Merkle and Binary Tree Inspired Method for Accelerated COVID-19 Testing: A Concept

Preprint · April 2020

DOI: 10.13140/RG.2.2.26116.88969

4 authors:

Aaqib Bashir Dar

1 PUBLICATION **0** CITATIONS

SEE PROFILE

Muzafar Ahmad Wani

National Institute of Technology, Srinagar

4 PUBLICATIONS **18** CITATIONS

SEE PROFILE

Auqib Hamid Lone

National Institute of Technology, Srinagar

14 PUBLICATIONS **31** CITATIONS

SEE PROFILE

Roohie Naaz Mir

National Institute of Technology, Srinagar

118 PUBLICATIONS **210** CITATIONS

SEE PROFILE

Merkle and Binary Tree Inspired Method for Accelerated COVID-19 Testing: A Concept

Aaqib Bashir Dar[a], Muzafar Ahmad Wani[b], Auqib Hamid Lone[c,*], Roohie Naaz[d]

[a]*Independent Researcher, Jammu and Kashmir,India,190015*
[b]*Department of Computer Science and Engineering, NIT Srinagar, Jammu and Kashmir, India,190006*
[c]*Department of Computer Science and Engineering, NIT Srinagar, Jammu and Kashmir, India,190006*
[d]*Department of Computer Science and Engineering, NIT Srinagar, Jammu and Kashmir, India,190006*

Abstract

In this concept paper, we propose a theoretical model inspired by Merkle and Binary trees for accelerated COVID-19 testing. We have formulated and simplified the idea of how the testing can be improved drastically while tracing the infected individual with supposedly very less complexity. Our approach reduces the complexity in terms of the number of tests that are to be carried out. We believe that the idea is promising but has no practical foundations as yet. For now, no practical examinations on real samples are carried out as per this model. However, our model in theory has some amazing implications and will help in carrying out more tests while covering a huge number of individuals. Since researchers at Technion and Rambam [1, 2, 3, 4] has improved the traditional testing procedure by simply pooling multiple samples in a single test tube which are then screened using the normal PCR testing procedure which takes several hours. Although there lies some logistical challenges while implementing this method, the number of samples tested per day will surely increase by large and will help in identifying the asymptotic carriers.

Keywords: COVID-19, Rapid Testing, Pooling, Sampling.

1. Proposed Model

Our proposed model works in two phases during which all the processing and evaluation is done. The first of which is called the Sampling Phase, we call the second as the Testing Phase. In our Merkle Tree [5] inspired approach, for sampling the bottom-up approach is used (Merkle Tree construction) while as the testing is done in the top-down manner (Binary tree). For brevity, we label the individuals as $1, 2, 3, \ldots, N$ and their corresponding samples as S_1, S_2,... S_N. The Merkle tree in the model is constructed in a bottom-up fashion where the individual's samples have height 0 i.e.,

*Auqib Hamid Lone

Email addresses: aaqibb13@gmail.com (Aaqib Bashir Dar), muzu28iust@gmail.com (Muzafar Ahmad Wani), ahl@nitsri.net (Auqib Hamid Lone), naaz310@nitsri.net (Roohie Naaz)

they are leaves which are then grouped in two so as to form a tree. In a similar fashion, the tree keeps growing upwards until a common single root is formed which is actually the test-tube that contains all the mixed samples. We call that as a Merkle sample. The simplified view of the entities used in our model can be found in the Figure 1.

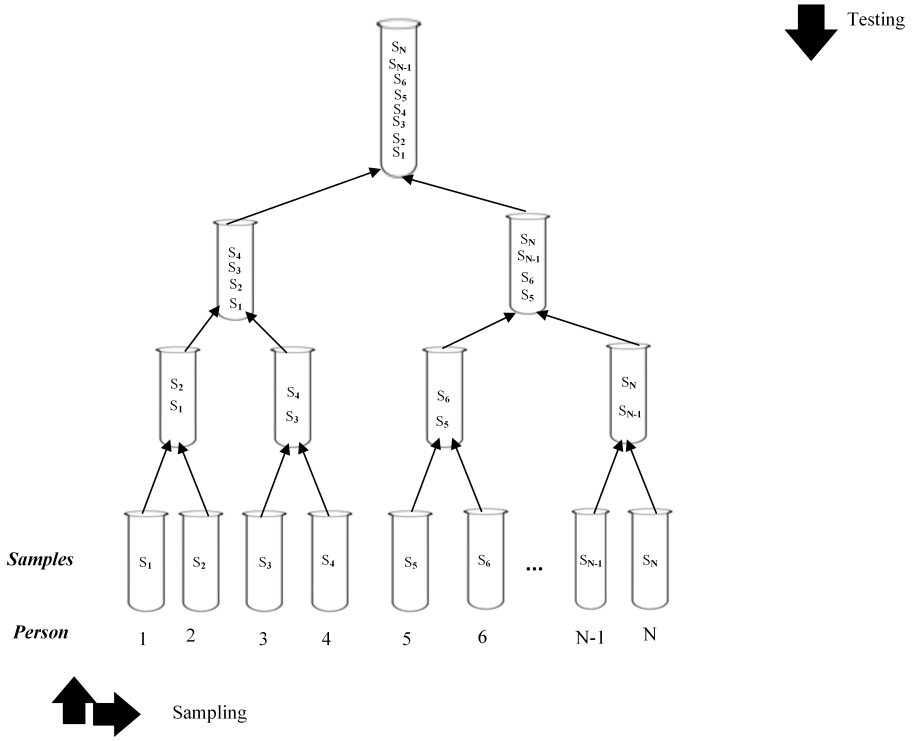

Figure 1: Simplified Architecture of Proposed Model

1.1. Sampling

It is clear from the figure that a part of the sample is collected and then stored in the test-tubes which form the base for the process of sampling. The process of sampling starts with these samples being grouped into two's, which is done at random since no prior knowledge about a particular sample being positive is known. However, we will discuss about some notions in a separate section which shed some light on the adaption of certain measures in order to increase efficiency in terms of the maximum numbers of tests being conducted. The grouping is done repeatedly in upward direction until a common root is formed (Merkle root in our case). The Merkle root contains all the individual's samples which are pooled together in order to perform molecular test to determine whether an individual is infected with novel Corona Virus or not.

Stepwise Preparation of Samples

1. Collect samples of all individuals visiting for check-up.

2. Divide a large set of samples into batches each having 64 (variable) samples
3. Pool the samples in test-tubes into groups of two.
4. Apply step 3 repeatedly on the mixed samples until a single sample (Merkle Root Sample) is prepared.

1.2. Testing

The core of our theoretical model is the Testing phase or procedure. The testing phase starts right at the top, at the Merkle Root itself and starts moving downwards. For better understanding, we will explain the testing phase in various cases. The simplest of which is, where only one single positive case is found among a group of individuals. We take some small cases as a basis of this model and discuss about the favorability of our model along with some thorough analysis of the best and worst cases and finally compare them with the traditional testing procedures. Testing procedure is summarized in Algorithm 1.

Algorithm 1: Testing Procedure

```
1  Test Root Sample
2  if -ve then
3  |  Stop ;
4  else
5  |  Go one level down;
6  |  Test both Left Child ad Right Child samples of Root;
7  |  if Both Left Child and Right Child are +ve then
8  |  |  Stop and Test all Samples Individually;
9  |  else
10 |  |  Test Child samples of +ve Sample only;
11 |  |  if Both Left Child and Right Child are +ve then
12 |  |  |  Test child samples of both ;
13 |  |  |  if Child samples are leaves then
14 |  |  |  |  return +ve sample;
15 |  |  |  else
16 |  |  |  |  Go one level down;
17 |  |  |  |  Goto step 11 ;
18 |  |  else
19 |  |  |  Test Child samples of +ve Sample only;
20 |  |  |  if Child samples are leaves then
21 |  |  |  |  return +ve sample;
22 |  |  |  else
23 |  |  |  |  Go one level down;
24 |  |  |  |  Goto step 11 ;
```

We present an example of the testing procedure as a Test-case in the analysis section.

Table 1: Results from Theoretical Analysis on Varying Sample Sizes

Samples		**No. of Tests required to be done**			
No. of original Samples (N) of the form 2^n from n=3,...,n	Samples after Pooling	Best Case	2^{nd} Best	2^{nd} Worst	Worst Case
8	15	1	7	9	11
16	31	1	9	17	19
32	63	1	11	33	35
64	127	1	13	65	67
128	255	1	15	129	131
2^n	$2^{n+1}-1$	1	$7+(n-1)*2$	2^n+1	2^n+3

2. Theoretical Analysis and Results

We have analysed the behaviour of our model and presented a numerical analysis of best and worst cases. We represent the number of samples by N and always group them in two such that they are in 2^n form. We start by taking (N=8 i.e., n=3) and go further till 128 (n=7) to check the feasibility and correctness of the model. After which, we then generalized the behaviour of our model. Finally, we consider a test-case in which we take an example scenario and show how our technique produces better results than the already available traditional technique. It is to be noted that no practical work was carried out to support the evidence. We also understand the fact that the process of sampling and testing are in practice quite complex which requires precision, calibration and controlled environments. However, we have mentioned that our model is conceptual and if found correct, should be carried out only after thorough analysis under the observation of a scientist who is an expert of the subject matter. Theoretical results are presented in Table 1.

2.1. Testcase

Consider batches of 64 samples are collected 100 times on a particular day.

Traditional Method: If we have batches each having 64 samples being collected 100 times a day. A total of 6400 tests are to be conducted to confirm the positive Corona virus patients. Though, we can consider the amount of time taken to test a patient. We do not consider that as a parameter in our model as it varies from one country to another. The model's primary goal is to reduce the number of tests being conducted which in turn will save time as well.

Proposed Method: We base our model on the fact that pooling of samples has been successfully done. However, what the pooling method lacks is, if the pooled sample turns out to be positive. All the individuals are to be checked one by one. We went a step further and used what one can call multi-level pooling (if that is the right term to use and is actually possible in practice). At every step during the sampling phase we pooled the samples in two so that during testing, we'll be able to track down the patient while lowering the complexity. For clarity, we consider a small example in this Test-case (say N=16 i.e., n=4) as shown in Figure 2.

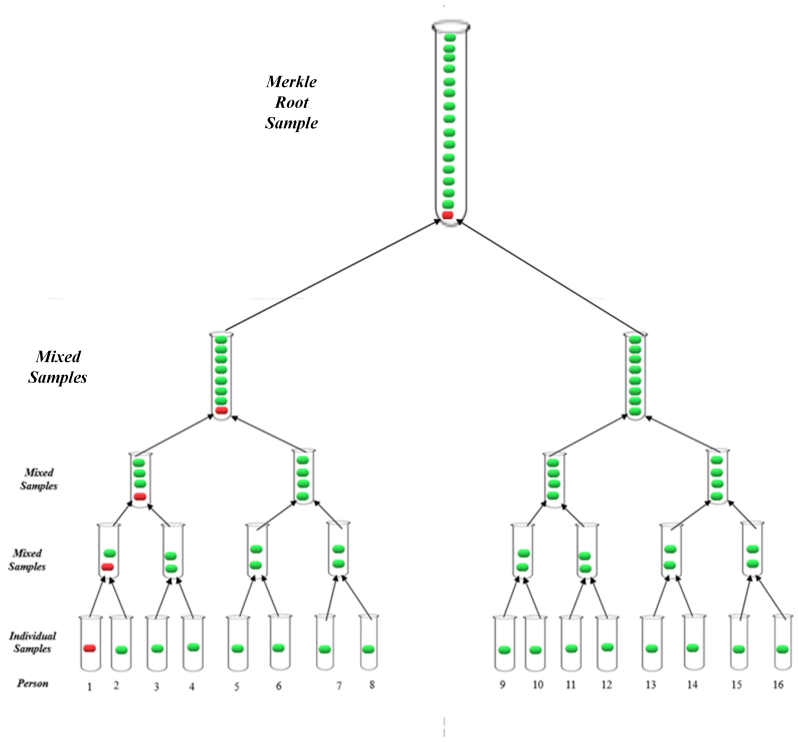

Figure 2: Test-case with 16 samples

For the sake of simplicity, we have marked the positive cases as Red and the negative cases as green. We now assume that there is only one positive case among a total of 16 samples and can randomly be anywhere in between these 16 samples since it is equally likely because we have no prior knowledge about the positive case. After collecting samples from 16 individuals, we consider the case where Person 1 is positive. The Bottom-up approach in the figure above explains the sampling phase. After getting the Merkle Sample (Root) which contains all the pooled samples. We go for the testing phase which is done in the Top-down manner. We now start with our algorithm of the testing phase. We tested the root which is found positive, we then move towards the root's children. We test them both and found only the left child to be positive. We only take left child into consideration since only the left child is found positive. Without loss of generality, we can clearly see that the case we have seen for (N=16), the total number of tests to be conducted is 9. Our model reduces the complexity by almost half in this case and is believed to produce better results for larger number of samples (i.e., when mass testing is done). We have a done a detailed analysis up to 128 individuals and have mentioned how our model can behave in the best and worst of cases.

3. Conclusions and Future Work

Our approach, if found suitable and correct should reduce the chances of infection spread and flatten the infection curve. Currently the pooling protocol that was successfully tested by Israeli researchers is designed as a molecular design to determine whether or not an individual has been infected with Corona virus but is believed to be applied to serological tests which will require calibration and careful testing to see whether it works. They even conducted a joint examination of 64 samples in which only one was a positive carrier[1, 2, 3, 4]. Their system was able to identify that one particular sample. We have done some analysis and discussed that in analysis section to determine how much testing benefit will be derived from our method where testing certain pooled specimens reduces the complexity and retesting everyone individually is only required if both the children of the Merkle root are positive. Likely this would produce greater benefit for mass screening at a time and place where the great majority is negative. This might save some testing resources, at a time when they are limited. We further believe that there are some other factors that can drastically increase the benefits of our model which is the categorization or grouping of individuals based on their travel history, severity of symptoms and type of contact. We consider that as a matter of future work to determine the effect of grouping of individuals by keeping these factors into consideration. This in essence is believed to help in reducing the number of tests being carried out so as to encourage the idea of mass testing at large and very efficiently. As per our knowledge, no such work has been carried out anywhere else as yet.

References

[1] "Israelis introduce method for accelerated covid-19 testing," accessed: 09-04-2020. [Online]. Available: https://www.israel21c.org/israelis-introduce-method-for-accelerated-covid-19-testing/

[2] "Israeli researchers announce new, faster testing method for covid-19." accessed: 09-04-2020. [Online]. Available: https://www.israelhayom.com/2020/03/22/israeli-researchers-announce-new-faster-testing-method-for-covid-19/

[3] "Israelis introduce method for accelerated covid-19 testing," accessed: 09-04-2020. [Online]. Available: https://www.tabletmag.com/sections/science/articles/pooling-covid-19-israel

[4] "Israeli researchers introduce pooling method for covid-19 testing of over 60 patients simultaneously," accessed: 09-04-2020. [Online]. Available: https://www.hospimedica.com/coronavirus/articles/294781273/israeli-researchers-introduce-pooling-method-for-covid-19-testing-of-over-60-patients-simultaneously.html

[5] R. C. Merkle, "A digital signature based on a conventional encryption function," in *Conference on the theory and application of cryptographic techniques*. Springer, 1987, pp. 369–378.