqslant y \leqslant 1)$ denote the ratio of positive samples to the total number of samples. y is usually unknown in practical applications. Since samples are randomly chosen for the rRT-PCR tests, we can roughly estimate y as the ratio of the number of samples in M to the total number of samples in both M and H. We have

$$\tilde{y} = \begin{cases} 0, & \text{if } M \cup H = \emptyset; \\ \frac{|M|}{|M|+|H|}, & \text{otherwise} \end{cases} \tag{1}$$

where $\tilde{y}$ denotes the estimate of y and $|\cdot|$ denotes the cardinality of a set.

Next, we demonstrate how the AdaGT algorithm adjusts the testing strategy between individual testing and a group testing strategy. Let $y_0 (0 \leqslant y_0 \leqslant 1)$ be a threshold parameter that is determined before the screening process. As shown in Fig. 2, if the estimated ratio $\tilde{y} \geqslant y_0$, we adopt the individual testing strategy, by which NP/OP swab samples are tested individually. If an rRT-PCR test gets a positive result, then the corresponding individual is infected with COVID-19; otherwise, this individual is not infected with COVID-19.

On the other hand, if $\tilde{y} < y_0$, we apply the group testing strategy, by which a group of NP/OP samples is mixed together to get one rRT-PCR test. If this rRT-PCR test gets a negative result, then all samples in this group are considered negative. Otherwise, there is at least one positive sample in this group, and more rRT-PCR tests should be conducted on these samples to find out the positive samples.

B. Maximum Group Size Analysis

When the group testing method is applied, we need first to determine the group size. In this subsection, we first analyze the maximum group size under some constraints to achieve this purpose.

Let s denote the number of the grouped samples examined in one rRT-PCR test. A positive rRT-PCR test is defined as probing a group of s samples containing the SARS-CoV-2

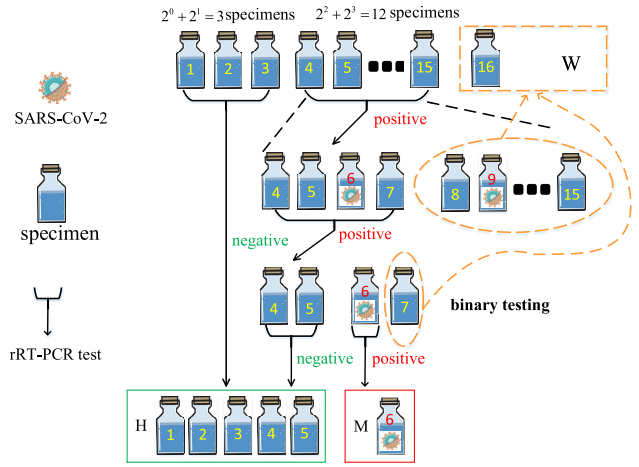


Fig. 2. An example to illustrate the AdaGT algorithm.

virus. A negative rRT-PCR test is defined as a test to probe a group of samples not containing the SARS-CoV-2 virus. We define sensitivity and specificity, as well as explain efficiency.

- *Sensitivity:* Let $f(s)$ $(0 \leqslant f(s) \leqslant 1)$ denote sensitivity which is defined as the probability that a positive rRT-PCR test correctly returns a positive result. Sensitivity is also called the true positive rate.
- *Specificity:* Let $h(s)$ $(0 \leqslant h(s) \leqslant 1)$ denote specificity which is the probability that a negative rRT-PCR test correctly returns a negative result. Specificity is also called the true negative rate.
- *Efficiency:* If an rRT-PCR test on a group of samples gets a negative result, all these samples can be identified as negative. This is the reason why group testing methods can help save rRT-PCR tests. Evidently, for achieving a high screening efficiency, we prefer that an rRT-PCR test with a negative result examine as many samples as possible. However, more samples examined by one rRT-PCR test imply a lower probability that this rRT-PCR test gets a negative result. Thus, to improve the screening efficiency, we should carefully select the group size s to balance between a large group size and a low probability for a negative result.

Mixing samples not containing the SARS-CoV-2 virus does not introduce the virus. This implies that if negative rRT-PCR tests are correctly performed according to the manufacture's instructions, before and after mixing samples, values of C_t exceed the cutoff value such that the rRT-PCR tests return negative results. In other words, before and after mixing samples, the specificity $h(s)$ does not change, regardless of the values of s. Thus, when determining the maximum group size, we do not involve the specificity of rRT-PCR tests.

For achieving high sensitivity, specificity, and efficiency for screening the SARS-CoV-2 virus using the rRT-PCR tests, in the AdaGT algorithm, we adopt the following three constraints to analyze the maximum group size of samples to be analyzed in the rRT-PCR tests during the screening process.

(1) *Non-Spillover Constraint*, s should not exceed the number of samples whose statuses have not been determined yet. Then, by the non-spillover constraint, we have

$$s \leqslant |W|, \quad \forall s \in \mathbf{N}^+, \tag{2}$$

where $\mathbf{N}^+$ denotes the set of positive integers.

(2) *Minimum Sensitivity Constraint*, by which we choose a group size s such that the sensitivity of rRT-PCR tests is not less than a pre-determined threshold, denoted as α_0, with $0 \leqslant \alpha_0 \leqslant 1$. For example, in applications, we can set α_0 as 0.95 or 0.9 [28]. As discussed in Section I, mixing samples usually leads to substantial dilution of viral RNA in the grouped samples. On the whole, the sensitivity of the rRT-PCR tests monotonically decreases with the group sizes [16], [27], [28]. Then, $f(s)$ is a decreasing function of s, and we have $f(s) \leq f(1) \leq 1$. By the minimum sensitivity constraint, we have $f(s) \geqslant \alpha_0$. Let s_{msc} denote the maximum group size satisfying the minimum sensitivity constraint. Then, we have

$$s \leqslant s_{msc} = \max\{s | \alpha_0 \leq f(s) \leq f(1) \leq 1, s \in \mathbf{N}^+\}, \tag{3}$$

where $\max(\cdot)$ returns the maximum value of a set.

(3) *One-Positive-Sample Constraint*, by which we choose a group size s such that, on average, there is at most one positive sample in this group to improve the method's efficiency. Technically, we have $sy \leqslant 1$, from which we can derive $s \leqslant \frac{1}{y}$. Let s_{ops} denote the maximum s satisfying the one-positive-sample constraint. Then, we have $s \leqslant s_{ops} = \left\lfloor \frac{1}{y} \right\rfloor$. If we know the ratio of positive samples, i.e., y, we can choose an appropriate s to satisfy the one-positive-sample constraint. However, as discussed early, in applications, we usually do not know the exact value of y, but can only roughly estimate it according to Equation (1). Thus, in applications, we have

$$s \leqslant s_{ops} = \begin{cases} +\infty, & \text{if } \tilde{y} = 0 \\ \left\lfloor \frac{1}{\tilde{y}} \right\rfloor, & \text{if } \tilde{y} > 0 \end{cases}, \quad \forall s \in \mathbf{N}^+, \tag{4}$$

where s_{ops} denotes the maximum s satisfying the one-positive-sample constraint.

Let s_{max} denote the maximum group size s that satisfies all the above three constraints. Then, we have $s_{max} \leqslant |W|$, $s_{max} \leqslant s_{msc}$ and $s_{max} \leqslant s_{ops}$, from which we can derive

$$s \leqslant s_{max} = \min\{s_{msc}, s_{ops}, |W|\}, \tag{5}$$

where $\min(\cdot)$ returns the minimum value of a set.

Next, we discuss false-negative and false-positive as follows.

- *False-Negative:* Sensitivity is related to false-negative rate since the summation of false-negative rate and sensitivity is 1. If a positive rRT-PCR test mistakenly returns a negative result, a false negative result occurs. Common reasons for false-negative results mainly include (1) inadequate laboratory rRT-PCR performance; (2) sample deficiency or degradation; (3) technical reasons relating to kit primers, probes, and fluorescence type; (4) SARS-CoV-2 mutations; and (5) RT-PCR inhibitors [33]. When samples are tested in groups, the viral RNA is diluted. According to the working principles of rRT-PCR tests in Section III, the dilution effect will increase the cycle threshold value (i.e., C_t). If C_t is increased to be a number greater

than the cutoff value, all the s samples, including those with the virus, are identified as negative samples. On the whole, a larger s value implies a more severe dilution and thereby a larger C_t. A larger C_t is more likely to exceed the cutoff value. Thus, we can conclude that under the dilution effect, the false-negative rate monotonically increases with the group size s. Therefore, the sensitivity $f(s)$ monotonically decreases with the group size s as mentioned before.

- *False-Positive:* Specificity is related to false-positive rate since the summation of false-positive rate and specificity is 1. Based upon reports released by the Centers for Disease Control and Prevention, the specificity of rRT-PCR tests is usually very high such that if rRT-PCR tests are performed strictly by manufacturer's instructions, false-positive results are almost impossible [34], [35]. However, in practical applications, false-positive results can still occur due to the following reasons: (1) inadequate laboratory rRT-PCR experience; (2) SARS-CoV-2 cross-contamination; (3) detection of unspecified coronaviruses; (4) SARS-CoV-2 inactive/residual detections; (5) cross-reaction with nucleic acids from other pathogens or tissue cells; and (6) technical reasons relating to kit primers, probes, and fluorescence type [33]. In this paper, we assume that the probability that an rRT-PCR test mistakenly returns a positive result is ε_1, where ε_1 is a small constant between 0 and 1. Since the summation of specificity and the false positive rate equals 1, we have $h(s) = 1 - \varepsilon_1$.

C. Group Size Determination

With the knowledge of the maximum group size under the above constraints, we demonstrate how to determine the group size during the screening process in this subsection.

To better understand, we first introduce a doubling strategy and a jumping strategy, which are commonly used to guide the group testing procedure. Both of the above two strategies greatly increase the group size if the results of previous rRT-PCR tests are negative, which helps to save the rRT-PCR tests for identifying negative samples. The basic idea of the doubling strategy is to double the previous group size every time. Specifically, by the doubling strategy, disjoint groups of $2^0, 2^1, 2^2, \cdots$ samples are probed sequentially until one rRT-PCR test returns a positive result [36]. Let s_{db} denote the group size, which has the same form as in the doubling strategy. Then, we have $s_{db} = 2^{k_2}$, with k_2 being a natural number. In contrast, the basic idea of the jumping strategy is to merge every two subsequent groups in the doubling strategy into one group. Specifically, by the jumping strategy, disjoint groups of $2^0 + 2^1, 2^2 + 2^3, 2^4 + 2^5, \cdots$ samples are probed until an rRT-PCR test returns a positive result [36]. Let s_{jp} denote a group size that has the same form as in the jumping strategy. Then, we have $s_{jp} = 2^{k_1} + 2^{k_1+1}$, with k_1 being an even natural number. In the following, we apply the jumping/doubling strategy if the group size is of the same form as in the jumping/doubling strategy.

We first demonstrate how to determine the group size in the case $\tilde{y} = 0$, where no samples have already been identified as positive. When there are no positive samples in a group of samples, the jumping strategy improves over the doubling strategy [37]. Thus, in the case $\tilde{y} = 0$, we apply the jumping strategy to guide the group testing procedure. Specifically, we initialize the value of k_1 as 0. If the rRT-PCR test returns a negative result, then the value of k_1 increases by 2. That is to say, at the j-th rRT-PCR test of the jumping strategy, we have $k_1 = 2j - 2$, with $j \in \mathbf{N}^+$. Since during the screening process, the maximum group size is s_{max}, we have $s_{jp} = 2^{k_1} + 2^{k_1+1} = 2^{2j-2} + 2^{2j-1} \leqslant s_{max}$, from which we can derive $k_1 \leqslant \lfloor \log_2 \frac{1}{3} s_{max} \rfloor$ and $j \leqslant \lfloor \frac{1}{2} \log_2 \frac{s_{max}}{3} \rfloor + 1$. This implies that the value of k_1 can increase to at most $\lfloor \log_2 \frac{1}{3} s_{max} \rfloor$. Besides, the group size increases until at most the $(\lfloor \frac{1}{2} \log_2 \frac{s_{max}}{3} \rfloor + 1)$-th rRT-PCR test.

We next demonstrate in the case $\tilde{y} = 0$ how to determine the group size after the $(\lfloor \frac{1}{2} \log_2 \frac{s_{max}}{3} \rfloor + 1)$-th rRT-PCR test (if any). In this case, for examining as many samples as possible with one rRT-PCR test, instead of sticking to either the doubling strategy or the jumping strategy as in [21], [22], we adaptively adjust the strategies between the jumping strategy and the doubling strategy such that the group sizes are as large as possible and do not exceed s_{max}. Specifically, we set $k_1 = \lfloor \log_2 \frac{1}{3} s_{max} \rfloor$, with which we can then obtain the maximum s_{jp} that does not exceed s_{max}. Afterwards, we set $k_2 = \lceil \log_2 s_{jp} \rceil = \lceil \log_2 (2^{k_1} + 2^{k_1+1}) \rceil = \lceil \log_2 3 \cdot 2^{k_1} \rceil = \lceil \log_2 3 + \log_2 2^{k_1} \rceil = k_1 + 2$, with which we can then obtain the smallest s_{db} that is larger than the above s_{jp}. Technically, we have $s_{db} = 2^{k_1+2} = 4 \cdot 2^{k_1} > s_{jp} = 2^{k_1} + 2^{k_1+1} = 3 \cdot 2^{k_1}$. If $s_{db} \leqslant s_{max}$, we set the group size $s = s_{db}$; otherwise, we set the group size $s = s_{jp}$. For example, in the case $s_{max} = 18$, we set $k_1 = \lfloor \log_2 \frac{1}{3} s_{max} \rfloor = 2$, according to which we can then calculate $s_{jp} = 2^2 + 2^3 = 12$. Afterwards, we set $k_2 = \lceil \log_2 12 \rceil = 4$, according to which we can calculate $s_{db} = 2^{k_2} = 16$. Since $s_{db} = 16 < s_{max} = 18$, we then set $s = 16$.

For the cases $\tilde{y} > 0$, we also determine the group size by adjusting the strategies between the jumping strategy and the doubling strategy adaptively as above.

Regardless of the case $\tilde{y} = 0$ or the case $\tilde{y} > 0$, we have $s_{jp} = 2^{k_1} + 2^{k_1+1} \geqslant 2^0 + 2^1 = 3$. Thus, the above strategies can only be used when $|W| \geqslant 3$. With regard to the case $|W| < 3$, we set the group size $s = s_{max}$. Evidently, in this case, s equals either 1 or 2.

D. Group Testing Strategies

After the group size, s, is determined, one rRT-PCR test is immediately performed on a group of s randomly chosen samples in W. If the result of the rRT-PCR test is "negative", then these s samples are identified as negative samples; otherwise, there is at least one positive sample among these s samples.

We first consider the case where the group size s has the same form as in the jumping strategy (i.e., $2^{k_1} + 2^{k_1+1}$). In this case, we perform another rRT-PCR test on a subset of 2^{k_1} samples of these samples, which are randomly chosen from

the $2^{k_1} + 2^{k_1+1}$ samples. The testing procedure proceeds as follows:

(1) If this rRT-PCR test gets a positive result, then among this subset of 2^{k_1} samples, there is at least one positive sample. In this case, we apply a binary testing strategy (whose working strategies are explained later) to find out at least one positive sample from this subset of 2^{k_1} samples. We do not know whether there are samples among the remaining 2^{k_1+1} samples containing the SARS-CoV-2 virus. Besides, the testing results of the 2^{k_1} samples lead to an update of the estimated ratio of positive samples (i.e., $\tilde{y}$), which further results in an update of the maximum group size satisfying the one-positive-sample constraint (i.e., s_{ops}). If s_{ops} is decreased to be smaller than 2^{k_1+1}, it violates the one-positive-sample constraint to perform the next rRT-PCR test on the remaining 2^{k_1+1} samples. As stated in subsection IV-C, the group size should be as close to (but not exceed) s_{max} as possible. Hence, if s_{ops} is increased to be much larger than 2^{k_1+1}, it is also not appropriate to perform the next rRT-PCR test on the remaining 2^{k_1+1} samples. In this paper, we update the group size for the next rRT-PCR test based upon the newly updated $\tilde{y}$, and the remaining 2^{k_1+1} samples are put back with the samples in W, waiting for further rRT-PCR tests shortly.

(2) Otherwise, if the rRT-PCR test on the subset of 2^{k_1} samples gets a negative result, we can infer that these 2^{k_1} samples are negative, and hence they are put into set H. We can also infer that there are positive samples among the remaining 2^{k_1+1} samples. Thus, we apply the binary testing strategy to find out at least one positive sample from the remaining 2^{k_1+1} samples. On the other hand, when the group size s has the same form as in the doubling strategy, we also apply the binary testing strategy to identify at least one positive sample.

The binary testing strategy proceeds as follows: (1) First, we perform one rRT-PCR test on half of the samples in this subset of 2^{k_1}, i.e., 2^{k_1-1} samples. (2) If this rRT-PCR test gets a positive result, we perform another rRT-PCR test on these 2^{k_1-1} samples, and the remaining 2^{k_1-1} samples are put back into W. (3) Otherwise, if this rRT-PCR test gets a negative result, the 2^{k_1-1} samples are identified as negative samples and are put into set H. Simultaneously, we perform another rRT-PCR test on the remaining 2^{k_1-1} samples. (4) The above processes (2) $\sim$ (3) repeat k_1 times. Particularly, at the (k_1-1)-th rRT-PCR test of the processes (2) $\sim$ (4), two samples are probed; and at the k_1-th rRT-PCR test of the processes (2) $\sim$ (4), only one sample is probed. If the k_1-th rRT-PCR test gets a positive result, this sample is identified as a positive sample, and the other sample examined at the (k_1-1)-th rRT-PCR test is put back into W. Otherwise, if the k_1-th rRT-PCR test gets a negative result, this sample is identified as a negative sample, and the other sample examined at the (k_1-1)-th rRT-PCR test is identified as a positive sample. To sum up, by the above binary testing strategy, if there is at least one positive sample among a total number of 2^{k_1} samples, we need to perform k_1+1 rRT-PCR tests to locate a positive sample. During this process, some negative samples may also be identified.

We conclude the above strategies in Algorithm 1, which is referred to as the Adaptive Group Testing (AdaGT) Algorithm.

Algorithm 1 Adaptive Group Testing (AdaGT)

Input: $N = \{1, 2, \ldots, n\}$, y_0
Output: M, H

$W \leftarrow N$, $M \leftarrow \emptyset$, $H \leftarrow \emptyset$, $\tilde{y} \leftarrow 0$; // $W, M, H, \tilde{y}$ are global variables
$k_1 \leftarrow 0$, $s_{msc} \leftarrow \max\{s | f(s) \geqslant \alpha_0, s \in \mathbf{N}^+\}$;
while $|W| > 0$ **do**
  if $\tilde{y} \geqslant y_0$ **then** // individual testing
    $X \leftarrow$ pop one sample j out of W;
    if rRT $-$ PCR(X) == *"positive"* **then**
      $M \leftarrow M \cup X$;
    else $H \leftarrow H \cup X$;
  else
    $s \leftarrow$ DetermineGroupSize $(\tilde{y}, s_{msc})$;
    $X \leftarrow$ pop s samples out of W;
    if rRT $-$ PCR(X) == *"positive"* **then**
      if $|X|\%3 == 0$ **then**
        $X' \leftarrow \frac{|X|}{3}$ samples from X;
        $X \leftarrow$ UpdateSet(X, X')
      while $|X| > 1$ **do** // binary testing
        $X' \leftarrow \frac{|X|}{2}$ samples from X;
        $X \leftarrow$ UpdateSet(X, X')
      $M \leftarrow M \cup X$;
    else $H \leftarrow H \cup X$;
  Update $\tilde{y}$ according to Equation (1);
Function DetermineGroupSize$(\tilde{y}, s_{msc})$**:**
  Update s_{ops} according to Equation (4);
  Update $s_{max} \leftarrow \min\{s_{msc}, s_{ops}, |W|\}$;
  if $|W| \geqslant 3$ **then**
    if $\tilde{y} == 0$ **then** // jumping strategy
      if $k_1 \leqslant \lfloor \log_2 \frac{1}{3} s_{max} \rfloor$ **then return** Jump (k_1);
      else return AdaJumpDouble (s_{max});
    else return AdaJumpDouble (s_{max}); // $\tilde{y} > 0$
  else return s_{max};
Function Jump(k_1)**:**
  $s_{jp} \leftarrow 2^{k_1} + 2^{k_1+1}$;
  $k_1 \leftarrow k_1 + 2$;
  return s_{jp};
Function AdaJumpDouble(s_{max})**:**
  $k_1 \leftarrow \lfloor \log_2 \frac{1}{3} s_{max} \rfloor$, $s_{jp} \leftarrow 2^{k_1} + 2^{k_1+1}$;
  $k_2 \leftarrow \lceil \log_2 s_{jp} \rceil$, $s_{db} \leftarrow 2^{k_2}$;
  if $s_{db} \leq s_{max}$ **then return** s_{db};
  else return s_{jp};
Function UpdateSet(X, X')**:**
  if rRT $-$ PCR(X') == *"positive"* **then**
    $W \leftarrow W \cup (X \setminus X')$, $X \leftarrow X'$;
  else $H \leftarrow H \cup X'$, $X \leftarrow X \setminus X'$;
  return X;

In Algorithm 1, we use X to denote a set of NP/OP swab samples to be tested by an rRT-PCR test and use X' to denote a subset of NP/OP swab samples in X. The function rRT-PCR represents performing one rRT-PCR test. The function DetermineGroupSize describes how to determine the group size. The function Jump describes how to apply the jumping strategy at the beginning of the screening process. The function AdaJumpDouble describes how to adaptively adjust the testing strategy between the jumping strategy and the doubling strategy. The function UpdateSet describes how to narrow down the search area.

E. Case Study

For better understanding, we take the example in Fig. 2 to illustrate how the AdaGT algorithm works. Note that in

Fig. 2, we assume $s_{msc} = 20$. As shown, for the first rRT-PCR test (when $\tilde{y} = 0$), the jumping strategy is applied to test $2^0 + 2^1 = 3$ samples. Since the rRT-PCR test returns a negative result, a total number of $2^2 + 2^3 = 12$ samples (i.e., samples $\{x_4, x_5, \cdots, x_{15}\}$ are tested at the second rRT-PCR test, which gets a positive result. Thus, at the third rRT-PCR test, four samples $\{x_4, x_5, x_6, x_7\}$ are examined. Since the third rRT-PCR test returns a positive result, there is at least one positive sample among the samples $\{x_4, x_5, x_6, x_7\}$, on which more rRT-PCR tests should be subsequently conducted by the binary testing strategy. For the remaining eight samples $\{x_8, x_9, \cdots, x_{15}\}$, they are put back with the sample 16 which remains in set W, waiting for further tests. At this time, W is updated as $\{x_{16}, x_8, x_9, \cdots, x_{15}\}$.

We next demonstrate how we apply the binary testing strategy to find out at least one positive sample from the samples $\{x_4, x_5, x_6, x_7\}$ after the third rRT-PCR test. Specifically, the fourth rRT-PCR test is performed on samples $\{x_4, x_5\}$. Since the fourth rRT-PCR test returns a "negative" result, samples $\{x_4, x_5\}$ are identified as negative samples and are put into H. We can also infer that there is at least one positive sample among the samples $\{x_6, x_7\}$. Next, we perform the fifth rRT-PCR test on the sample 6, which returns a "positive" result. Thus, we can identify sample 6 as a positive sample and then put it in to set M. At the same time, we put back the sample 7 into set W.

V. Performance Analysis

In this section, we provide a performance analysis of the AdaGT algorithm. Specifically, we first analyze the bounds of the number of rRT-PCR tests of the AdaGT algorithm. Then, we give out an appropriate selection of the parameter y_0, which is previously defined as a user-specified threshold for $\tilde{y}$ in Subsection IV-A.

A. Bounds of the Number of rRT-PCR Tests

Let $t(n, m)$ denote the number of rRT-PCR tests when we apply the AdaGT algorithm to screen out m positive samples among a total number of n samples.

Theorem 1: Assume that there are no positive samples among a total number of n samples. Then, we have $\left\lceil \frac{3(n+1)}{4s_{msc}} - \frac{3}{4} \right\rceil \leqslant t(n, 0) \leqslant \left\lfloor \frac{1}{2}\log_2 \frac{s_{msc}}{3} \right\rfloor + \left\lfloor \frac{2n}{s_{msc}} \right\rfloor + \left\lfloor \log_2 s_{msc} \right\rfloor$.

Proof: If there are no positive samples among a total number of n samples, then the estimated ratio $\tilde{y}$ is always equal to 0, which implies that $s_{ops} = +\infty$. This implies that in the case $\tilde{y} = 0$, we have $\min\{s_{msc}, s_{ops}, |W|\} = \min\{s_{msc}, |W|\}$. By substituting this into Inequality (5), we can derive $s \leqslant s_{max} = \min\{s_{msc}, |W|\} \leqslant s_{msc}$. Based upon how the group sizes are determined, the screening process can be divided into the following three stages:

(1) Stage I in which the jumping strategy and increases determine the group sizes after every rRT-PCR test. As discussed early, when the screening process begins, we examine disjoint groups of NP/OP swab samples of sizes $2^0 + 2^1, 2^2 + 2^3, \cdots, 2^{2j-2} + 2^{2j-1}$ at the first, second, $\cdots$, and the j-th rRT-PCR tests, respectively. Since $s \leqslant s_{max} = \min\{s_{msc}, |W|\} \leqslant s_{msc}$, we have $2^{2j-2} + 2^{2j-1} = 3 \cdot 2^{2j-2} \leqslant s_{msc}$, from which we can derive $0 \leqslant j \leqslant \left\lfloor \frac{1}{2}\log_2 \frac{s_{msc}}{3} \right\rfloor + 1$.

With the increase of the value of j, the value of $2^{2j-2} + 2^{2j-1}$ finally exceeds s_{max}. In this case, we determine the group size by adaptively adjusting the strategies between the jumping strategy and the doubling strategy. That is to say, the group size equals either $s_{jp} = 2^{k_1} + 2^{k_1+1}$ or $s_{db} = 2^{k_2}$, with $k_1 = \left\lfloor \log_2 \frac{s_{max}}{3} \right\rfloor$ and $k_2 = \left\lceil \log_2 s_{jp} \right\rceil = k_1 + 2$. Based upon the value of s_{max}, the screening process after Stage I can be divided into the following two stages:

(2) Stage II in which W contains at least s_{msc} samples, i.e., $|W| \geqslant s_{msc}$, and hence we have $s_{max} = \min\{s_{msc}, |W|\} = s_{msc}$. Let j' denote the number of rRT-PCR tests in Stage II. Since the total number of NP/OP swab samples examined in Stage I is $2^0 + 2^1 + \cdots + 2^{2j-2} + 2^{2j-1} = 4^j - 1$ and there is at least s_{msc} samples in W, in Stage II we examine at most $n - (4^j - 1) - s_{msc}$ NP/OP swab samples. Since the group size equals either $s_{jp} = 2^{k_1} + 2^{k_1+1} = 3 \cdot 2^{k_1}$ or $s_{db} = 2^{k_2} = 2^{k_1+2} = 4 \cdot 2^{k_1}$, we have

$$\left\lceil \frac{n - \left(4^j - 1\right) - s_{msc}}{4 \cdot 2^{k_1}} \right\rceil \leqslant j' \leqslant \left\lfloor \frac{n - \left(4^j - 1\right) - s_{msc}}{3 \cdot 2^{k_1}} \right\rfloor,$$

with $k_1 = \left\lfloor \log_2 \frac{s_{max}}{3} \right\rfloor = \left\lfloor \log_2 \frac{s_{msc}}{3} \right\rfloor$. Since $0 \leqslant j \leqslant \left\lfloor \frac{1}{2}\log_2 \frac{s_{msc}}{3} \right\rfloor + 1 \leq \log_4 \frac{s_{msc}}{3} + 1$,

we have

$$\begin{aligned}
&\left\lceil \frac{n - \left(4^j - 1\right) - s_{msc}}{4 \cdot 2^{k_1}} \right\rceil \\
&\quad \geqslant \left\lceil \frac{n + 1 - s_{msc} - 4^{1+\log_4 \frac{s_{msc}}{3}}}{4 \cdot 2^{\left\lfloor \log_2 \frac{s_{msc}}{3} \right\rfloor}} \right\rceil \\
&\quad \geqslant \left\lceil \frac{n + 1 - s_{msc} - 4^{1+\log_4 \frac{s_{msc}}{3}}}{4 \cdot 2^{\log_2 \frac{s_{msc}}{3}}} \right\rceil \\
&\quad = \left\lceil \frac{3(n+1) - s_{msc} - 4 \cdot \frac{s_{msc}}{3}}{4 \cdot \frac{s_{msc}}{3}} \right\rceil \\
&\quad = \left\lceil \frac{3(n+1) - 7s_{msc}}{4s_{msc}} \right\rceil \\
&\quad = \left\lceil \frac{3(n+1)}{4s_{msc}} - \frac{7}{4} \right\rceil \\
&\quad = \left\lceil \frac{3(n+1)}{4s_{msc}} - \frac{3}{4} \right\rceil - 1,
\end{aligned}$$

and

$$\begin{aligned}
&\left\lfloor \frac{n - \left(4^j - 1\right) - s_{msc}}{3 \cdot 2^{k_1}} \right\rfloor \\
&\quad \leqslant \left\lfloor \frac{n - s_{msc}}{3 \cdot 2^{\left\lfloor \log_2 \frac{s_{msc}}{3} \right\rfloor}} \right\rfloor \\
&\quad \leqslant \left\lfloor \frac{n - s_{msc}}{3 \cdot 2^{(\log_2 \frac{s_{msc}}{3}) - 1}} \right\rfloor \\
&\quad \leqslant \left\lfloor \frac{n - s_{msc}}{3 \cdot \frac{s_{msc}}{3} \cdot \frac{1}{2}} \right\rfloor \\
&\quad = \left\lfloor \frac{2(n - s_{msc})}{s_{msc}} \right\rfloor \\
&\quad = \left\lfloor \frac{2n}{s_{msc}} \right\rfloor - 2.
\end{aligned}$$

Thus, we have

$$\left\lceil \frac{3(n+1)}{4s_{msc}} - \frac{3}{4} \right\rceil - 4 \leqslant j' \leqslant \left\lfloor \frac{2n}{s_{msc}} \right\rfloor - 2.$$

(3) Stage III: With more NP/OP swab samples examined by the rRT-PCR tests, there are fewer samples in set W. In Stage III, we consider the case where the number of samples in W is less than s_{msc} samples, i.e., $|W| < s_{msc}$. We have $s_{max} = \min\{s_{msc}, |W|\} = |W|$. Thus, in this stage, the group size s is adaptively adjusted based upon the value of $|W|$. Specifically, the value of k_1 is adjusted as $k_1 = \left\lfloor \log_2 \frac{1}{3} s_{max} \right\rfloor = \left\lfloor \log_2 \frac{|W|}{3} \right\rfloor$. For one rRT-PCR test in Stage III, the minimum number of NP/OP swab samples to be examined is $s_{jp} = 2^{k_1} + 2^{k_1+1} = 3 \cdot 2^{k_1}$. Since $k_1 = \left\lfloor \log_2 \frac{|W|}{3} \right\rfloor \geqslant \log_2 \frac{|W|}{3} - 1$, we have $s_{jp} \geqslant 3 \cdot 2^{\log_2 \frac{|W|}{3} - 1} \geqslant \frac{|W|}{2}$. This means that every rRT-PCR test in Stage III probes at least one half of NP/OP swab samples remained in W. Let j'' denote the number of rRT-PCR tests in Stage III. Then, we have

$$1 \leqslant j'' \leqslant \lceil \log_2 |W| \rceil \leqslant \lceil \log_2 s_{msc} \rceil = \lfloor \log_2 s_{msc} \rfloor + 1.$$

Thus, combining Stages I, II, and III, we have $\left\lceil \frac{3(n+1)}{4s_{msc}} - \frac{3}{4} \right\rceil \leqslant t(n,0) = j + j' + j'' \leqslant \left\lfloor \frac{1}{2} \log_2 \frac{s_{msc}}{3} \right\rfloor + \left\lfloor \frac{2n}{s_{msc}} \right\rfloor + \lfloor \log_2 s_{msc} \rfloor$. This completes the proof. □

Theorem 2: Assume that there are m positive samples among a total number of n samples. Then, for $0 < m \leqslant \frac{n-1}{e}$ *where e is the natural constant, we have* $t(n,m) \leqslant \beta_0 + m \log_2 \frac{n-1}{m} + 1.42(m-1)$, *with* $\beta_0 = \left\lfloor \frac{1}{2} \log_2 \frac{s_{msc}}{3} \right\rfloor + \left\lfloor \frac{2n}{s_{msc}} \right\rfloor + \lfloor \log_2 s_{msc} \rfloor + \left\lfloor \log_2 \frac{s_{msc}}{3} \right\rfloor + 2$.

Proof: Based upon the value of $\tilde{y}$, the screening process by the proposed AdaGT Algorithm 1 can be divided into the following three procedures:

(1) Procedure 1: $\tilde{y} = 0$. This procedure involves at least one of the three stages in Theorem 1 and stops when the results of the rRT-PCR tests become positive. According to the proof analysis in Theorem 1, we can easily understand that for all rRT-PCR tests in the three stages, an rRT-PCR test in Stage II examines most NP/OP swab samples. This implies that compared to the cases where we get the first positive result at one rRT-PCR test in Stage I or Stage III, we need to perform more rRT-PCR tests to identify one positive sample when we get the first positive result at one rRT-PCR test in Stage II.

We next consider the case where we get the first positive result at one rRT-PCR test in Stage II; how many more rRT-PCR tests should be conducted to identify one positive sample. As discussed early, in this case, there are a total number of $s_{jp} = 2^{k_1} + 2^{k_1+1}$ or $s_{db} = 2^{k_2} = 2^{k_1+1} + 2^{k_1+1}$ samples, with $k_1 = \left\lfloor \log_2 \frac{s_{msc}}{3} \right\rfloor$. Without loss of generality, we next assume the group size is $s_{jp} = 2^{k_1} + 2^{k_1+1}$. By the AdaGT algorithm, after we find that there is at least one positive sample among these $2^{k_1} + 2^{k_1+1}$ samples, another rRT-PCR test is conducted to check whether there are positive samples among the first 2^{k_1} samples. (a) If this rRT-PCR test gets a positive result, then there is at least one positive sample among these 2^{k_1} samples. Consequently, a total of k_1 rRT-PCR tests are conducted on these samples to identify one positive sample. (b) Otherwise, if this rRT-PCR test gets a negative result, we can infer that there is at least a positive sample among the remaining 2^{k_1+1} samples. For finding out at least one positive sample from these 2^{k_1+1} samples, $k_1 + 1$ rRT-PCR tests are subsequently conducted. To sum up, if we perform one rRT-PCR test on $2^{k_1} + 2^{k_1+1}$ samples and it returns a positive result, we need to conduct $1 + k_1$ or $1 + (k_1 + 1) = k_1 + 2$ more rRT-PCR tests on this group of samples to find out at least one positive sample. Similarly, if a rRT-PCR test on a group of $s_{db} = 2^{k_2} = 2^{k_1+1} + 2^{k_1+1}$ samples returns a positive result, we need to conduct $k_1 + 2$ tests for finding out at least one positive sample. Let n' denote the number of negative samples identified before the first positive result. Then, the maximum number of rRT-PCR tests for identifying one positive sample in Procedure 1 is $t(n', 0) + 1 + (k_1 + 1) \leqslant t(n, 0) + k_1 + 2 \leqslant \left\lfloor \frac{1}{2} \log_2 \frac{s_{msc}}{3} \right\rfloor + \left\lfloor \frac{2n}{s_{msc}} \right\rfloor + \lfloor \log_2 s_{msc} \rfloor + \left\lfloor \log_2 \frac{s_{msc}}{3} \right\rfloor + 2$.

(2) Procedure 2: $0 < \tilde{y} \leqslant y_0$. In this procedure, the group size s is adjusted as either $s_{jp} = 2^{k_1} + 2^{k_1+1}$ or $s_{db} = 2^{k_2} = 2^{k_1+2}$, with $k_1 = \left\lfloor \log_2 \frac{s_{max}}{3} \right\rfloor$ and $s_{max} = \min\{s_{msc}, \frac{1}{\tilde{y}}, |W|\}$. Similar to the analysis in Procedure 1, if we perform one rRT-PCR test on this group of samples which returns a positive result, we need to conduct $k_1 + 1$ or $k_1 + 2$ more rRT-PCR tests on this group of samples to find out at least one positive sample. To sum up, for identifying one positive sample among $s_{jp} = 2^{k_1} + 2^{k_1+1}$ or $s_{db} = 2^{k_1+1} + 2^{k_1+1}$ samples, we need to conduct at most $1 + (k_1 + 2) = k_1 + 3$ rRT-PCR tests.

(3) Procedure 3: $\tilde{y} \geqslant y_0$. The NP/OP swab samples are tested individually.

Let m_i denote the total number of positive samples that are detected in all occurrences of Procedure $i (i = 1, 2, 3)$. We have $m_1 + m_2 + m_3 = m$, with $m_1 = 1$. Let $A_1, A_2, \cdots, A_{m_2}$ be the m_2 subsets that are tested in the corresponding m_2 occurrences of Procedure 2. Let $a_i = |A_i|$, $i = 1, 2, \cdots, m_2$. Clearly, for any A_i, we have $a_i = s_{jp} = 3 \cdot 2^{k_1}$ or $a_i = s_{db} = 4 \cdot 2^{k_1}$. Since $s_{jp} < s_{db}$, for obtaining the maximum number of rRT-PCR tests, we only need to consider the extreme case where all group sizes in Procedure 2 are of the form $s_{jp} = 3 \cdot 2^{k_1}$. In this case, to locate one malicious user from A_i, we conduct at most $\log_2 \frac{a_i}{3} + 3$ inspection steps. Through the above analyses, we can derive $t(n,m) \leqslant \beta_0 + \sum_{i=1}^{m_2} (\log_2 \frac{a_i}{3} + 3) + m_3 \leqslant \beta_0 + \sum_{i=1}^{m_2} (\log_2 a_i) + (3 - \log_2 3) m_2 + m_3$, where $\beta_0 = \left\lfloor \frac{1}{2} \log_2 \frac{s_{msc}}{3} \right\rfloor + \left\lfloor \frac{2n}{s_{msc}} \right\rfloor + \lfloor \log_2 s_{msc} \rfloor + \left\lfloor \log_2 \frac{s_{msc}}{3} \right\rfloor + 2$.

From the convexity of $\log_2(x)$, it follows that $\sum_{i=1}^{m_2} \log_2 a_i \leqslant m_2 \log_2 \frac{\sum_{i=1}^{m_2} a_i}{m_2} \leqslant m_2 \log_2 \frac{n - m_1 - m_3}{m_2} \leqslant m_2 \log_2 \frac{n-1}{m_2}$. Let $f(x) = x \log_2 \frac{n-1}{x}$. Then, we have $f'(x) = \log_2 \frac{n-1}{x} - \frac{1}{\ln 2}$, where $\ln(\cdot)$ denotes the logarithm with the base of the natural constant $e = 2.71828 \cdots$. When $x < 2^{-\frac{1}{\ln 2}} (n-1) = \frac{n-1}{e}$, the function $f(x)$ increases monotonically. Thus, when $m \leqslant \frac{1}{e}(n-1)$, we can derive from the above equation that $\sum_{i=1}^{m_2} \log_2 a_i \leqslant m_2 \log_2 \frac{n-1}{m_2} \leqslant m \log_2 \frac{n-1}{m}$. Substituting $\sum_{i=1}^{m_2} \log_2 a_i \leqslant m \log_2 \frac{n-1}{m}$ into the previous result $t(n,m) \leqslant \beta_0 + \sum_{i=1}^{m_2} (\log_2 a_i) + (3 - \log_2 3) m_2 + m_3$,, we can derive $t(n,m) \leqslant \beta_0 + \sum_{i=1}^{m_2} (\log_2 \frac{a_i}{3} + 3) + m_3 \leqslant \beta_0 + m \log_2 \frac{n-1}{m} +$

$(3-\log_2 3)(m_2+m_3) \leqslant \beta_0 + m\log_2\frac{n-1}{m} + 1.42(m-1)$. This completes the proof. □

Theorem 3: Assume that there are m positive samples among a total number of n samples. Then, for any $0 < m \leqslant n$, *we have* $t(n,m) \leqslant \beta_0 + \frac{\log_2 e}{e}(n-1) + 1.42(m-1)$, *where e is the natural constant and* $\beta_0 = \left\lfloor \frac{1}{2}\log_2\frac{s_{msc}}{3} \right\rfloor + \left\lfloor \frac{2n}{s_{msc}} \right\rfloor + \left\lfloor \log_2 s_{msc} \right\rfloor + \left\lfloor \log_2\frac{s_{msc}}{3} \right\rfloor + 2$.

Proof: From the analysis in the proof of Theorem 2, we can know that the function $f(x) = x\log_2\frac{n-1}{x}$ obtains the maximum value when $x = \frac{n-1}{e}$. Thus, from the inequality in Theorem 1, we can derive $t(n,m) \leqslant \beta_0 + m\log_2\frac{n-1}{m} + 1.42(m-1) \leqslant \beta_0 + \frac{\log_2 e}{e}(n-1) + 1.42(m-1)$. This completes the proof. □

B. Selection of Parameter y_0

As discussed in Section IV, on the whole, the AdaGT algorithm applies the following two testing strategies: (1) an individual testing strategy by which NP/OP swab samples are tested one by one; (2) a group testing strategy whereby a group of NP/OP swab samples are mixed and then tested with one rRT-PCR test. During the screening process, the value of y_0 determines which testing strategy is to be applied. Specifically, if $\tilde{y} \geqslant y_0$, the individual testing strategy is applied; otherwise, if $0 \leqslant \tilde{y} < y_0$, the group testing strategy is employed. We next discuss choosing the parameter y_0 such that the AdaGT algorithm can achieve the minimum average number of rRT-PCR tests.

Theorem 4: Assume that we apply the AdaGT algorithm to screen out m positive samples among a total number of n samples. Then, the average number of rRT-PCR tests achieves the minimum when $y_0 = \frac{1}{3}$.

Proof: We first calculate the average number of rRT-PCR tests to locate one positive NP/OP swab sample.

Case 1: the individual testing strategy is applied. If the estimated ratio of positive samples is $\tilde{y}$, $\tilde{y} \neq 0$, then on average we need to perform $\frac{1}{\tilde{y}}$ rRT-PCR tests to locate one positive NP/OP swab samples.

Case 2: the group testing strategy is applied. As discussed in Section IV-C, when $|W| > 3$, then the group size is either $s_{jp} = 2^{k_1} + 2^{k_1+1}$ or $s_{db} = 2^{k_2} = 2^{k_1+1} + 2^{k_1+1}$. As indicated in Procedure 2 of the proof in Theorem 2, for a group of $s_{jp} = 2^{k_1} + 2^{k_1+1}$ NP/OP swab samples which contain SARS-CoV-2 virus, if the rRT-PCR test gets on the first 2^{k_1} NP/OP swab samples gets a positive result, we need to conduct k_1 more rRT-PCR tests to locate one positive sample from the 2^{k_1} NP/OP swab samples; otherwise, we need to conduct more $k_1 + 1$ rRT-PCR tests to locate one positive sample from the remaining 2^{k_1+1} NP/OP swab samples. Since the NP/OP swab samples are randomly permutated, the average number of rRT-PCR tests to locate one infected NP/OP swab samples from $s_{jp} = 2^{k_1} + 2^{k_1+1}$ samples is $\frac{1}{3}(1+1+k_1) + \frac{2}{3}(1+1+k_1+1) = k_1 + 3$. Similarly, we can conclude that the average number of rRT-PCR tests to locate one infected NP/OP swab samples from $s_{db} = 2^{k_1+1} + 2^{k_1+1}$ samples is $\frac{1}{2}(1+1+k_1+1) + \frac{1}{2}(1+1+k_1+1) = k_1 + 3$.

For achieving a minimum average number of rRT-PCR tests, we can only apply the group testing strategy when its average number of rRT-PCR tests to locate one positive sample is less than that of the individual testing, i.e.,

$$k_1 + 3 = \left\lfloor \log_2\frac{1}{3}s_{max} \right\rfloor + 3 \leqslant \frac{1}{\tilde{y}}. \tag{6}$$

Assume that $|W|$ is large enough such that $s_{max} = \min\left\{s_{msc}, \left\lfloor\frac{1}{\tilde{y}}\right\rfloor, |W|\right\} = \min\left\{s_{msc}, \left\lfloor\frac{1}{\tilde{y}}\right\rfloor\right\}$. **Case A**: $s_{msc} \leqslant \left\lfloor\frac{1}{\tilde{y}}\right\rfloor$. Obviously, in this case, we have $s_{max} = s_{msc}$, which means that the Inequality (6) can be transformed as $\left\lfloor \log_2\frac{1}{3}s_{msc} \right\rfloor + 3 \leqslant \frac{1}{\tilde{y}}$. Thus, we can derive $\tilde{y} \leqslant \frac{1}{\left\lfloor \log_2\frac{1}{3}s_{msc} \right\rfloor + 3}$. This means that: (1) when $s_{msc} = 1$, we have $\tilde{y} \leqslant 1$; (2) when $s_{msc} = 2$, we have $\tilde{y} \leqslant \frac{1}{2}$; (3) when $s_{msc} \geqslant 3$, we have $\tilde{y} \leqslant \frac{1}{3}$. **Case B**: $s_{msc} > \left\lfloor\frac{1}{\tilde{y}}\right\rfloor$. In this case, we have $s_{max} = \left\lfloor\frac{1}{\tilde{y}}\right\rfloor$, and Inequality (6) can be transformed as $\left\lfloor \log_2\frac{1}{3}\left\lfloor\frac{1}{\tilde{y}}\right\rfloor \right\rfloor + 3 \leqslant \frac{1}{\tilde{y}}$, from which we can derive $\tilde{y} \leqslant \frac{1}{3}$. As discussed early, when $y \leqslant y_0$, the AdaGT algorithm applies the group testing. Thus, combining **Case A** and **Case B**, we should choose $y_0 = \frac{1}{3}$ such that the choice between the group testing and the individual testing can always lead to the minimum average number of rRT-PCR tests for identifying a positive sample, regardless of the ratio of positive samples. This completes the proof. □

VI. Experiments

In this section, we report the results of experiments, conducted in Python 3.8.3 on an integrated development environment platform - Jupyter Notebook 6.0.3. Note that each piece of data in the following figures is averaged over 50 times of repeated experiments.

A. Impacts of False Negative and False Positive

As discussed in subsection IV-B, the summation of the false-negative rate and the sensitivity (i.e., the true positive rate) is 1. Thus, the problem of investigating the impacts of false negative can be transformed into the problem of how the sensitivity of rRT-PCR tests impacts the performance of the proposed AdaGT algorithm. As discussed in Section I, mixing samples usually leads to substantial dilution of viral RNA in the grouped samples; and on the whole, the sensitivity of the rRT-PCR tests (i.e., $f(s)$) monotonically decreases with the group sizes size s [16], [27], [28]. Note that the specific form of $f(s)$ is not the scope of this paper. In the experiments, without loss of generality and easy implementation, we simply assume a linear relationship between the group size and the sensitivity. Specifically, in Fig. 4, for investigating the impacts of sensitivity on experimental results, we assume the following three relationships between the group size and the sensitivity:

$$f_1(s) = \begin{cases} -\frac{1}{400}s + 1; & \forall s \leqslant 16, s \in N^+ \\ -\frac{3}{800}s + 1.02; & \forall s > 16, s \in N^+, \end{cases} \tag{7}$$

$$f_2(s) = -\frac{1}{320}s + 1, \tag{8}$$

and

$$f_3(s) = -\frac{1}{640}s + 1. \tag{9}$$

We show curves of $f_1(s)$, $f_2(s)$ and $f_3(s)$ in Fig. 3.

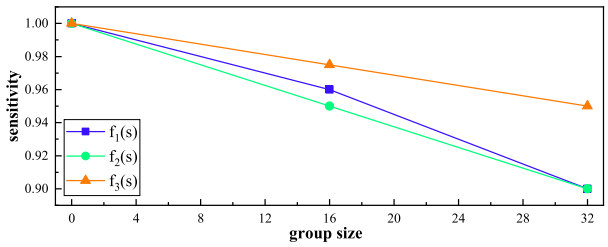


Fig. 3. Three linear relationships between the group size and the sensitivity.

The above Equation (7) is consistent with the observations in [27], introduced in Section II. In the following, if not otherwise specified, we assume $f(s) = f_1(s)$.

In Fig. 4, we assume that there are a total number of 10,000 NP/OP swab samples to be tested. We set $\alpha_0 = 0.9$ and $y_0 = 0.33$. As mentioned in Section IV-B, the specificity of an rRT-PCR test (i.e., $h(s)$) is the probability that a negative rRT-PCR test probing s samples correctly returns a negative result, and $h(s)$ does not change with the group size s. In the following, if not otherwise specified, we assume that the probability that a negative rRT-PCR test mistakenly returns a positive result is $\varepsilon_1 = 0.004$ due to reasons such as inadequate laboratory rRT-PCR tests. That is to say, in Fig. 4, we assume $h(s) = 1 - \varepsilon_1 = 0.096$.

As shown in the left of Fig. 4(a), when the ratio of positive samples y ranges from 0.005 to 0.1, the average numbers of rRT-PCR tests under relationships $f_1(s)$ and $f_2(s)$ almost coincide. This is mainly due to the following reasons: (a) when $\alpha_0 = 0.9$, the values of s_{msc} under both relationships $f_1(s)$ and $f_2(s)$ are equal to 32; (b) when $0 < y \leq 0.03$, we have $s_{ops} = \lfloor \frac{1}{y} \rfloor \geq \lfloor \frac{1}{0.03} \rfloor = 33 > 32$; thus, according to Equation (5), we can know that when $0 < y \leq 0.03$, we have $s_{max} = s_{msc} = 32$; (c) similarly, we can infer that when $0.03 < y \leq 0.1$, we have $s_{max} = s_{ops}$. Combining (b) and (c), we can know that when $y < 0.1$, the values of s_{max} under relationships $f_1(s)$ and $f_2(s)$ are equal, resulting in an almost equal average number of rRT-PCR tests.

We can also infer that when $y \leq 0.015$, we have $s_{max} = 64$ under the relationship $f_3(s)$. It is larger than the value of s_{max} under relationships $f_1(s)$ and $f_2(s)$ (i.e., 32). This leads that the average number of rRT-PCR tests under relationship $f_3(s)$ is slightly smaller than those under relationships $f_1(s)$ and $f_2(s)$, as shown in the right of Fig. 4(a). When $y \geq 0.021$, we have $s_{ops} = \lfloor \frac{1}{y} \rfloor \leq \lfloor \frac{1}{0.021} \rfloor = 47$. As aforementioned, the group size s is of the same form as in the jumping/doubling strategy. Thus, when $0.021 \leq y \leq 0.03$, we can choose the group size s to be at most 32 under relationships $f_1(s)$, $f_2(s)$, and $f_3(s)$. As shown in Fig. 3, when $s = 32$, the probability that an rRT-PCR test correctly returns a positive result under relationship $f_3(s)$ is much higher than those under relationships $f_1(s)$ and $f_2(s)$. This implies that more positive results will be returned under relationship $f_3(s)$. Thus, when $0.021 \leq y \leq 0.03$, the average number of rRT-PCR tests under relationship $f_3(s)$ is slightly more than those under relationships $f_1(s)$ and $f_2(s)$, as shown in the right of Fig. 4(a).

In overall settings, the sensitivity is calculated as the ratio of the number of positive samples correctly identified to the total number of positive samples. As shown in Fig. 4(b), for a given ratio of positive samples, the sensitivity usually achieves the highest value under relationship $f_3(s)$, followed by relationship $f_1(s)$, and achieves the lowest value under relationship $f_2(s)$, which is consistent with Fig. 3. For any given relationship among $f_1(s)$, $f_2(s)$, and $f_3(s)$, the sensitivity has a tendency to increase with the ratio of positive samples.

In overall settings, the specificity is calculated as the ratio of the number of negative samples correctly identified to the total number of negative samples. As shown in Fig. 4(c), for any of the above three relationships, the specificity is very high. Specifically, it is larger than 99.8%. For any given relationship among $f_1(s)$, $f_2(s)$, and $f_3(s)$, the specificity has a tendency to decline with the ratio of positive samples.

We next investigate the impacts of false positive. As aforementioned, we use ε_1 to denote the probability that an rRT-PCR test mistakenly returns a positive result. Thus, in Fig. 5, we investigate the performance of the AdaGT algorithm when the value of ε_1 increases from 0.002 to 0.02. We set $\alpha_0 = 0.9$ and $y_0 = 0.33$. As shown in Fig. 5(a), with ε_1 increasing, the average number of rRT-PCR tests increases slowly, regardless of the value of the ratio of positive samples y. In Fig. 5(b), we can observe that for any given y, with ε_1 increasing, the sensitivity tends to decline. For any given ε_1, the AdaGT algorithm usually achieves the highest sensitivity under $y = 0.04$, followed by $y = 0.03$, and achieves the lowest sensitivity under $y = 0.02$. In Fig. 5(c), we can observe that for any given y, with ε_1 increasing, the specificity tends to decline. As shown in Fig. 5(c), the AdaGT algorithm usually achieves the highest specificity under $y = 0.02$. When $\varepsilon_1 < 0.06$, the specificity under $y = 0.04$ is larger than that under $y = 0.03$; and vice versa, when $\varepsilon_1 > 0.06$.

B. Selection of Parameters α_0 and y_0

As defined earlier in subsection IV-B, the parameter α_0 is a pre-determined threshold that the sensitivity of rRT-PCR tests should not be less than it. In Fig. 6, we investigate how different values of parameter α_0 impact the performance of the AdaGT algorithm in terms of the average number of rRT-PCR tests and the sensitivity. We assume that there are a total number of 10,000 NP/OP swab samples to be tested. Although in the real world, the value of α_0 can be set by medical professionals as any value between 0 and 1, in our experiments, without loss of generality, we set α_0 as 0.9 and 0.95, respectively. The ratio of positive samples, denoted by y, ranges from 0.005 to 0.07. As indicated in Theorem 4, the AdaGT algorithm achieves the minimum average number of rRT-PCR tests when $y_0 = \frac{1}{3}$. Thus, in Fig. 6, the threshold for the estimated ratio of positive samples, i.e., y_0, is set as 0.33.

According to Equations (3) and (7), in the case $\alpha_0 = 0.95$, we have $s_{msc} = 18$; and in the case $\alpha_0 = 0.9$, we have $s_{msc} = 32$. Thus, when $y < 0.03$, in both of the above cases, we usually have $s_{ops} > s_{msc}$. According to the analysis in Subsection IV-C, when $s_{ops} > s_{msc}$, it is s_{msc} that determines the maximum group size in the group testing strategy. Since s_{msc} in the case $\alpha_0 = 0.9$ is greater than s_{msc} in the case $\alpha_0 = 0.95$, we can infer that the maximum group size in the

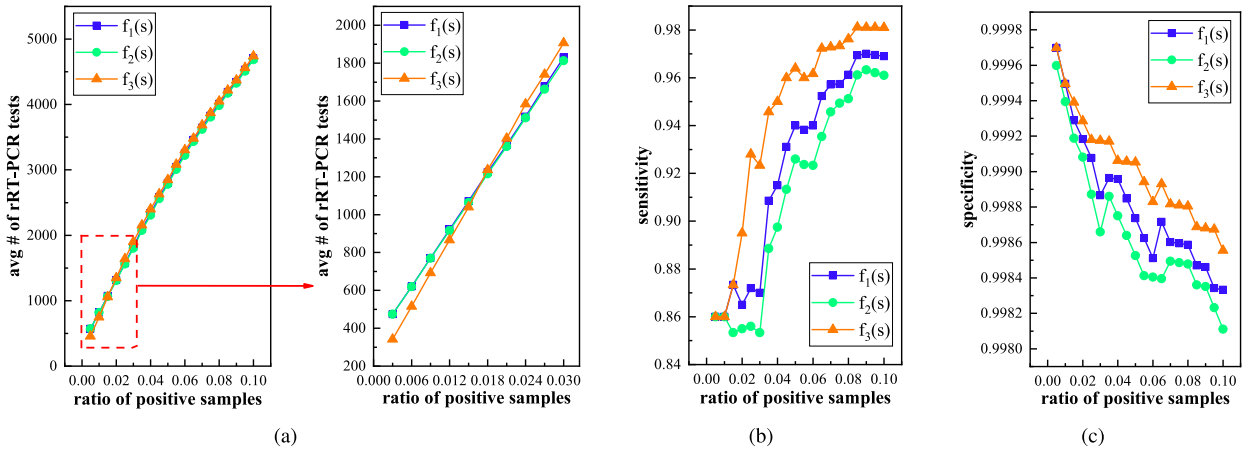


Fig. 4. Experiment results under three different relationships $f_1(s)$, $f_2(s)$, and $f_3(s)$ when the ratio of positive samples ranges from 0.005 to 0.1.

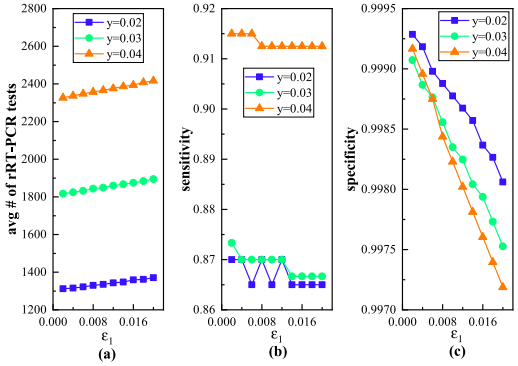


Fig. 5. Experiment results under the relationship $f_1(s)$ when ε_1 ranges from 0.002 to 0.02. Note that ε_1 denotes the probability that an rRT-PCR test mistakenly returns a positive result.

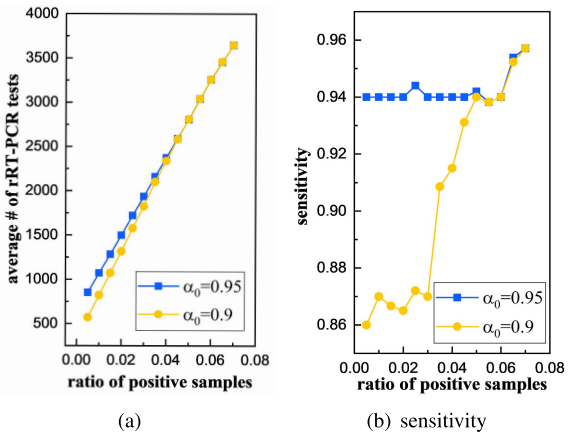


Fig. 6. Experiment results under different values of α_0: (a) average number of rRT-PCR tests; (b) sensitivity.

case $\alpha_0 = 0.9$ is greater than that in the case $\alpha_0 = 0.95$. When $0.03 < y < 0.05$, in the case $\alpha_0 = 0.95$ we usually have $s_{ops} > s_{msc} = 18$, which means the maximum group size is less than 18. In contrast, in the case $\alpha_0 = 0.9$ we usually have $20 < s_{ops} < s_{msc} = 32$, which means that the maximum group size is larger than 20. Thus, when $0.03 < y < 0.05$, the maximum group size in the case $\alpha_0 = 0.9$ is larger than that in the case $\alpha_0 = 0.95$. Since a larger maximum group size, on the whole, implies more samples examined in an rRT-PCR test, we can derive that when $y < 0.05$, on the whole, an rRT-PCR test examines more NP/OP samples in the case $\alpha_0 = 0.9$ than in the case $\alpha_0 = 0.95$. Thus, when $y < 0.05$, fewer rRT-PCR tests are conducted in the case $\alpha_0 = 0.9$ than in the case $\alpha_0 = 0.95$. The above analysis is validated in Fig. 6(a). As can be seen, when $y < 0.05$, the average number of rRT-PCR tests in the case $\alpha_0 = 0.95$ is larger than that in the case $\alpha_0 = 0.9$. Furthermore, with the increase of y, the difference between the average rRT-PCR tests in these two cases gets smaller constantly.

When $y > 0.05$, in both of the above cases, we have $s_{ops} < s_{msc}$, which implies that in both cases, it is the estimated ratio of positive samples that determines the maximum group size. For a given y, its estimated value is not impacted by values of α_0. Thus, for a given ratio of positive samples, the maximum group size in both of the above two cases is almost the same. This implies that the average number of rRT-PCR tests conducted in both of the above two cases is almost the same. The above analysis is also validated in Fig. 6(a). As can be seen, when $y > 0.05$, the two curves of the average numbers of rRT-PCR tests under the above two cases almost coincide with each other.

In Fig. 6(b), we have the following observations: when $y < 0.05$, the sensitivity in the case $\alpha_0 = 0.9$ is less than that in the case $\alpha_0 = 0.95$. Besides, with the increase of the ratio of positive samples, the gap between values of sensitivity in the above two cases narrows, and finally, the two curves almost overlap with each other. The reason behind this phenomenon is analyzed as follows: (1) As discussed early, when $y < 0.05$, on the whole, an rRT-PCR test examines more NP/OP samples in the case $\alpha_0 = 0.9$ than in the case $\alpha_0 = 0.95$; (2) The sensitivity of the rRT-PCR tests is seriously impacted by the dilution effect. In other words, when a positive sample is diluted in a group of negative samples, the sensitivity of rRT-PCR tests lowers substantially. (3) On the whole, more rRT-PCR tests intuitively imply smaller sensitivity. Nevertheless, in the experiments, the differences between rRT-PCR tests in the case $\alpha_0 = 0.9$ and $\alpha_0 = 0.95$ are not large enough to compensate for the dilution effect of the group size. Note that if it is not otherwise stated in the following, we set $\alpha_0 = 0.9$.

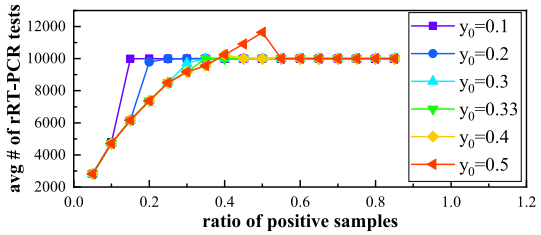


Fig. 7. Experiment results under different values of y_0 in terms of the average number of rRT-PCR tests.

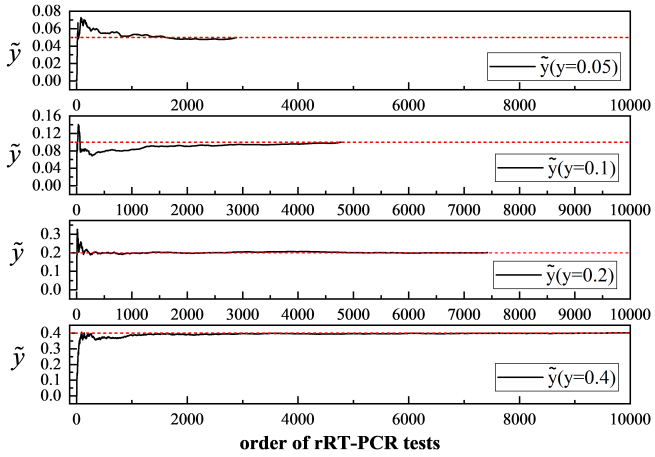


Fig. 8. Experiment results regarding how the value of $\tilde{y}$ changes during the screening process.

In Fig. 7, we investigate how different values of parameter y_0 impact the performance of the AdaGT algorithm in terms of the average number of rRT-PCR tests. We also assume that the total number of NP/OP swab samples to be tested is 10,000, i.e., $n = 10,000$. As shown, for any given y_0, when the ratio of positive samples does not exceed y_0, the average number of rRT-PCR tests increases monotonically. Particularly, when $y_0 \leqslant 0.33$, the average number of rRT-PCR tests first grows with the increase of the ratio of positive samples and then remains stable. When $y_0 > 0.33$, the average number of rRT-PCR tests first increases to a number larger than $10,000$ and then declines to $10,000$. Regardless of the ratio of positive samples, the average number of rRT-PCR tests achieves the minimum in the case $y_0 = 0.33$. This validates Theorem 4. In the following, if it is not other stated, we set $y_0 = 0.33$.

C. Estimation Accuracy of y

In Fig. 8 and Fig. 9, we investigate the following two issues, respectively: (1) how does the estimation accuracy of y change during the screening process? (2) how does the estimation accuracy of y impact the performance of the proposed AdaGT algorithm in terms of the average number of rRT-PCR tests. In the experiments, we set $\alpha_0 = 0.9$, $y_0 = 0.33$, and $\varepsilon_1 = 0.04$.

In Fig. 8, we consider four cases where the ratio of positive samples (i.e., y) is assumed to be 0.05, 0.1, 0.2, and 0.4, respectively. As shown in the figure, in all of the above cases, as the screening process proceeds, the bias between $\tilde{y}$ and y tends to become smaller. The reasons behind this phenomenon are analyzed as follows: (1) When the screening process begins, the estimation accuracy is low due to the limited information. (2) With more and more rRT-PCR tests being performed, the knowledge about negative or positive samples increases, leading to an improvement in the estimation accuracy.

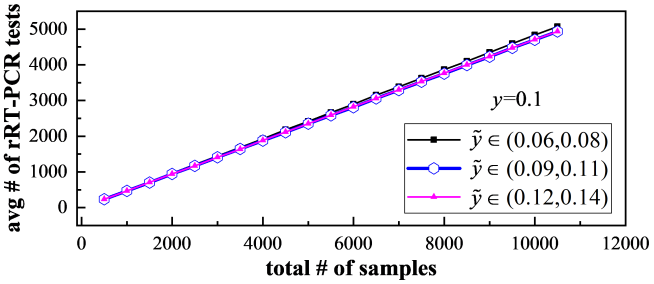


Fig. 9. Experiment results regarding how estimation accuracy of y impacts the performance of the proposed AdaGT algorithm in terms of the average number of rRT-PCR tests.

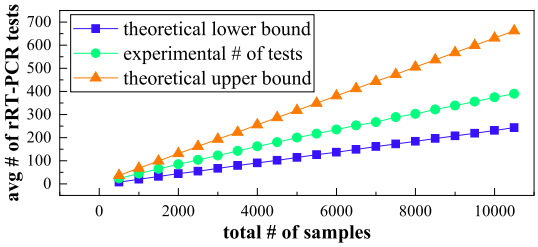


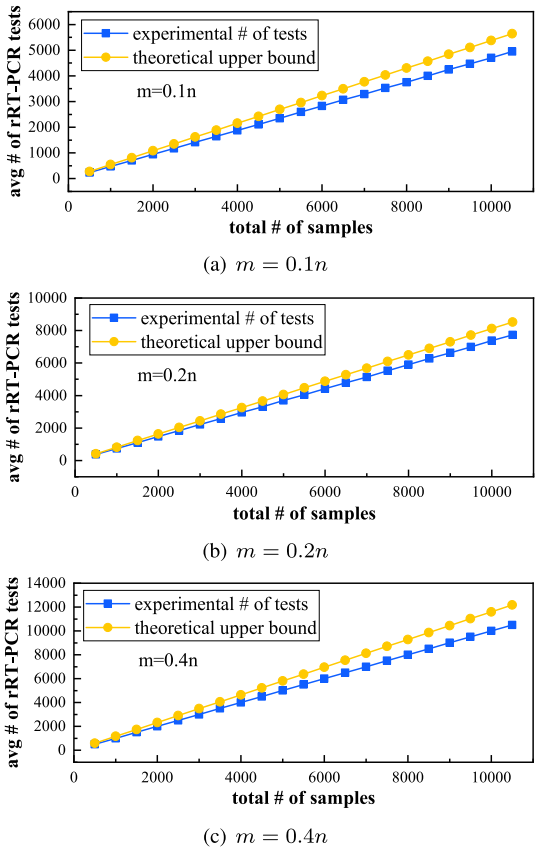
Fig. 10. Experiment results under $m = 0$.

In Fig. 9, we assume that the ratio of positive samples is 0.1. For investigating how estimation accuracy of y impacts the performance of the proposed AdaGT algorithm in terms of the average number of rRT-PCR tests, we consider three cases where $\tilde{y}$ is estimated to be valued in the intervals $(0.06, 0.08)$, $(0.09, 0.11)$, and $(0.12, 0.14)$, respectively, during the screening process. As shown in the figure, with the total number of samples to be tested ranging from 500 to 10500, the average number of rRT-PCR tests does not differ much in the above three cases. To conclude, although the estimation accuracy of y is low at the beginning of the screening process, this does not deteriorate the screening efficiency. Thus, it is reasonable to use the estimated $\tilde{y}$ at the beginning of the screening process, and it is not necessary to use the estimated $\tilde{y}$ after enough rRT-PCR tests have been done.

D. Bounds of Number of rRT-PCR Tests

In Fig. 10, we investigate how the average number of rRT-PCR tests changes when there are no positive samples among a total number of n samples. As shown, with n ranging from 1000 to 11000, the average number of rRT-PCR tests increases monotonically. However, it is always between the theoretical lower bound, and upper bound given out in Theorem 1. This validates Theorem 1.

In Fig. 11, we investigate how the average number of rRT-PCR tests changes when the ratio of positive samples is 0.1, 0.2, and 0.4, respectively. The theoretical upper bounds in Fig. 11(a) and Fig. 11(b) are calculated based upon Theorem 2, and the theoretical upper bound in Fig, 11(c) is calculated based upon Theorem 3. As shown, in all of the above three cases, with n ranging from 1000 to 11000, the



(a) $m = 0.1n$

(b) $m = 0.2n$

(c) $m = 0.4n$

Fig. 11. Experiment results: (a) $m = 0.1n$; (b) $m = 0.2n$; (c) $m = 0.4n$.

average number of rRT-PCR tests increases monotonically. Still, it never exceeds the theoretical upper bounds given out in Theorem 2 or Theorem 3. This validates the correctness of Theorem 2 and Theorem 3.

E. Comparison With Existing Schemes

In this subsection, we assume that there are a total number of 10,000 Np/OP samples to be tested. We compare the proposed AdaGT algorithm with the following group testing methods against COVID-19 in [23], [24]: (1) The BT32 method means the binary testing protocol in [23], by which at the first rRT-PCR test, a group of 32 samples is mixed together to be examined. If the result of an rRT-PCR test is negative, then all samples being tested are clear; otherwise, half of these samples are further tested in the next rRT-PCR test. (2) Both of the P9S3 and P4S2 methods are multi-stage group testing methods in [24]: (2a) Specifically, the P9S3 method is performed with three stages as follows: At the first stage, the NP/OP samples are divided into groups of 9 samples. For the tested positive groups at the first stage, their elements are further divided into groups of 3 samples at the second stage. For the groups that tested positive at the second stage, their elements are tested individually at the third stage. (2b) The P4S2 method is performed with two stages as follows: At the first stage, the NP/OP samples are divided into groups of 4 samples. For the groups which are tested positive, their elements are further tested individually at the second stage.

In Fig. 12, the ratio of positive samples ranges from 0.005 to 0.1. Particularly, for the AdaGT algorithm, we consider the following two cases $\alpha_0 = 0.9$ and $\alpha_0 = 0.95$. As shown at the left-side of Fig. 13(a), with the increasing ratio of positive samples, the average numbers of the rRT-PCR tests of the AdaGT, the BT32, the P9S3, and the P4S2 methods in [23] increase monotonically. As shown at the right-side of Fig. 13(a), when the ratio of positive samples is less than 0.05, both the AdaGT ($\alpha_0 = 0.9$) and the BT32 algorithms achieve the minimum average number of the rRT-PCR tests, followed by the AdaGT ($\alpha_0 = 0.95$). In contrast, on average, the P9S3 and the P4S2 algorithms conduct much more rRT-PCR tests than the proposed AdaGT algorithm with both $\alpha_0 = 0.9$ and $\alpha_0 = 0.95$. As shown at the right-side of Fig. 13(a), when the ratio of positive samples is larger than 0.05, the AdaGT algorithm conducts the fewest rRT-PCR tests. From the above observations, we can conclude that when the ratio of positive samples ranges from 0.005 to 0.1, the proposed AdaGT algorithm outperforms the P9S3, the P4S2, and the BT32 methods in terms of screening efficiency.

In Fig. 12(b), we have the following observations: (1) The P4S2 method achieves the greatest sensitivity, followed by the P9S3 method. (2) When the ratio of positive samples is less than 0.035, the AdaGT ($\alpha_0 = 0.9$) and the BT32 methods have comparable sensitivity. (3) When the ratio of positive samples is larger than 0.035, the sensitivity of the AdaGT method ($\alpha_0 = 0.9$) increases quickly. When the ratio of positive samples is larger than 0.07, the AdaGT method has a comparable sensitivity with the P9S3 method. When α_0 is set as 0.9, the lowest sensitivity of the proposed AdaGT algorithm is about 0.87. In contrast, when α_0 is set as 0.95, the lowest sensitivity of the proposed AdaGT algorithm is about 0.94. This implies that we can control the sensitivity of the proposed AdaGT algorithm by adjusting the value of α_0. Particularly, when α_0 is set as 0.95, the sensitivity of the proposed AdaGT algorithm is distributed between 0.94 and 0.97, which is comparable to those of the P9S3 and the P4S2 algorithms.

In Fig. 12(c), we investigate how the specificity of all rRT-PCR tests changes with the ratio of positive samples. As shown in Fig. 12, for a given ratio of positive samples, the specificity values of the P9S3 and P4S2 methods are much higher than those of the AdaGT and BT32 methods. This is mainly because the P9S3 and P4S2 methods do not infer samples' states ("positive" or "negative") through other samples' states, whereas both the AdaGT and BT32 methods indeed do this. In addition, for any given method among the above four methods, as the ratio of positive samples increases, the specificity value tends to decrease monotonically. On the whole, the specificity values of all the above algorithms are very high.

For better understanding, in Fig. 13, we compare the proposed AdaGT algorithm with the BT32, the P9S3, and the P4S2 algorithms when the ratio of positive samples ranges from 0.05 to 0.9. From the above analysis, we know that when the ratio of positive samples is larger than 0.05, the curves of the average number of rRT-PCR tests conducted by the AdaGT algorithm under $\alpha_0 = 0.9$ and $\alpha_0 = 0.95$ coincide with each other. Thus, in Fig. 13, we only consider the case $\alpha_0 = 0.9$. As shown in Fig. 13(a), with the increasing of the

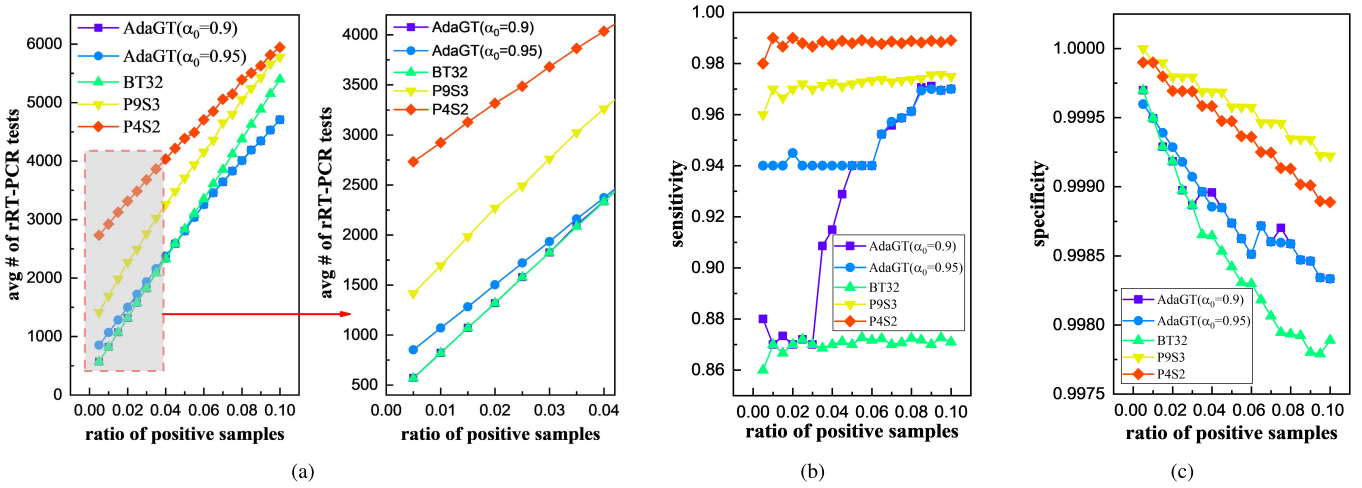


Fig. 12. Comparing the AdaGT algorithm with existing group testing methods against COVID-19 when the ratio of positive samples ranges from 0.005 to 0.1.

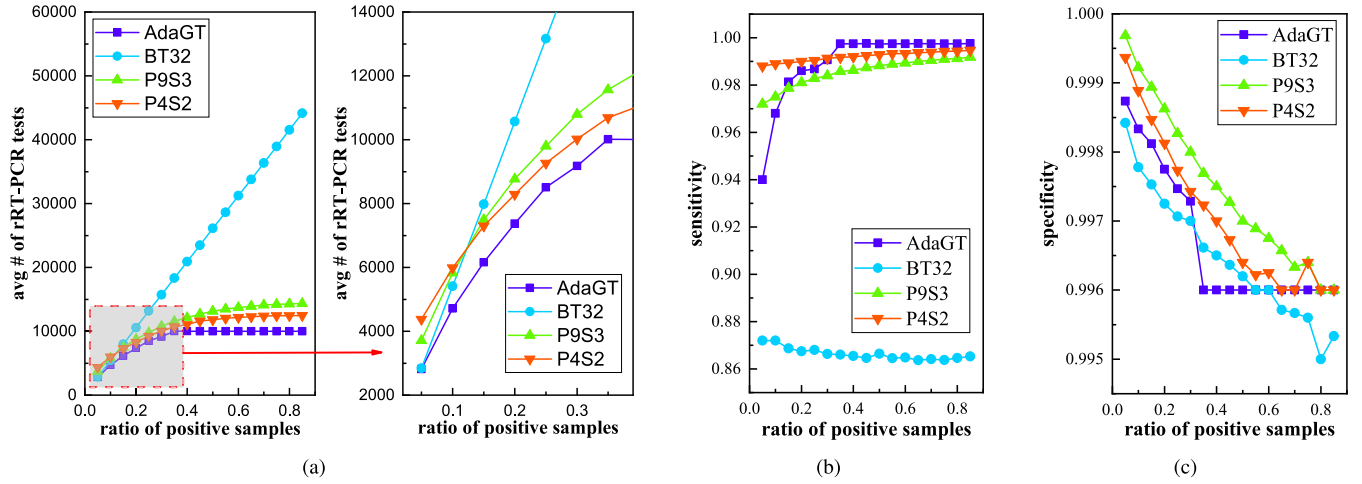


Fig. 13. Comparing the AdaGT algorithm with existing group testing methods against COVID-19 when the ratio of positive samples ranges from 0.05 to 0.9.

ratio of positive samples, the average number of rRT-PCR tests of the BT32 method in [23] increases linearly. For P9S3 and P4S2 methods in [24], the average number of rRT-PCR tests increases quickly at first; when the ratio of positive samples is larger than about 0.3, the increasing speed decreases. For the proposed AdaGT method, the average number of rRT-PCR tests increases monotonically at first; when the ratio of positive samples is larger than 0.35, the average number of rRT-PCR tests remains constant. Regardless of the ratio of positive samples, the AdaGT algorithm always takes the fewest rRT-PCR tests to screen out all positive samples.

In Fig. 13(b), we have the following observations: (1) With the ratio of positive samples increasing from 0.05 to 0.9, the sensitivity of the BT32 methods remains around 0.87 The sensitivity of the BT32 methods is always less than that of the AdaGT, the P9S3, and the P4S2 methods. (2) The sensitivity of the P9S3 and P4S2 methods in [24] increases slightly. (3) As for the AdaGT method, with the increase of the ratio of positive samples, the sensitivity first increases quickly and then remains almost constant. (4) When the ratio of positive samples is less than 0.1, the sensitivity of the AdaGT algorithm is less than that of the P9S3 and P4S2 methods. (5) When the ratio of positive samples is greater than 0.1, the sensitivity of the AdaGT algorithm is comparable or higher than that of the P9S3 and P4S2 methods.

In Fig. 13(c), We compare the specificity value of the proposed AdaGT algorithm with those of the P9S3, the P2S4, and the BT32 methods. As shown in the figure, with the increase of the ratio of positive samples, the specificity values of the above four methods tend to decreases monotonically and quickly. Particularly, after the ratio of positive samples increases to 0.35 or more, the specificity value of the AdaGT algorithm remains constant at 99.6%. Although the specificity values of the P9S3 and the P4S2 methods are a little bit higher than those of the proposed AdaGT and the BT32 methods, on the whole, the specificity values of the above four methods are very high.

Based upon the statistics released in [4], the ratio of confirmed COVID-19 cases to the total population of the world is about 0.02. However, this ratio varies among different countries. For example, the ratio of confirmed COVID-19 cases to the total population of the USA is about 0.1. We can easily infer that in some areas of the USA, this ratio is above 0.1, while in other areas of the USA, this ratio is below 0.1.

The AdaGT algorithm is practically useful, as analyzed in the following:

- As discussed above, regardless of the ratio of positive samples, the AdaGT algorithm can always perform fewer rRT-PCR tests to screen all positive samples than existing methods like the BT32, the P9S3, and the P4S2. In other words, regardless of the positive sample ratios, the screening efficiency of the AdaGT algorithm outperforms those of those methods.
- The AdaGT algorithm usually has higher sensitivity than the BT32 method. When the ratio of positive samples is below 0.1, by controlling the value of α_0, the proposed AdaGT algorithm can achieve comparable sensitivity with the P9S3 and the P4S2 methods. When the positive sample ratio is above 0.1, the AdaGT algorithm has comparable or even higher sensitivity with the P9S3 and the P4S2 methods.
- The specificity of the AdaGT method is very high.

VII. Conclusion

In this paper, we aim to screen the SARS-CoV-2 virus with as few rRT-PCR tests as possible, under the premise that the sensitivity of rRT-PCR tests is larger than a predetermined threshold. To achieve this goal, we propose the AdaGT algorithm. Based upon some information collected during the testing process, the AdaGT algorithm can estimate the ratio of positive samples during the screening process. If this estimated ratio is larger than a user-specified threshold, an individual testing strategy is applied to test the NP/OP swab samples separately. Otherwise, the group testing strategy is employed to test a group of NP/OP swab samples. In this case, the group size is carefully selected to guarantee that the sensitivity of the rRT-PCR test is higher than a predetermined threshold and that, on average, there is at most one positive sample in this group. If there are positive samples among this group, a binary testing strategy is further employed to identify one positive sample from these samples. We analyze the theoretical bounds of the number of rRT-PCR tests, and the theoretical performance analysis also indicates that when $y_0 = \frac{1}{3}$, the AdaGT algorithm can achieve a minimum average number of rRT-PCR tests. Experimental results show that the AdaGT algorithm outperforms existing group testing methods against COVID-19 efficiency and sensitivity.

As our future work with potential funding, we hope to cooperate with some hospitals to validate our model with real testing data to justify its advantages.

Acknowledgment

The authors would like to thank WeiXing Wu, who works in Cancer Hospital Chinese Academy of Medical Sciences, Shenzhen Center, for her generosity in sharing professional knowledge in screening the SARS-CoV-2 virus.

References

[1] World Health Organization. (2020). *Naming the Coronavirus Disease (COVID-19) and the Virus That Causes It*. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it

[2] World Health Organization. (2020). *Archived: WHO Timeline-COVID-19*. [Online]. Available: https://www.who.int/news/item/27-04-2020-who-timeline-covid-19

[3] A. M. Manoharan and M. G. Sumithra, "Secure data communication IoT and wireless sensor network for COVID-19," *Int. J. Sensor Netw.*, vol. 36, no. 1, p. 11, 2021.

[4] Worldometer. (2020). *COVID-19 Coronavirus Pandemic*. [Online]. Available: https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1?"

[5] World Health Organization. (2020). *Who Coronavirus Disease (COVID-19) Dashboard*. [Online]. Available: https://covid19.who.int/

[6] A. Sandford. (2020). *Coronavirus: Half of Humanity Now on Lockdown as 90 Countries Call for Confinement*. [Online]. Available: https://www.euronews.com/2020/04/02/coronavirus-in-europe-spain-s-death-toll-hits-10-000-after-record-950-new-deaths-in-24-hou

[7] World Health Organization. (2020). *Impact of COVID-19 on People's Livelihoods, Their Health and Our Food Systems*. [Online]. Available: https://www.who.int/news/item/13-10-2020-impact-of-covid-19-on-people's-livelihoods-their-health-and-our-food-systems

[8] (2020). *International Monetary Fund, the Great Lockdown: Worst Economic Downturn Since the Great Depression*. [Online]. Available: https://blogs.imf.org/2020/04/14/the-great-lockdown-worst-economic-downturn-since-the-great-depression/

[9] NIH Leadership. (2020). *Why COVID-19 Testing Is the Key to Getting Back to Normal*. [Online]. Available: https://www.nia.nih.gov/news/why-covid-19-testing-key-getting-back-normal#:~:text=Testingcanhelppeopledetermine,backtoworkandschool

[10] E. Sanchez. (2020). *COVID-19 Science: Why Testing Is so Important*. [Online]. Available: https://www.heart.org/en/news/2020/04/02/covid-19-science-why-testing-is-so-important

[11] D. Lakdawalla and E. Trish. (2020). *Group Testing for Coronavirus—Called Pooled Testing—Could Be the Fastest and Cheapest Way to Increase Screening Nationwide*. [Online]. Available: https://theconversation.com/group-testing-for-coronavirus-called-pooled-testing-could-be-the-fastest-and-cheapest-way-to-increase-screening-nationwide-141579

[12] American Society for Microbiology. (2020). *supply Shortages Impacting COVID-19 and Non-Covid Testing*. [Online]. Available: https://asm.org/Articles/2020/September/Clinical-Microbiology-Supply-Shortage-Collecti-1

[13] Krishna Kumar and ET Bureau. (2021). *Now, Shortage of RT-PCR Kits Adds to Maharashtra's COVID-19 Woes*. [Online]. Available: https://economictimes.indiatimes.com/news/india/now-shortage-of-rt-pcr-kits-adds-to-maharashtras-covid-19-woes/articleshow/81998594.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst

[14] Wikipedia. (2021). *List of Countries Implementing Pool Testing Strategy Against COVID-19*. [Online]. Available: https://en.wikipedia.org/wiki/List_of_countries_implementing_pool_testing_strategy_against_COVID-19#Screening_strategies

[15] B. D. Hitt. *A Shiny APP for Pooled Testing*. Accessed: Jan. 20, 2021. [Online]. Available: https://bilder.shinyapps.io/PooledTesting/

[16] A. P. Christoff *et al.*, "Swab pooling for large-scale RT-qPCR screening of SARS-CoV-2," *MedRxiv*, vol. 4, pp. 1–22, Sep. 2020.

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

[18] T. C. Quinn *et al.*, "Feasibility of pooling sera for HIV-1 viral RNA to diagnose acute primary HIV-1 infection and estimate HIV incidence," *Aids*, vol. 14, no. 17, pp. 2751–2757, Dec. 2000.

[19] G. K. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1880–1901, Mar. 2012.

[20] M. T. Goodrich and D. S. Hirschberg, "Improved adaptive group testing algorithms with applications to multiple access channels and dead sensor diagnosis," *J. Combinat. Optim.*, vol. 15, no. 1, pp. 95–121, 2008.

[21] X. Xia, Y. Xiao, and W. Liang, "ABSI: An adaptive binary splitting algorithm for malicious meter inspection in smart grid," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 2, pp. 445–458, Feb. 2019.

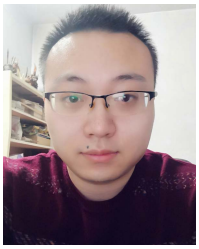
[22] X. Xia, Y. Xiao, W. Liang, and M. Zheng, "GTHI: A heuristic algorithm to detect malicious users in smart grids," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 2, pp. 805–816, Apr. 2020.

[23] C. Gollier and O. Gossner, "Group testing against COVID-19," *Covid Econ.*, vol. 1, no. 2, pp. 32–42, 2020.

[24] J. N. Eberhardt, N. P. Breuckmann, and C. S. Eberhardt, "Multi-stage group testing improves efficiency of large-scale COVID-19 screening," *J. Clin. Virol.*, vol. 128, Jul. 2020, Art. no. 104382.

[25] B. Netta. (2020). *Lessons From Applied Large-Scale Pooling of 133,816 SARS-COV-2 RT-PCR Tests*. [Online]. Available: https://www.medrxiv.org/content/early/2020/10/20/2020.10.16.20213405
[26] B. D. Hitt, C. R. Bilder, J. M. Tebbs, and C. S. Mcmahan, "The objective function controversy for group testing: Much ado about nothing?" *Statist. Med.*, vol. 38, no. 24, pp. 4912–4923, Oct. 2019.
[27] I. Yelin *et al.*, "Evaluation of COVID-19 RT-qPCR test in multi sample pools," *Clin. Infectious Diseases*, vol. 71, no. 16, pp. 2073–2078, Nov. 2020.
[28] S. Y. Kim *et al.*, "Pooling upper respiratory specimens for rapid mass screening of COVID-19 by real-time RT-PCR," *Emerg. Infectious Diseases*, vol. 26, no. 10, pp. 2469–2472, Oct. 2020.
[29] S. Shukla, V. Upadhyay, and V. K. Maurya, "Evaluating the efficiency of specimen (sample) pooling for real-time PCR based diagnosis of COVID-19," *Indian J. Med. Microbiology*, pp. 2193–2199, Mar. 2021.
[30] N. Jawerth. (2020). *How Is the COVID-19 Virus Detected Using Real Time RT-PCR*. [Online]. Available: https://www.iaea.org/bulletin/infectious-diseases/how-is-the-covid-19-virus-detected-using-real-time-rt-pcr
[31] Wisconsin Veterinary Diagnostic Laboratory. (2021). *Real Time PCR Ct Values*. [Online]. Available: https://www.wvdl.wisc.edu/wp-content/uploads/2013/01/WVDL.Info_.PCR_Ct_Values1.pdf
[32] M. R. Tom and M. J. Mina, "To interpret the SARS-CoV-2 test, consider the cycle threshold value," *Clin. Infectious Diseases*, vol. 71, no. 16, pp. 2252–2254, Nov. 2020, doi: 10.1093/cid/ciaa619.
[33] D. S. Mouliou and K. I. Gourgoulianis, "False-positive and false-negative covid-19 cases: Respiratory prevention and management strategies, vaccination, and further perspectives," *Expert Rev. Respiratory Med.*, vol. 4, pp. 1–10, Apr. 2021.
[34] Centers for disease control and prevention. (2021). *Interim Guidance for Antigen Testing for SARS-CoV-2*. [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antigen-tests-guidelines.html
[35] Centers for Disease Control and Prevention. (2021). *Nucleic Acid Amplification Tests (NAATs)*. [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/lab/naats.html
[36] D.-Z. Du and F. K. Hwang, *Series on Applied Mathematics: Combinatorial Group Testing and Its Applications*, vol. 12. Singapore: World Scientific, 2000.
[37] A. Bar-Noy, F. K. Hwang, I. Kessler, and S. Kutten, "Competitive group testing in high speed networks," *Discrete Appl. Math.*, vol. 52, pp. 29–38, May 1994.

Xiaofang Xia (Member, IEEE) received the Ph.D. degree in control theory and control engineering from the Shenyang Institute of Automation, Chinese Academy of Sciences, China, in 2019. She is currently an Assistant Professor with the School of Computer Science and Technology, Xidian University, China. She was a Visiting Student with the Department of Computer Science, University of Alabama, USA, from August 2016 to February 2018. Her research interests are mainly in cyber-physical systems, smart grid security, database management systems, and anomaly detection.



Yang Liu received the B.S. degree in computer science from the Harbin Institute of Technology in 2016. He is currently pursuing the Ph.D. degree with the School of Computer Science and Technology, Xidian University, China. His research interests are mainly in database management systems, anomaly detection, and graph query.

Yang Xiao (Fellow, IEEE) received the B.S. and M.S. degrees in computational mathematics from Jilin University, Changchun, China, and the M.S. and Ph.D. degrees in computer science and engineering from Wright State University, Dayton, OH, USA. He is currently a Full Professor with the Department of Computer Science, The University of Alabama, Tuscaloosa, AL, USA. He had directed 20 doctoral dissertations and supervised 19 M.S. theses/projects in the past. He has published over 300 SCI-indexed journal papers (including over 60 IEEE/ACM TRANSACTIONS papers) and 300 EI indexed refereed conference papers and book chapters related to these research areas. His current research interests include cyber-physical systems, the Internet of Things, security, wireless networks, smart grid, and telemedicine.

He is a fellow of IET and AAIA. He currently serves as the Editor-in-Chief for *Cyber-Physical Systems* journal, *International Journal of Sensor Networks* (IJSNet), and *International Journal of Security and Networks* (IJSN). He has been serving as an Editorial Board or an Associate Editor for 20 international journals, including the IEEE TRANSACTIONS ON CYBERNETICS since 2020, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS from 2014 to 2015, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from 2007 to 2009, and *IEEE Communications Survey and Tutorials* from 2007 to 2014. He has served as a Guest Editor over 30 times for different international journals, including the IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING in 2021, IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING in 2021, IEEE NETWORK in 2007, IEEE WIRELESS COMMUNICATIONS in 2006 and 2021, *IEEE Communications Standards Magazine* in 2021, and ACM/Springer Mobile Networks and Applications (MONET) in 2008. He was a Voting Member of the IEEE 802.11 Working Group from 2001 to 2004, involving the IEEE 802.11 (WIFI) standardization work.

Jiangtao Cui (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in computer science from Xidian University, China, in 1998, 2001, and 2005, respectively. He is currently a Professor with the School of Computer Science and Technology, Xidian University. He was a Visiting Scholar with The University of Queensland, Australia, from 2007 to 2008. In these fields, he has published over 60 academic papers in top-level conferences and important international academic journals, including ICDE, SIGMOD, VLDB, TKDE, and the *VLDB Journal*. His research interests mainly include data security, privacy-preserving, blockchain application technology, and complex data query and indexing.

Bo Yang (Member, IEEE) received the Ph.D. degree in control science and engineering from the University of Chinese Academy of Sciences, Beijing, China, in 2017. He is currently a Post-Doctoral Fellow with the University of Toronto. His research interests include wireless networks, edge computing, and multi-agent learning.

Yanguo Peng (Member, IEEE) received the B.S. degree in network engineering from the North University of China, Taiyuan, China, in 2009, the M.S. degree in computer software and theory from Guizhou University, Guiyang, China, in 2012, and the Ph.D. degree in computer systems organization from Xidian University, China, in 2016. He is currently an Associate Professor with the School of Computer Science and Technology, Xidian University. He is also a Master Supervisor. His research interests include cloud security, data privacy protection, and blockchain.