

Using The Divide-and-Conquer Technique in RT-PCR Group Test to Increase Coronavirus Testing Rate in Indonesia

Fakhrurrida Clarendia Widodo 13518091

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia

13518091@std.stei.itb.ac.id

Abstract—With only 40 in every million are being tested for the coronavirus, Indonesia is ranked the fourth worst in testing rate among countries with a 50 million population or above. This is mainly due to the fact that large quantities testing of samples to detect one or more positive samples is high-priced and time consuming. Pooled RT-PCR testing similar to divide and conquer algorithm can optimize the coronavirus testing process. The present paper proves that the pooled RT-PCR test could inflate the coronavirus testing rate in Indonesia using a simulation run by a Python code.

Keywords—Pool Testing, Group Testing, Coronavirus 2019, Divide and Conquer, RT-PCR.

I. INTRODUCTION

On December 31, 2019, the World Health Organization's (WHO) office in China received the first reports of an unknown virus behind the skyrocketing number of pneumonia cases in Wuhan, a city in Eastern China with a population of 11.08 million. What started as a local epidemic only limited to this region has now become a truly worldwide pandemic. According to John Hopkins University Coronavirus dashboard, by May 1 2020, there have been over 3,117,880 cases and 217,212 deaths related to this virus. Those number are collated from national health authorities all over the globe. The disease has been discovered in more than 200 countries, with Spain, the US, and Italy experiencing the worst outbreaks outside of China.^[1]

Meanwhile, from January until February 2020, Indonesia reported zero case of coronavirus even being encircled by infected countries such as Singapore, Malaysia, Philippines, and Australia. Flights from highly infectious countries like Thailand, China, and South Korea continued operating. Medical researchers and experts at Harvard University of the US stated that Indonesia is unprepared for an outbreak and there could be dozens of undetected coronavirus cases that time. Until, on the March 2020, Indonesian president Joko Widodo reported the first cases of coronavirus in the country: a dance instructor and her mother in Depok, West Java. They both had held a dance class at Kemang, Jakarta on 14 February, which was attended by more than a dozen of people. One of them was a Japanese man, who was later tested positive for coronavirus test in Malaysia.^[2]

Since then, the number of cases in Indonesia was skyrocketing. Health experts all over the globe are concerned that Indonesia is failing to trace the transmission of the virus. Marc Lipsitch of Harvard T.H. Chan School of Public Health suggested that Indonesia might have missed cases.^[3] Diplomats as well as regional and international news outlets hypothesized that the lack of coronavirus cases in Indonesia is a result from inadequate testing and under reporting, as opposed to sheer luck and cleric intervention. On 22 March 2020, a study proposed that the official number of cases may only reflect 2% of the real cases of infections in Indonesia. By 5 April, the central government has only conducted a daily average of 240 RT-PCR tests since 2 March.^[3]

II. BASIC THEORY

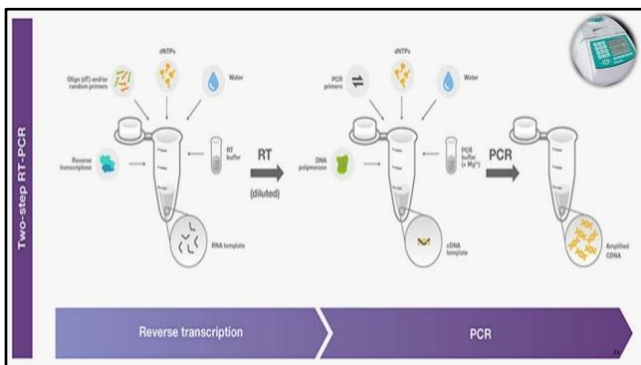
A. RT-PCR Test

RT-PCR (Reverse Transcription Polymerase Chain Reaction) is a laboratory technique of amplification of specific DNA (Deoxyribonucleic acid) using polymerase chain reaction and combining the reverse transcription of RNA (Ribonucleic acid) into DNA. This technique is mainly used to quantify the amount of certain RNA. This procedure is done by the help of a method called quantitative PCR, monitoring the amplification reaction using fluorescence or a visible radiation emitted by the substances as a result of incident radiation of ultraviolet light. The combination of RT-PCR and quantitative PCR are regularly applied for analysis of quantification and expression of viral RNA in research and medical environment. This close association between quantitative PCR and RT-PCR made the use of the term quantitative PCR to mean RT-PCR.

Oftentimes, RT-PCR is applied in research procedures to determine gene expression. For example, Lin et alia used RT-PCR and quantitative PCR to measure Gal genes' expression in yeast cells. First, they engineered designed a mutation of protein suspected to join in the regulation of Gal genes. That mutation was postulated to selectively eliminate Gal expression. To support this postulate, gene expression levels of yeast cells containing this mutation were examined using both RT-PCR and

quantitative PCR. They were able to discover that the mutation of this regulatory protein reduced Gal expression.

Beside of that, RT-PCR can be very advantageous in the insertion of eukaryotic genes into prokaryotes. The fact that most eukaryotic genes contain introns, a segment of DNA or RNA molecule which is present in the genome but not in the mature RNA, implies that the cDNA generated from a RT-PCR procedure is the exact DNA sequence that would be directly translated into protein after transcription. When these genes are expressed in prokaryotic cells in order to preserve the protein production or purification, the RNA fabricated right from transcription does not need to go through the step of splicing as the transcript only contains exons.^[4]



Picture 2.1. The Procedure of Two-Step RT-PCR Testing
<https://bioline.com/>

RT-PCR can also be used to identify genetic disease such as Lesch-Nyhan Syndrome. This genetic disease is triggered by a malfunction in the hypoxanthine phosphoribosyltransferase-1 (HPRT1) gene. This condition medically leads to the fatal uric acid urinary stone and symptoms almost identical to gout. Examining a pregnant mother and a fetus for mRNA expression levels of HRPT1 will divulge if the mother is a carrier or if the fetus will plausibly to develop Lesch-Nyhan Syndrome.^[5]

B. Pooled RT-PCR Testing



Picture 2.2. Drive-thru Coronavirus Testing was Set Up in Las Cruces at the Doña Ana County Health Services Center, Scheduled from on Friday, March 20, 2020, from 9 a.m. to 3 p.m. or Until Supplies Ran Out.
<https://lcsun.news.com/>

Pooled testing is a procedure of examining samples of multiple individuals which are put together in a tube and screened through the PCR test. This testing often refers to combining multiple samples that are to be assessed in a single test serum and then testing the sample to look for virus or contagion. If the entire pool of individual samples that was tested turns out to be negative, that suspected virus is not present in all the individuals who submitted their samples in this pool. But, if the pool of samples is found to be positive, the pool will be split into two sub-samples to be retested.^[6]

C. Molecular Testing



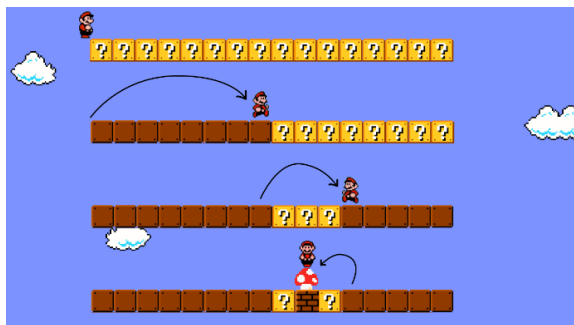
Picture 2.3. A Researcher in Vancouver’s Molecular Testing Labs (MTL) Performs The Procedure of Molecular Testing
<https://vbjusa.com/>

Molecular testing is a laboratory techniques practiced to study biological markers in the proteome and genome. Proteome is the individual's genetic code and how their cells express their genes as unique proteins by applying molecular biology to clinical testing. This testing technique is applied to identify and observe disease, trace risks, and determine which therapies will work best for individual patients. By analyzing the specifics treatment for specific patients and specific diseases, molecular diagnostics offers the future of personalized medicine.

These tests are quite beneficial in a range of medical specialization, including cancer detection, infectious disease, human leucocyte antigen typing, coagulation, and pharmacogenomics. Molecular testing utilizes *in vitro* biological assessment such as Fluorescence *in situ* hybridization or PCR-ELISA (Polymerase chain reaction-enzyme linked immune sorbent assay). The assessment will detect a molecule that is a marker of certain disease or risk in a sample taken from the patient.^[7]

Molecular testing is applied to identify many infectious diseases such as influenza virus, tuberculosis, H1N1 virus, or chlamydia. Molecular testing is also used to comprehend the particular strain of the pathogen. For example, it is used for detecting which drug resistance genes an organism possesses and hence which therapies to avoid.

D. Binary Search Algorithm

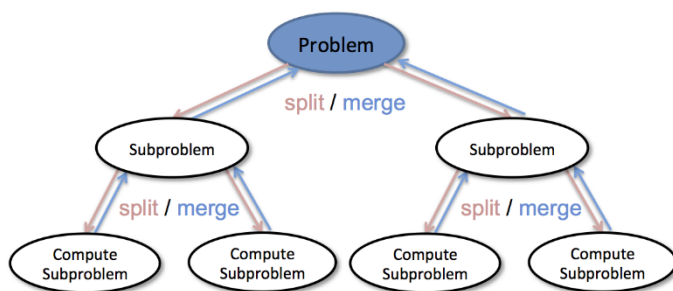


Picture 2.4. Binary Search as Illustrated by the Gameplay of Mario Bros Video Game
<https://medium.com/>

In the field of computer science or computer engineering, binary search or often called as half-interval search, logarithmic search, or binary chop, is a technique that finds the position of a target value within a sorted list. Binary search algorithm collate the target value to the middle element of the list. If they are not equal, the half in which the target does not lie is eliminated from the candidates and the search for the target value continues on the remaining half. The process is repeating again by taking the middle element to compare to the target value until the target value is found. If the search ends with the candidates being empty, the target is not in the list. [8]

Binary search algorithm runs in logarithmic time, in the worst case, making $O(\log n)$ comparisons, where n is the number of elements in the list. Binary search algorithm is faster than almost every naïve linear search except for small lists. However, to be able to apply binary search algorithm, the list must be sorted first. Actually, there are specialized data structures designed for faster searching, such as hash tables, that can locate the target value more efficiently than binary search algorithm. However, binary search algorithm can be applied to solve a wider range of problems, such as finding the next-smallest or next-largest element in the array relative to the target even if it is absent from the list.

E. Divide-and-Conquer Strategy



Picture 2.5. The Flowchart of Divide-and-Conquer Technique
<https://slideshare.com/>

In the field of computer science, divide-and-conquer strategy is a problem solving technique based on branched recursion. A divide-and-conquer strategy works by

recursively breaking down a problem into sub-problems of the same or related problem type, until these sub-problems become simple enough to be solved directly. The solutions to the sub-problems are then concurrently conjoined to give a solution to the main problem. This divide-and-conquer strategy is the basis of efficient algorithms for all kinds of problems, such as sorting, multiplying large numbers, discovering the closest pair of points, and calculating the discrete Fourier transform (FFT).

Understanding the divide-and-conquer strategies is a complex skill that requires a good concept of the nature of the original problem to be solved. The validity of a divide-and-conquer strategy is usually verified by mathematical induction, and its computational cost is often calculated by resolving recurrence relations. [9]

The divide-and-conquer strategies are often used to find a prime solution to a problem. Its basic idea is to disentangle a given problem into simpler, sub-problems, to solve them in turn, and to compose their solutions to solve the given problem. Some problems have sufficient simplicity to be solved directly. For example, to sort a given array of n natural numbers, split it into two lists of in a size of $n/2$ numbers each, sort each of them in turn, and conjoin both results appropriately to obtain the sorted version of the given list. Such technique is known as the merge sort algorithm. [10]

A significant application of divide-and-conquer strategy is in optimization, where if the search space is pruned by a constant factor in every step, the overall algorithm has the identical asymptotic complexity as the pruning step, with the constant relying on the pruning factor. In many literature this technique is known as prune and search. The divide-and-conquer strategy often helps in the invention of more efficient algorithms. For example, it was the key to Karatsuba's fast multiplication for large numbers method, the merge sort and quick sort algorithms, and the Strassen's algorithm for large matrix multiplication. In all those instances, the approach of divide-and-conquer strategy led to an improvement in the asymptotic cost of the solution.

III. ANALYSIS

A. Testing Method

The method used to analyze the impact of RT-PCR group testing is by comparing the results of regular RT-PCR testing simulation to the results of divide-and-conquer RT-PCR group testing. Compared in this analysis is the amount of the tests need to be held by each technique in order to reveal all samples' test results. To represent the test population, this analysis uses an array of tuple of boolean and string. The boolean used as an attribute of a test subjects to indicate whether they are infected or not are generated randomly with the probability of being infected is 4.3%. This figure of probability is used based on the ratio of positive cases to number of people tested worldwide.

```

def generatePopulation(number):
    a = random.choices(
        population=[[True], [False]],
        weights=[0.043, 0.957],
        k=number)
    new_a = []
    for i in range (0, number):
        new_a.append(a[i].copy())
        new_a[i].append(i)

```

Picture 3.1. The Algorithm of Random Test Samples Generation as Written in Python Programming Language

Meanwhile, in simulating the regular RT-PCR testing, this analysis uses a procedure called *normalTesting* that receives an array of test subjects and evaluates each and every element of the array. Thus, making the number of tests held in order to reveal all samples' test results equals to the length of the array or the amount of people participated in the test.

```

def normalTesting(population):
    for i in range (len(population)):
        if (population[i][0] == True):
            print(str(population[0][1]) +
                "is negative.")
        else:
            print(str(population[0][1]) + "
                "is positive.")
    print("Number of Tests: ", len
        (population))

```

Picture 3.2. The Algorithm of Regular RT-PCR Testing Simulation as Written in Python Programming Language

Whereas, in simulating the divide-and-conquer RT-PCR group testing, this analysis uses recursive procedure called *pooledTesting* that receives an array of test subjects and uses the basic idea of divide-and-conquer strategy. Thus, making the amount of the tests held in order to reveal each and every samples' test results equals to the number of times this procedure is being called.

```

def pooledTesting(population):
    global called
    called+=1
    if (len(population)==1):
        if (population[0][0]):
            print(str(population[0][1]) +
                "is negative.")
        else:
            print(str(population[0][1]) +
                "is positive.")
    else:
        if (!isContainingPos(population):
            for i in range (len(population)):
                print(str(population[i][1]) +
                    " is negative.")
        else:
            a = splitList(population, 1)
            b = splitList(population, 2)
            if (isContainingPos(a)):
                pooledTesting(a)
            else:
                for i in range (len(a)):
                    print(str(a[i][1]) +
                        " is negative.")
            if (isContainingPos(b)):
                pooledTesting(b)
            else:
                for i in range (len(b)):
                    print(str(b[i][1]) +
                        " is negative.")

```

Picture 3.3. The Algorithm of The Divide-and-Conquer RT-PCR Group Testing Simulation as Written in Python Programming Language

Also, this analysis uses a procedure called *isContainingPos* to simulate the process of grouping the samples into a pool to later be tested as a whole.

```

def isContainingPos(array_pop):
    pos_found = False
    i = 0
    while ((not pos_found) and i < len(array_pop)):
        if (array_pop[i][0]):
            i += 1
        else:
            pos_found = True
    return pos_found

```

Picture 3.4. The Algorithm of Pooled Sample Testing as Written in Python Programming Language

B. Testing Result

The table shown below are the results obtained from the simulation for each test technique if the amount of people tested is 10.

Test Nr.	Number of Test Held (Non-group Test)	Number of Test Held (Group Test)
1	10	1
2	10	1
3	10	1
4	10	4
5	10	1

Table 3.1. The Results of RT-PCR Testing Simulation if The Amount of People Tested is 10

The table shown below are the results obtained from the simulation for each test technique if the amount of people tested is 50.

Test Nr.	Number of Test Held (Non-group Test)	Number of Test Held (Group Test)
1	50	1
2	50	11
3	50	6
4	50	7
5	50	7

Table 3.2. The Results of RT-PCR Testing Simulation if The Amount of People Tested is 50

The table shown below are the results obtained from the simulation for each test technique if the amount of people tested is 500.

Test Nr.	Number of Test Held (Non-group Test)	Number of Test Held (Group Test)
1	500	140
2	500	83
3	500	60
4	500	123
5	500	84

Table 3.3. The Results of RT-PCR Testing Simulation if The Amount of People Tested is 500

The table shown below are the results obtained from the simulation for each test technique if the amount of people tested is 10,000.

Test Nr.	Number of Test Held (Non-group Test)	Number of Test Held (Group Test)
1	10,000	1,798
2	10,000	1,921
3	10,000	1,933
4	10,000	1,786
5	10,000	1,907

Table 3.4. The Results of RT-PCR Testing Simulation if The Amount of People Tested is 10,000

The table shown below are the results obtained from the simulation for each test technique if the amount of people tested is 100,000.

Test Nr.	Number of Test Held (Non-group Test)	Number of Test Held (Group Test)
1	100,000	19,270
2	100,000	18,631
3	100,000	19,063
4	100,000	18,892
5	100,000	19,337

Table 3.5. The Results of RT-PCR Testing Simulation if The Amount of People Tested is 100,000

C. Result Analysis

From the results of simulations run on each technique and population, we get the mean of the amount of tests held for each technique as shown below.

Populations	Mean of Number of Tests(s) Held (Non-group Test)	Mean of Number of Test(s) Held (Group Test)
10	10	1.6
50	50	6.4
500	500	98
10,000	10,000	1,869
100,000	100,000	19,038.6

Table 3.5. The Average Number of Test(s) Held by Each Technique and Each Number of Population

Based on the table above, the divide-and-conquer could respectively increase the testing rate by the factor of 6.35, 7.8125, 5.1020, 5.3505, and 5.2525 or increasing the testing rate by the factor of 5.9735 on average.

IV. CONCLUSION

After analyzing the simulation run for every test technique, we can see the contrast result between the number of tests held by regular RT-PCR testing and by the divide-and-conquer RT-PCR group testing in order to reveal each samples' test result. Thus, we can conclude that the divide-and-conquer RT-PCR group testing is both theoretically and experimentally proven to increase the coronavirus testing rate by up to a factor of 7.8.

VIDEO LINK AT YOUTUBE

<https://youtu.be/Q9kACozQ4OE>

ACKNOWLEDGMENT

First and foremost, I would like to thank God Almighty for giving me the strength, knowledge, ability and opportunity to finish this paper and to persevere and complete it satisfactorily. Without his blessings, this achievement would not have been possible.

I would also like to thank Mrs. Masayu Leylia Khodra, Mr. Rinaldi Munir, and Mrs. Nur Ulfa Maulidevi as our lecturers in this course for the meritorious knowledge that they have shared to us over the semester. Also, I would like to express my deepest gratitude to my family and friends who have helped and supported me while this paper was created. In addition, I would also like to thank *Worldometers*, as their contribution in providing verified COVID-19 cases datasets and an in-depth statistical data analysis.

REFERENCES

- [1] Weiss, Sabrina (2020). *How Coronavirus Started and What Happens Next, Explained*. Retrieved from <https://www.wired.co.uk/article/china-coronavirus> 1st May 2020
- [2] Limited, Bangkok Post Public Company. *Indonesia confirms first cases of coronavirus*. Bangkok Post. Retrieved 30th April 2020.
- [3] *Using a delay-adjusted case fatality ratio to estimate under-reporting*. CMMID Repository. 22nd March 2020. Retrieved 1st May 2020.

- [4] Freeman WM, Walker SJ, Vrana KE (January 1999). *Quantitative RT-PCR: pitfalls and potential*. BioTechniques.
- [5] Torres RJ, Garcia MG, Puig JG (December 2012). *Carrier and prenatal diagnosis of Lesch-Nyhan disease due to a defect in HPRT gene expression regulation*. Gene.
- [6] Indian Council of Medical Research. (2020). *Advisory on feasibility of using pooled samples for molecular testing of COVID-19*. Retrieved from <https://www.mohfw.gov.in/> on 2nd May 2020
- [7] Poste G (May 2001). *Molecular diagnostics: a powerful new component of the healthcare value chain*. Expert Review of Molecular Diagnostics.
- [8] Williams, Jr., Louis F. (22 April 1976). *A modification to the half-interval search (binary search) method*. Proceedings of the 14th ACM Southeast Conference. ACM.
- [9] Anany V. Levitin. (2002). *Introduction to the Design and Analysis of Algorithms*. Addison Wesley.
- [10] Karatsuba, Anatolii A. Yuri P. Ofman. (1962). "Умножение многозначных чисел на автоматах". Doklady Akademii Nauk SSSR. 146: 293–294. Translated in *Multiplication of Multidigit Numbers on Automata*. Soviet Physics Doklady.

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Surabaya, 2 Mei 2020



Fakhruddin Clarendia Widodo
13518091