

The prediction of mobile data traffic based on the ARIMA model and disruptive formula in industry 4.0: A case study in Jakarta, Indonesia

Ajib Setyo Arifin, Muhammad Idham Habibie

Department of Electrical Engineering, Faculty of Engineering, Universitas Indonesia, Indonesia

Article Info

Article history:

Received Apr 23, 2019

Revised Jan 5, 2020

Accepted Feb 21, 2020

Keywords:

Capacity planning

Disruptive formula

Industry 4.0

IoT

Prediction methods

ABSTRACT

Disruptive technologies, which are caused by the cellular evolution including the Internet of Things (IoT), have significantly contributed data traffic to the mobile telecommunication network in the era of Industry 4.0. These technologies cause erroneous predictions prompting mobile operators to upgrade their network, which leads to revenue loss. Besides, the inaccuracy of network prediction also creates a bottleneck problem that affects the performance of the telecommunication network, especially on the mobile backhaul. We propose a new technique to predict more accurate data traffic. This research used a univariate Autoregressive Integrated Moving Average (ARIMA) model combined with a new disruptive formula. Another model, called a disruptive formula, uses a judgmental approach based on four variables: Political, Economic, Social, Technological (PEST), cost, time to market, and market share. The disruptive formula amplifies the ARIMA calculation as a new combination formula from the judgmental and statistical approach. The results show that the disruptive formula combined with the ARIMA model has a low error in mobile data forecasting compared to the conventional ARIMA. The conventional ARIMA shows the average mobile data traffic to be 49.19 Mb/s and 156.93 Mb/s for the 3G and 4G, respectively; whereas the ARIMA with disruptive formula shows more optimized traffic, reaching 56.72 Mb/s and 199.73 Mb/s. The higher values in the ARIMA with disruptive formula are closest to the prediction of the mobile data forecast. This result suggests that the combination of statistical and computational approach provide more accurate prediction method for the mobile backhaul networks.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Ajib Setyo Arifin,

Department of Electrical Engineering,

Faculty of Engineering, Universitas Indonesia,

Depok 16424, Indonesia.

Email: ajib@eng.ui.ac.id

1. INTRODUCTION

The total mobile data traffic generated by telecommunication technologies has been significantly contributing to the core network in recent years. This has led to a congestion problem, especially in mobile backhaul technologies, which play a significant role in bringing traffic to the core network. If the mobile backhaul is congested, the operator performance may return a packet drop or higher latency, where affecting the end user indirectly. Besides, as the upcoming Industry 4.0 has already been introduced in several countries, mobile

data traffic now might be difficult to predict. This issue can lead to erroneous predictions that might exert a negative impact, such as high Capital Expenditure (CAPEX) investment caused by incorrect capacity planning.

Disruptive technologies created by innovative business models have changed existing business to evaluate their business model by adapting new technologies to monetize new revenue [1][2]. According to study, the probability of creating a successful business with an innovate approach is 10 times higher than using a non-innovative approach [3]. This knowledge has motivated incumbent businesses to change their business models according to Industry 4.0, which introduces more automation and data-exchange systems. Therefore, in recent years, new or existing industries offer new small business model, simplified products and services, and efficient solution. This has contributed significantly to mobile data traffic as well.

The large capacity pioneered by The Third Generation (3G) and Fourth Generation (4G) have contributed significantly to mobile backhaul capacity recently. Another upcoming technology is The Fifth Generation (5G), which is the latest generation of telecommunication networking and may officially be launched in 2020. The 5G, which is named as New Radio (NR) in radio access, has a downlink and uplink rate of 20 Gb/s and 10 Gb/s respectively [4], according to the International Telecommunication Union (ITU) [5]. The 5G systems also have working frequencies in the millimetre wave range; such minimal distances are causing infrastructure to become denser. The last and most critical factors that might contribute to the network involve population statistics. The Indonesian population has increased by 5.6% over the last five years and is expected to continue to rise [6]. Accordingly, the number of people assumed to be the active users of application/web services will also rise, in turn increasing traffic in the network. This can be correlated with the number of smart phones and International Mobile Subscriber Identities (IMSI) which have already provides much traffic to the traffic congestion [6].

Based on these factors, a huge capacity is urgently needed on the backhaul to support this kind of technology. By summing the traffic peak rate of these integrated technologies, we expected a 21 Gb/s pipeline to be supported in the backhaul. Current mobile backhaul technologies, which use an outdated microwave system, support around 150-300 Mb/s in each link; this should be changed to support a higher capacity by, for instance, using fiber optics or millimeter-wave links of up to 1 Gb/s.

This paper aims to predict the forecast traffic based on current technologies using second generation (2G), third generation (3G), fourth generation (4G), and to implement some IoT devices to predict the short forecast for the next few years of mixed traffic (2G, 3G, 4G, LTE, IoT, and 5G). Some researchers have shown that ARIMA model has been widely used in many traffic forecasting cases, for instance, determining Traffic Channels (TCH) in GSM obtained in NMS in China [7] and determine Time Series value for Wi-Fi data [8].

Other references show that ARIMA able to be combined with different techniques to improve the accuracy of the predicted traffic in the session: For example, the prediction traffic of the 802.11 Wireless Local Access Network (WLAN) using Seasonal Autoregressive Integrated Moving Average (SARIMA) [9]. In another case, the ARIMA model was combined with the Travel Distance Algorithm (TDE) to predict road vehicle traffic to obtain more accurate real-time traffic data [10]. Besides, It is also well-known that ARIMA is also used in for currency prediction [11].

In this paper, we attempt to predict the forecast using a univariate ARIMA model based on autoregression, moving average, and differentiated principles combined with disruptive formulas to make the forecast more accurate. Combining a statistical method with a judgmental technique (using disruptive equations) might improve the accuracy of the predicted traffic, but sometimes this combination might leads to the erroneous trend [12, 13]. However, research has shown that using a conventional ARIMA model yields a a large error rate compared with the ARIMA combination model with another technique [14, 15]. Besides, the ARIMA model is categorized as a short forecast, where it predicts a few time-series after data set. The disruptive formula analysis was based on the time to market, value proposition, cost, and PEST analysis [3].

The main motivation of this research was to analyze and predict the disruptiveness of integrated traffic by considering a low cost and secure bandwidth solution. Ideally, preparing an unexpected traffic in the radio access network mobile operators should invest a larger pipeline. But these operators are mostly saving investment cost in the radio access network. Therefore, this research attempted to predict the traffic in the mobile backhaul precisely with the cost and bandwidth consideration.

This paper is organized as follows: Section 2 gives a literature review of forecasting methods that support a basic understanding of this research. Section 3 explains the methodological points of traffic forecasting that were used in this research. Section 4 and 5 present the results and discussion of this research, respectively, including the analysis and some findings. The last section provides a conclusion of this work.

2. FORECASTING METHODS

Several techniques have been proposed to predict technological disruptiveness, especially in Industry 4.0. The first technique is called Forward-Citation Node Pair Algorithm (FCNA), which was introduced by Changwoo Choi and Yongtae Park [16], and uses a patent-citation matrix consisting of a set of nodes connected by arcs. This technique aims to identify the main development path of the complex patent-citation by understanding both present and past technologies [17], where the leading technology possesses the main patents that are linked to the selected arcs. The other technique, to improve FCNA, is K-core analysis, which concentrates on the sub-groups' nodes rather than on the main patents [17]. This technique aims to remove the central patents (which are assumed to be the most disruptive technologies) by distributing them into different subgroups that can help identify the essential data [18]. The last technique to identify disruptiveness in Industry 4.0 is called topic modelling and is similar to a search engine. Topic modelling is mostly used in the cluster that has been defined in the K-core analysis. The highest number of repetitive words in each cluster defined in K-core analysis leads to the most important aspects of the technology. Afterward, to validate the results raised by this technique, it is recommended that two experts review them [17]. All these techniques are mostly aimed at identifying the most important disruptive technologies in the market through clustering. Such techniques help market leaders identify which technologies are more disruptive, but they do not determine how much traffic each will contribute. Based on the analysis of these techniques, the major contribution technologies are IoT, artificial intelligence, financial technology (including blockchain), virtual reality, and autonomous vehicles.

2.1. Types of forecasts

Judgmental methods are based on intuition, personal interest, and user experiences [19]. One example of a judgmental method is the Delphi method, which employs a panel of experts to analyze research results to ensure validity. A judgmental method will also be used in this prediction to analyze the disruptive formula using a risk-factor technique. Univariate methods depend on past and present values that have been forecasted in a single series [19]. Univariate methods are used in this research to analyze the predicted traffic forecast. The ARIMA model consists of both univariate and multivariate models. Multivariate models use more than one independent variable (time series) simultaneously to predict the forecast. These variables might comprise interrelationships using a different time variable.

2.2. The ARIMA model

ARIMA stands for Autoregressive Integrated Moving Average [19]. The ARIMA model is based on the Box and Jenkins method of using three different concepts: Autoregression (AR), Moving Average (MA), and integration, together classified as an ARIMA(p, d, q); p defines the AR; d defines the differential; and q defines the MA. AR is a technique for analyzing the past and present values of a data set. AR is denoted as p , where it shows the weighted linear of sum p values based on ARIMA (p, d, q) terminology. The p value indicates the number of order. The formula to denote this AR is shown in (1):

$$y_t = \Theta_1 y_{t-1} + \Theta_2 y_{t-2} + \dots + \Theta_p y_{t-p} + \epsilon \quad (1)$$

where p is used to determine the number of orders of past values; t is the time series; Θ is the slope coefficient of the weighted past values; and y is the time-series function of the ARIMA model. The error term ϵ is normally distributed with mean zero and variance ϵ^2 .

The MA process is denoted by order q in the ARIMA (p, d, q) classification, which shows an error value ϵ in (1). The error term ϵ is normally distributed with mean zero and variance ϵ^2 . MA also uses the number of orders in the past values, as denoted in (2):

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} \quad (2)$$

where t is the time series; θ is the slope coefficient of the weighted past value; θ is the number of orders needed to identify the past values; and y is the time-series function of the ARIMA model. To identify how many orders are in the calculation of AR, the parameter of q is used. MA has been used for stock trading. MA aims to eliminate the noises or peaks from random noise fluctuations in the graph, where it leads to the erroneous prediction. To calculate the average value in the chart, MA takes a certain period, such as seven values, to calculate the average as shown in Figure 1.

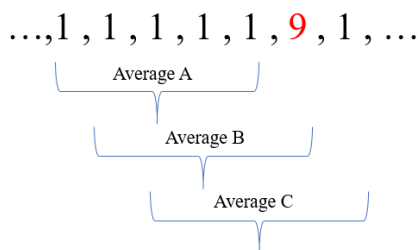


Figure 1. The MA concept example

Figure 1 describes an MA concept where average A, B, and C are calculated. The results of A, B, and C are 1, 2.6, and 2.6, respectively. The average A shares the same average value with each value in the sequence, which is one. Due to inconsistent sequences in series six, the averages B and C have different values compared to A. However, instead of using an original sequence value that has a significant peak, among others, the MA shows a smooth graph. Otherwise, Integrated or differentiated versions are denoted as d in ARIMA (p, d, q) , which is defined as the parameter that checks whether the graph is stationary [20]. As a best practice case, the time-series graph is mostly non-stationary. Therefore, MA and AR are not sufficient to determine the prediction. Being non-stationary might cause problems that can lead to error prediction. Therefore, a differential or integrated model is one of the best techniques available to make graphs stationary [19].

3. TRAFFIC FORECASTING FORMULATION

The ARIMA model is a statistical model that predicts the forecast based on past and present values. The statistical model might be inaccurate if new technologies or trends are affecting the graph, making them more disruptive. Since a judgmental approach can predict the future more accurately [13], this paper proposes a new closest prediction approach to identify traffic using a combination of judgmental and statistical approaches. To use both approaches, three important procedures must be completed: Analysis of the data set for 3G and LTE traffic using the ARIMA model; generation of a disruptive formula based on environmental technologies in the particular country; validate the disruptive formula combined with the ARIMA model.

3.1. Determining the data set

The data set was obtained from one of the busiest traffic site in Indonesia where 2G, 3G, and LTE have been installed. We took an example from the capital city of Indonesia, Jakarta, where LTE was installed in the first month of 2017. This site is strategically located in the capital city, where internet penetration is relatively greater than other locations. Besides, this site also represents the possibility to implement the IoT and the 5G network. LTE and 3G mobile data traffic data were collected for one year. Because of mobile operators have not invested in 2G technology anymore, this was assumed to be stagnant and excluded in the ARIMA and disruptive calculations. Based on the Cisco Visual Networking Index, the penetration of 2G devices has been decreasing gradually over the last five years [21]. Therefore, we assumed that 2G mobile traffic consumes a maximum of 2 Mb/s at each site based on maximum capacity, where traffic remains stable. IoT sensors were also excluded in the calculation because Indonesia just will launch the Narrow-band IoT License in Indonesia in the latest 2018, where IoT just officially starting in the next year. However, it is predicted that 400,000 IoT sensors might be installed by 2022 [22]. This trend is similar to that for 5G systems, which are expected to expand by 2020, especially in Indonesia [23]. Therefore, this research made several assumptions about these technologies (2G, 5G, and IoT), whereas 3G and LTE were more clearly demonstrated.

3.2. A new disruptive formula

The peak data rate of LTE and 3G, which are 300 Mb/s downlink and 63 Mb/s downlink (with 3rd carriers) respectively, have not shown a real fact that full throughput is used. Even 4G and 3G have 300 Mb/s and 63 Mb/s in the downlink, only approximately 20% of the full throughput would be used [24]. As a result, traffic is unpredictable and might halt one day, depending on human behaviour and company profiles. This paper proposes a new modification to the ARIMA formula. A new disruptive formula was defined as a judgemental method that might involve experts to identify the results. The combination of the judgemental and analytical methods can improve the accuracy of the prediction [13]. We proposed a new formulation, to predict disruptiveness in the future:

$$y_t = (\Theta_1 y_{t-1} + \Theta_2 y_{t-2} + \dots + \Theta_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}) 2^D. \quad (3)$$

where D depends on four variables: time to market (TTM), Cost, Politics, Economics, Social, and Technological (PEST), and Market Share. The value of disruptiveness ranges by $0 \leq D \leq 1$. The value of two and the disruptiveness range were inspired by the Global mobile data traffic forecast in [28], where it defines data traffic which never reached more than two times compared to the previous year every year; this motivated the creation of Equation (3) to define the D formula, which applies to both mobile backhaul and mobile backbone traffic. In this paper, we proposed an ARIMA model as a legacy formula for modification. As a consequence, the results obtained from the ARIMA model were amplified using the disruptiveness formula based on four variables. As a result, the result of the time-series function returns Equation (3), which is based on the disruptiveness and legacy formula.

We can immediately see that the Global mobile data traffic forecast until 2021 increases gradually each year [25]. The average increase over five years is around 47.4%. We extract the incremental traffic each year as illustrated in Table 1. The incremental traffic each year motivated us to analyze more deeply how to determine the disruptive formula. Table 1 shows that the mobile traffic trend never reached around two times compared to the previous year, which was later defined as the maximum disruptiveness. However, due to unpredicted technologies, mobile data traffic might not increase at all. Therefore, the range of disruptiveness values applies from 0 to 1, where 1 defines the maximum disruptiveness by two times the formula, and 0 defines the minimum disruptiveness, which remains the same.

Table 1. Incremental traffic based on the Global mobile data traffic forecast [25]

| Year | Incremental Traffic (%) |
|------|-------------------------|
| 2017 | 54.14 |
| 2018 | 54.54 |
| 2019 | 41.18 |
| 2020 | 45.83 |
| 2021 | 40 |

The disruptiveness value is a judgemental method based on current behaviour represented in the TTM, cost, PEST, and market share variables. These variables might affect the disruptiveness value in the formula. TTM is the period during which a product has been agreed upon and resources have been committed to a project. The TTM is divided into two variables: impact and probability. The length of the TTM gives it flexibility to decrease and increase significantly, depending on time-related processes [26]. The simpler the products or services provided, the shorter the TTM will be. The impact of the implementation products/services and the probability that it will be developed affect the value of the TTM. Table 2 shows how scoring the TTM to define the disruptiveness formula. Cost defines the total cost, including the variable and fixed costs and even operational and capital costs. The main general cost discussed here is the amount needed to create a new project [27]. The cost is divided into two different variables: impact and probability. The cost variable's value ranges from 0 to 1; its definition is provided in Table 2. PEST is considered in this disruptive technique to determine the performance and activities of businesses, especially in the long term [28]. It is clear that PEST might affect technology, especially when the technology is legalized. For example, a new technology might not be implemented in a new project if the relevant authorities do not allow it; this will in turn affect the implementation of the new technology. Table 2 shows the PEST correlation in this variable which is divided by two; PEST Impact and PEST Probability. As with TTM and cost, this PEST value ranges from 0 to 1. Market Share variable represents the percentage of market share of an industry markets total sales over a certain period. The market share is very important to determining the level of competitiveness among competitors. Table 2 shows that the market share was divided into two variables: market share impact and market share probability. The market share ranges from 0 to 1; each definition is shown in Table 2.

The formula used to identify disruptiveness incorporates the four variables as shown in Table 2. These four variables were used to identify the value of the disruptiveness, as shown in (3). Each variable has its own weight which affect the disruptiveness formula in (4).

$$D = \left(\frac{4\text{Cost} + 3\text{TTM} + 2\text{PEST} + \text{Market Share}}{10} \right)_{\text{Impact}} + \left(\frac{4\text{Cost} + 3\text{TTM} + 2\text{PEST} + \text{Market Share}}{10} \right)_{\text{Probability}} \quad (4)$$

The range of the disruptiveness value is defined in Table 2. Each variable has its own weight or priority, as shown in Table 2. The priority of every variable such as Market Share, PEST, TTM, and Cost have different values as shown in Table 2, which will affect the formula of disruptiveness in (4). It is assumed that the variable of Cost leads as a first priority to affect disruptiveness, whereas market share has the lowest priority. The first or last priority identify the weight value in each variable which leads to the final formula of disruptiveness, as expressed in (4).

Table 2. The risk factor and variables in detail

| 2*Variable | 2*Priority | 2*Disruptiveness | Risk Factor | |
|----------------|------------|------------------|---|---|
| | | | Impact | Probability |
| 2*TTM | 2*2 | Score 1 | TTM runs shorter as the impact of the implementation product/services is LARGE | The probability that a SHORT TTM for these products/services are implemented in the network |
| | | Score 1 | TTM runs longer as the impact of the implementation product/services is SMALL | The probability that a LONG TTM for these products/services are implemented in the network |
| 2*Cost | 2*1 | Score 1 | Total Cost compared to the impact of the Value Proposition. HIGH means still affordable | The probability that the products will sell to the customer. HIGH means still affordable |
| | | Score 0 | Total Cost compared to the impact of the Value Proposition. LOW means relatively not affordable | The LOW probability of the technology to offer support in terms of PEST |
| 2*PEST | 2*3 | Score 1 | Having a LARGE impact on PEST for supporting these technologies | The HIGH probability of the technology to offer support in terms of PEST |
| | | Score 0 | Having a SMALL impact on PEST for supporting these technologies | The LOW probability of the technology to offer support in terms of PEST |
| 2*Market Share | 2*4 | Score 1 | The HIGH impact of the market share on subscribers | The percentage of market share to lead others |
| | | Score 0 | The LOW impact of market share on subscribers | The percentage of market share to lead others |

3.3. Forecasting error management

The analysis of this formula compares the global mobile data forecast prediction with the model calculation used in this paper. The disruptive formula was analyzed using a percentage error comparing between conventional ARIMA and ARIMA with disruptive formula, which are defined as follows:

$$\%Error = \left| \frac{x - y}{y} \right| 100\%. \quad (5)$$

where x refers to the Global mobile data traffic forecast, and y refers to the type of model used—in this case, the conventional ARIMA and the ARIMA with disruptive formula. The percentage error defined the significance of an error compared to the mobile data traffic, where the lowest error rate led to better performance in deciding the forecast value.

4. 3G AND 4G FORECASTING

The formula of disruptiveness combined with ARIMA has been defined. This section will explain two main analyses: 3G and 4G forecasting. Based on the mobile data traffic data set in Figure 2, the 3G traffic reached the highest payload traffic in September 2016. However, during the subsequent year, 3G traffic reduced slowly in March 2017, and remains stable afterwards.

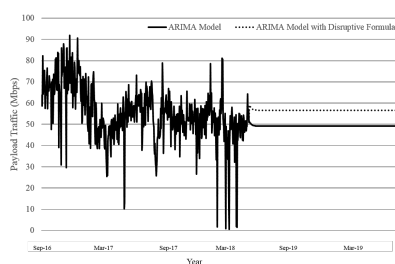


Figure 2. Comparison between ARIMA model and ARIMA with disruptive formula

Based on the ARIMA calculation, defined in the black line in Figure 2, the prediction of this traffic shows a small decrease in the first month, after which it remained stable until the end of March 2019. The ARIMA calculation used an ordered ARIMA (2,1,3), where it defined order two for AR, one for differentiating, and three for MA. The order calculation is calculated by data analysis tools, which is called R studio, to support predictive analysis using a Akaike Information Criterion (AIC).

Based on the combined ARIMA with a disruptiveness formula, as shown in the dotted line in Figure 2, the graph shows a significant value on ARIMA with a disruptive formula compared to the conventional ARIMA. The disruptiveness value was based on the cost, TTM, PEST, and market share variables, which are defined in Table 3.

Table 3. The disruptive variables in the 3G and 4G Networks

| 2*Variable | 3G Network | | 4G Network | |
|--------------|------------|-------------|------------|-------------|
| | Impact | Probability | Impact | Probability |
| TTM | 0.2 | 0.2 | 0.5 | 0.7 |
| Cost | 0.4 | 0.8 | 0.7 | 0.5 |
| PEST | 0.3 | 0.8 | 0.5 | 0.5 |
| Market Share | 0.6 | 0.6 | 0.7 | 0.7 |
| <i>D</i> | 0.34 | 0.6 | 0.6 | 0.58 |

Based on the results in Figure 2, there was a difference in traffic of around 8 Mb/s over the year. We conclude that from 2018 to 2019, based on Table 3, the PEST and cost variables, especially with respect to probability, showed a significant value, which reached 0.8 from 1.0. This might have caused the 3G mobile network penetration, which is still promised for technology implementation across the Indonesian islands. In fact, Base Station (BS) are still unavailable on several islands especially rural areas. Therefore, 3G might be preferred for application compared to other technologies regarding PEST and cost probability. However, PEST might not affect the overall disruptiveness values much, since it is a third priority, after cost and TTM.

The TTM variable illustrated in Table 3 shows a small value of 0.2 from 1.0, which was caused by the 3G network trend in 2019. Since mobile network trends are moving towards LTE and 5G networks in 2019 to support low-power devices, 3G might not be preferable for implementation in mobile traffic.

Figure 3 shows from the trends in the data set that LTE mobile data traffic significantly increased from March 2017 until the end of November 2017, but then remained stable until March 2018. The prediction shows constant data traffic over a year starting from March 18. Based on two calculations, the trend prediction analysis using ARIMA and the ARIMA with disruptive formula as depicted in the graph in Figure 4 after March 2018, shows different payload traffic around 40 Mb/s over a year.

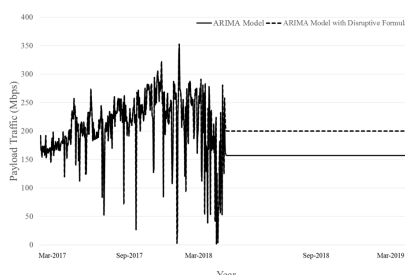


Figure 3. Comparison of 4G mobile traffic with a prediction

Based on ARIMA as shown in the black line in Figure 3, the prediction traffic shows a small decrease at first and then remains stable over a year. The ARIMA model used in this formula used ARIMA (1,1,1) based on the AIC calculation in R Studio. Based on ARIMA with disruptive formula in the dotted line from Figure 3, the effects of four variables introduced in Section IV.B show a significant contrast with the ARIMA formula. The difference in value between ARIMA with disruptive formula and the ARIMA model is approximately 40 Mb/s over the time series. During this year, the four variables mostly had the same average for the Impact and Probability.

In the 4G network, based on Table 3, the market share leads in 4G using a relative higher value compared to other variables, which is 0.7 for both probability and impact. This is mainly caused by 4G network penetration, which increased relatively from 2018 to 2019. The TTM Probability and cost Impact had a value of 0.7 from 1.0, since 4G is an affordable technology to support higher traffic. A higher cost for 4G might still be preferable if compared to the impact of this technology on users, where most people and some devices are using more data than in previous years. This might cause the cost impact to be higher than others.

The TTM Probability, showing 0.7 from 1.0 in Table 3, was caused by the impact of this technology as well. The behaviour of people in 2019 is expected to support digitalization technology that consumes more data traffic in the network. This will lead to shorter TTM to support the mobility devices in the network.

5. RESULT AND DISCUSSION

The simulation results using ARIMA model and an ARIMA combination with disruptive formula have been described in Figure 2 and Figure 3, respectively. The increasing traffic using a disruptiveness formula for 3G and 4G technologies significantly escalated the base value of ARIMA model by 8 Mb/s and 40 Mb/s, respectively. The four variables were deemed more promising for accurate prediction compared with using the ARIMA model. The ARIMA calculation, which was based on past and present values and MA, might generate inaccurate predictions if disruptive technologies are not considered. To assess this issue, this study utilised a percentage error that compared between conventional ARIMA and ARIMA with disruptive formula with the global mobile data traffic forecast.

5.1. Error performance

The Error performance subsection aims to compare the percentage error between two models in 3G and 4G traffic based in the results obtained in Section 4. Based on (5), the global data traffic variable used a global mobile data forecast [28], which was identified using a data set multiplied by the incremental value in Table 1 in the year 2018-2019. Additionally, the model calculation in (5) applied the conventional ARIMA and ARIMA with disruptive formula.

The global mobile data traffic calculated the average mobile data traffic obtained from the data set in 3G and 4G—in this case, from 2016 to 2018. The data set was averaged over 2016 to 2018 and multiplied using an incremental value based on Table 1, which was 41.17% from 2018-2019. By obtaining the data set traffic and the incremental value from this data, the optimized prediction traffic was obtained, which is shown in Table 4. Table 4 shows optimized traffic based on Mobile Data Forecast, where it increases 41.17% from 2018 to 2019. The data traffic in 2019 based on Table 4 will be compared to the prediction based on Conventional ARIMA and ARIMA + Disruptive Formula. It assumed that the global mobile data traffic forecast has more accurate prediction based on several types of research.

Table 4. Traffic comparison between 3G and 4G in 2018-2019

| Technology | Year | |
|------------|-------------|-------------|
| | 2018 | 2019 |
| 3G | 55.06 Mb/s | 77.63 Mb/s |
| 4G | 207.87 Mb/s | 293.09 Mb/s |

In 3G, the ARIMA with disruptive formula reaches 56.67 Mb/s, which is almost optimized to 72.17 Mb/s in Table 4. The error rate seemed to be lower than in the conventional ARIMA. Besides, in 4G, the ARIMA with disruptive formula reaches 199.6 Mb/s, whereas conventional ARIMA reaches 156.93 Mb/s, which is far higher than the global mobile data traffic, 293.09 Mb/s in Table 4. The error rate in both values is caused by several factors, i.e.: backhaul traffic and developing country factors. The calculation of the ARIMA model used backhaul traffic, where it seemed to be more negative compared to the global data forecast.

However, the lower error rate in ARIMA with disruptive formula is more promising compared to the conventional ARIMA, where it defined more optimized values from 2018 to 2019. This demonstrates that the ARIMA with disruptive formula had more accurate prediction compared to the conventional ARIMA. Internet penetration in developing countries has increased relatively slowly compared to developed countries, which might affect mobile data traffic. In this case, compared to global mobile data traffic, the conventional ARIMA shows a small increase over a year for 3G and 4G, as shown in Figures 3 and 4. However, the ARIMA with disruptive formula was more positive, which is closely related to the mobile data traffic trend and the internet implementation program across Indonesia. This led to more accurate prediction.

5.2. Variable analysis

The four factors that have been defined-TTM, Cost, PEST, and Market Share-significantly affected the conventional ARIMA. As explained earlier, the variable cost had maximum priority over others, while market share was less important. To validate the variables, we identified the different levels of each variable using the maximum incremental traffic of Impact and Probability. The maximum order of Impact and Probability, M , can be expressed,

$$M = N \times \max(\text{Impact} \times \text{Probability}). \quad (6)$$

where N is equal to the number of weight, which is identified in Table 2. For example, the variable cost has four maximum weights, where the variable market share has 1 maximum weight. Besides, Impact and Probability are assumed to be at maximum probability, which is equal to 1. It also assumes that the other variables are zero if a particular variable is calculated.

Based on (6), Table 5 shows different maximum orders in each variable. Moreover, Figure 4 took an example in the 3G network, showing the different incremental traffic using the four variables identified in Table 5. The incremental traffic shows a different prediction was used in the ARIMA model.

| Disruptive Variables | Incremental Traffic (%) |
|----------------------|-------------------------|
| TTM | 9 |
| Cost | 16 |
| PEST | 4 |
| Market Share | 9 |

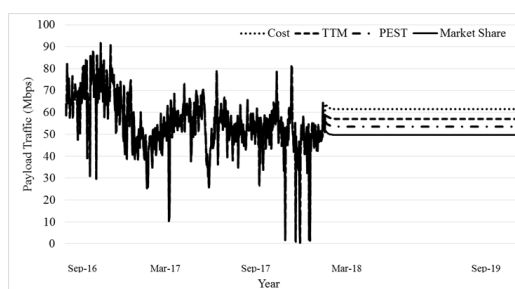


Figure 4. Comparison of the different mobile networks

Based on Figure 4 and Table 5, the variable cost will affect a maximum 16% incremental value compared to the conventional ARIMA. Cost is cross-related to the revenues of companies. This is reasonable since if we imagine that companies have a significant amount of revenues to utilize the initial and maintenance costs for Capital Expenditure (CAPEX) and Operational Expenditure (OPEX), they will prioritize customer demands, including data traffic speed and latency, which are essential for users. By this variable, in the Indonesian case, the operators have an opportunity to either spread the mobile base stations into different locations or to make regular networks denser to increase their capacity. The cost might correlate with other variables, such as PEST, TTM, and Market Share. If mobile operators have more revenues, this will affect disruptiveness and other variables. For example, the revenues will consider the frequency allocation that has been determined by the regulators in each particular country, where more revenues will probably decide more frequency allocations.

TTM and PEST are the second and third priorities in the disruptiveness variables. The maximum total incremental traffic is 16% and 9% , respectively, as shown in Figure 4. The example real-world case of this variable is the license readiness to implement new frequencies and technologies in the country. For example, the main challenge for the license readiness example is the millimetre wave in 5G technology, where several steps are needed to assess the frequency spectrum in their country. The new spectrum in millimetre wave should consider permission to open a new license spectrum. This license is also a dependent factor with the cost profile of the mobile operators, where having a new spectrum leads to higher risk of spending at more considerable cost.

Besides, technology readiness is also a dependent factor with the cost profile of mobile companies, where technology readiness might be delayed if traffic penetration in the country is not covered 100%. As an example, in Indonesia, the LTE mobile stations will be implemented later, since 3G stations have not yet been implemented across the whole of Indonesia. Therefore, to support more efficiency, 3G is still preferable to LTE, which reduces more CAPEX and OPEX, for new stations. Both cases might cause a delay in technology implementation for the country, where both variables still depend mostly on cost. As a result, cost is still the highest priority affecting mobile data traffic.

Market share, the lowest priority, determines only 1% of the incremental traffic. Market share does not affect mobile data traffic very much if traffic and revenues are relatively increasing inversely, which has occurred in the telecommunication environment. This makes market share the lowest priority in the disruptiveness value. To conclude, four variables are the main factors of disruptive traffic, with cost/revenues being the most dominant factors that affect disruptive traffic.

5.3. Backhaul analysis

The arrival of disruptive technologies will affect the total capacities in the mobile backhaul. As shown in Section 5, 4G and 3G traffic have amplified traffic in the conventional ARIMA, around 35 Mb/s and 10 Mb/s, respectively. Other technology, such as 2G systems, which is assumed to be stagnant, consumes around 2 Mb/s each year. Besides, the 5G network has not been implemented yet, and while IoT systems are increasing this year, we still assumed no IoT due to having the lowest bit rate and smallest total number of devices. Illustrating all these factors, Table 6 presents an overview of the predicted average mobile data traffic between 2017 and 2018, mainly in each site, using ARIMA with disruptive formula.

Table 6. Average Mobile Data Traffic 2017 and 2018

| 2*Technologies | Average Mobile Data Traffic(Mb/s) | |
|----------------|-----------------------------------|--------|
| | 2018 | 2019 |
| 2G | 2 | 2 |
| 3G | 54.62 | 56.67 |
| 4G | 198.03 | 199.6 |
| Total | 254.65 | 258.23 |

By calculating the ARIMA with disruptive formula model, we conclude that the mobile backhaul could support the old microwave technologies, where the future backhaul will need at least the average of 258.2 Mb/s each site based on Table 6. This capacity is basically could be supported by the existing microwave technologies. However, to anticipate the unexpected traffic in the future, this paper recommends a list of feature for the mobile backhaul, which are (from the most efficient): Using HOM that supports traffic greater than 300 Mb/s, for instance, with 1024 or 2048 QAM, implementing more antennas in the mobile backhaul systems to increase capacity, such as MIMO or Massive MIMO, migrating old microwave technologies to fiber optics. This paper finds an effective and accurate way to predict the traffic forecast based on statistical and judgemental approach. With the combination of ARIMA model and disruptive formula that this paper proposed, it has shown that ARIMA is more accurate if it is associated with a Judgemental approach to correct the errors.

6. CONCLUSION

The major contribution of the study is the development of a new formula in the ARIMA model to predict forecast traffic based on four variables: TTM, cost, PEST, and market share. Our research confirms that disruptive technology affect the mobile data traffic if: the telecommunication companies are profitable; the time to market to implement new project is acceptable, the environment of PEST supports new technologies

and mobilities, and market share has increased a new segment market. This combination also leads to the conclusion that mobile backhaul in Indonesia recommended using the existing microwave technologies in 2018. Future work will use short forecasts for the long-period method, such as the Long Short Term Memory (LSTM) network, where mobile traffic is predicted more accurately over a longer period.

ACKNOWLEDGEMENT

This work was supported by the United States Agency for International Development (USAID) through the Sustainable Higher Education Research Alliance (SHERA) Program for Universitas Indonesia's Scientific Modeling, Application, Research, and Training for City-centered Innovation and Technology (SMART CITY) Project.

REFERENCES

- [1] E. Robles, "How to Identify Disruptive New Businesses," *Rev. Buss. Fin. Stud.*, vol. 6, no. 1, pp. 121–130, 2015.
- [2] H. Taskin, M.R. Adali, and E. Ersin, "Technological intelligence and competitive strategies: An application study with fuzzy logic," *J. Intell. Manuf.*, vol. 15, no. 4, pp. 417–429, Aug. 2004.
- [3] A. Stoiciu, E. Szabo, M. Totev, K. Wittmann, and N. Hampf, "Assessing the Disruptiveness of New Energy Technologies - An Ex-Ante Perspective," *Inst. Strategic Manag. Energy and Strategy Think Tank*, Mar. 2014. [Online]. Available: http://epub.wu.ac.at/4310/1/ISM_ESTT_WP2_Assessing_the_Disruptiveness_of_New_Energy_Technologies.pdf, Accessed on: Feb. 28, 2018
- [4] S. Parkvall, E. Dahlman, A. Furuskar, and M. Frenne, "NR: The New 5G Radio Access Technology," *IEEE Commun. Stand. Mag.*, vol. 1, no. 4, pp. 24–30, Dec. 2017.
- [5] S. Anthony, "5G specs announced: 20Gb/s download, 1ms latency, 1M devices per square km," *Ars Technica*, London, UK. Feb 24, 2017. [Online]. Available: <https://arstechnica.com/information-technology/2017/02/5g-imt-2020-specs/>, Accessed on: Feb. 27, 2018.
- [6] C. Chen, Q. Pei, and L. Ning, "Forecasting 802.11 traffic using seasonal ARIMA model," *IFCSTA 2009 Proc. - 2009 Int. Forum Comput. Sci. Appl.*, vol. 2, no. 1, pp. 347–350, 2009.
- [7] R. Madan and P. Sarathimangipudi, "Predicting Computer Network Traffic: A Time Series Forecasting Approach Using DWT, ARIMA and RNN," *2018 11th Int. Conf. Contemp. Comput. IC3 2018*, no. 1, pp. 2–4, 2018.
- [8] B. P. Statistika, "Statistical Yearbook of Indonesia 2016," *Badan Pusat Statistika*. Jakarta, Indonesia. Jun. 2016. [Online]. Available: <https://media.neliti.com/media/publications/48410-ID-statistik-indonesia-2016.pdf>
- [9] C. Chen, Q. Pei, L. Ning, "Forecasting 802.11 traffic using seasonal ARIMA model," in *Proc. 2009 Int. Forum Comput. Sci. Appl.*, Chongqing, China, 2009, pp. 347–350.
- [10] A. Qonita, A. G. Pertiwi, and T. Widiyaningtyas, "Prediction of rupiah against us dollar by using arima," *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, vol. 4, no. September, pp. 746–750, 2017.
- [11] Z. Zhao, W. Chen, H. Yue, Z. Liu, "A novel short-term traffic forecast model based on travel distance estimation and ARIMA," in *Proc. 28th Chinese Control Decis. Conf.*, Yinchuan, China, 2016, pp. 6270–6275.
- [12] C. W. S. Chen, M. K. P. So, F. -C. Liu, "A review of threshold time series models in finance," *Statistics and Its Interface*, vol. 4, pp. 167–181, 2011.
- [13] J. S. Armstrong, F. Collopy, "Integration of statistical methods and judgment for time series forecasting: principles from empirical research," *Forecasting with Judgement*, New York, USA: John Wiley, 1998, pp. 269–293.
- [14] T. Hirata, T. Kuremoto, M. Obayashi, S. Mabu, K. Kobayashi, "Time series prediction using DBN and ARIMA," in *Proc. Int. Conf. Comput. Appl. Technol.*, Matsue, Japan, 2015, pp. 24–29.
- [15] P. Areekul, T. Senjyu, H. Toyama, A. Yona, "Combination of artificial neural network and ARIMA time series models for short term price forecasting in deregulated market," in *Proc. Transmission Distrib. Conf. Expo. Asia Pacific*, Seoul, South Korea, 2009, pp. 0–3.
- [16] C. Choi, Y. Park, "Monitoring the organic structure of technology based on the patent development paths," *Tech. Forecast. Soc. Change*, vol. 76, no. 6, pp. 754–768, July 2009.

- [17] A. Momeni, K. Rost, "Identification and monitoring of possible disruptive technologies by patent-development paths and topic modeling," *Tech. Forecast. Soc. Change*, vol. 104, pp. 16–29, Mar. 2016.
- [18] V. Batagelj, A. Mrvar, "Some analyses of Erdos collaboration graph," *Soc. Networks*, vol 22, no. 2, pp. 173–186 May 2000.
- [19] C. Chatfield, "Time-series forecasting," New York, USA: Chapman & Hall/CRC, 2005.
- [20] R. Nau, "Introduction to ARIMA models," Fuqua School of Business, Duke University, 2014. [Online]. Available: https://people.duke.edu/~rnau/Slides_on_ARIMA_models-Robert_Nau.pdf, Accessed on: Mar. 21, 2018.
- [21] G. Brandli and M. Dick, "Alternating current fed power supply," U.S. Patent 4 084 217, Nov. 4, 1978.
- [22] B. D. Atmoko, "Riset: Indonesia Memerlukan Laboratorium Internet of Things (IoT)," *Gizmologi*, Jakarta, Indonesia. Dec. 7 2017. [Online]. Available: <https://gizmologi.id/berita/indonesia-lab-iot/>. Accessed on: Mar. 2018.
- [23] O. Yusuf, "Negara Maju Tahun 2019, Kapan Indonesia Kebagian Internet 5G?," *Kompas*, Jakarta, Indonesia. Dec. 12, 2017. [Online]. Available: <https://tekno.kompas.com/read/2017/12/19/11040667/negara-maju-tahun-2019-kapan-indonesia-kebagian-internet-5g>. Accessed on: Mar. 22, 2018.
- [24] Motorola. "Realistic LTE performance from peak rate to subscriber experience," Motorola, Chicago, IL, USA. 2009. [Online]. Available: https://www.apwpt.org/downloads/realistic_lte_experience_wp_motorola_aug2009.pdf. Accessed on: Mar. 23, 2018.
- [25] Cisco, "Cisco visual networking index: Global mobile data traffic forecast, 2015–2020," Cisco, San Jose, CA, USA. Feb. 3, 2016. [Online]. Available: <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html>. Accessed on: Mar. 17, 2018.
- [26] J. Groh, "A 5-Step Approach to Reduce Time-to-Market," *New Product Visions*, Flat Rock, NC, USA. May 25, 2016. [Online]. Available: <http://newproductvisions.com/blog/?p=847>. Accessed on: Mar. 16, 2018.
- [27] C Harris, "Fixed and variable costs: Theory and practice in electricity," Palgrave Macmillan, USA. 2014.
- [28] J. Makos, "Understanding PEST analysis with definitions and examples," *Pestleanalysis*, USA. Dec. 31, 2013. [Online]. Available: <http://pestleanalysis.com/pest-analysis/>. Accessed on: Mar. 16, 2018.