Journal of the Korea Industrial Information Systems Research Vol. 22 No. 5, Oct.. 2017 :1-7
http://dx.doi.org./10.9723/jksiis.2017.22.5.001 ISSN:1229-3741

민간항공기 사이의 거리 분석 모델링

(Statistical Modeling of Inter-Aircraft Distances)

진 성 근[1)*], 김 진 경[2)]
(Jin Sunggeun and Kim Jinkyeong)

요 약 우리는 실제 항공 데이터를 토대로 가장 가까운 항공기간 거리를 조사하였다. 그리고, 항공기 사이의 거리가 감마파레토 분산을 따른 다는 것을 발견하였다. 우리의 발견은 항공기간 통신을 위한 무선통신시스템을 설계하거나 연구할 때 중요한 자료가 될 것으로 기대한다.

핵심주제어 : DIO가공시스템, 유전자 알고리즘

Abstract We analyze Inter-Aircraft Distances between Two Closest Flying Passenger Aircrafts on a Global Scale from Real Aviation Databases. Then, We Reveal that the Distances Follow a Gamma-Pareto Distribution. Our Finding is Useful for Designing Wireless Transceivers Since it Gives the Probability Distribution Regarding the Link Distances which the Wireless Transceivers should Cover for Providing Internet Services.

Key Words : Passenger Aircraft, Wireless Communication, Inter-Aircraft Distances

1. 서 론

There have been many efforts to provide Internet services for the passengers in passenger aircrafts such as OnAir, Gogo inflight Internet, and AeroMobile [1 - 3]. So far, most technologies have employed satellites together with ground stations or only satellites in order to maintain the wireless connectivity for the Internet services. However, thanks to recent evolution of wireless communication technologies covering very long distances [4 - 6], e.g., 200~300 km, it will eventually be possible to provide inexpensive satisfactory inflight Internet services by establishing wireless multihop links with only ground stations against the previous work in [1 - 3], [7,8].

For this purpose, we need to know how long the wireless transceivers should cover air-to-air links with practical statistics. However, there have been few studies regarding the distances. For this reason, we analyze the distances between the closest passenger aircrafts. We had collected worldwide flight data for four months. Then, we find statistical characteristics regarding the distance between an arbitrarily chosen aircraft

* Corresponding Author : sgjin@daegu.ac.kr
+ 이 논문은 2014학년도 대구대학교 학술연구비지원에 의한 논문임.
Manuscript received July 20, 2017 / revised Sep 11 ,2017 / accepted Sep 25, 2017
1) 대구대학교 컴퓨터소프트웨어학과, 제1저자
2) 한국전자통신연구원, 제2저자

and its closest aircraft all over the world. Surprisingly, we find it out that the distance follows a Gamma-Pareto distribution with negligible errors. It implies that the distances between the closest passenger aircrafts can be simply represented by the specific parameters of the Gamma-Pareto distribution.

Our finding can be further used for numerical analysis of mobility management or simulation studies when flying passenger aircrafts establish wireless relay networks for inflight Internet services. This paper is organized as follows: in Section II, we explain the statistics we have collected. In Section III, we search for the characteristics regarding the distances. Section IV concludes this paper.

2. 시스템 구성 및 실험

We had collected the flight information posted on “www.flightradar24.com” from April 13, 2014 to August 8,2014. Flightradar24 provides passenger aircraft flight information by gathering the information from three sources, i.e., automatic dependent surveillance-broadcast, multilateration, and federal aviation administration [9].

For the data collection, we have developed a software, written in python and bash script languages. It retrieves the flight data every minute and parses them for a proper analysis. The flight data includes passenger aircraft name, altitude, latitude, longitude, departure airport, destination airport etc, and hence, by using each pair of the latitude and longitude, we can calculate the distance(=D) between arbitrarily chosen two passenger aircrafts with the equation below [10]:

$$D = R\arccos(\cos(\phi_1)\cos(\phi_2)\cos(\lambda_1 - \lambda_2) + \sin(\cos(\phi_1)\sin(\phi_2) \quad (1)$$

where ϕ_1 and ϕ_2 are latitudes of two passenger aircrafts, respectively while λ_1 and λ_2 are longitudes of them, respectively. For the equation, latitudes and longitudes are converted in units of radians. R(≌6373 km) is the average radius of the earth.

For our analysis, we choose passenger aircrafts flying higher than 22000 feet since typical cruising altitude for domestic/international flight ranges between 22000/26000 and 28000/39000 feet. Note that 22000, 26000, 28000 and 39000 feet are 6.70, 7.92, 8.53, 11.88 km. In the rest of this paper, we deal with only flying passenger aircrafts higher than 22000 feet.

Meanwhile, the distances are derived based on the belief that the earth is round. However, passenger aircrafts communicate with each other on a straight-line distance. Besides, the passenger aircrafts may fly at different altitudes so that the different altitudes may cause an error for the calculation of the distance. In other words, there exist errors between the derived distances with Eq. (1) and the straight-line distances for the wireless communications.

Therefore, we derive the error ratio as follows. For the derivation, (1) we assume that one of two chosen aircrafts is located on the surface of the earth for worst case consideration. (2) We have the difference between the distance along the surface of the earth and the straight-line distance for two chosen aircrafts. After that, (3) we divide the difference by the straight-line distance derived by:

$$1 - \frac{\arccos(\frac{R}{R+h})}{\tan(\arccos(\frac{R}{R+h}))} \quad (2)$$

where h is a cruising altitude and 6:70 km ≤

$h \le$ 11:88 km. From this equation, the error ratio is maximized to about 0.001241 when h = 11:88 km. It means that we can ignore the errors in this work. For further explanations, we summarize the notations for the variables used in this paper as shown in Table I.

Table 1 Notations for the Variables

Notation	Meaning
d	Set of collected flight data
h	Altitude of an aircraft
k, θ	Parameters of the Gamma distribution
$p(n)$	Probability function for the number of aircrafts
D	Inter-aircraft distance
$G^{(n)}$	Group of aircrarts
R	The radius of the earth
α, c, θ	Parameters of the Gamma-Pareto distribution
ϕ_1, ϕ_2	Latitudes
λ_1, λ_2	Longitudes

3. Statistical Analysis

We observe the number of concurrent flying passenger aircrafts on a global scale. Fig. 1 illustrates how the average number of flying passenger aircrafts and its standard deviation varies from the beginning date, i.e., April 13, 2014 to the ending date, i.e., August 7, 2014.

In this figure, x-axis and y-axis represent the days elapsed from the beginning date and the number of passenger aircrafts in the sky, respectively. For this figure, we divide flight data for a day into two parts, i.e.,the data for 00:00 AM~12:00PM and 12:00PM~23:59:59AM, GMT+7, respectively, so that each point represents the average for each half day. In this figure, we can see a sharp drop on 27th day from the beginning date since the data collecting operation was halted for a few hours temporarily due to a technical problem of our server. However, the missing part is negligible compared with the entire data collections.

From the averages and the standard deviations, we can see large standard deviations on each day. For instance, we observe the total number of flying passenger aircrafts change every minute on June 7, 2014 in Fig. 2. This figure shows that the number of flying passenger aircrafts varies in a wide range between about 3500 and about 5200 at that day. Meanwhile, we observe that the average moves from about 3500 to about 4500 incrementally for three month. From Figs. 1 and 2, we classify the collected flight data according to the number of passenger aircrafts in the sky for further analysis, and hence, we continue to make groups of passenger aircrafts with the equation below:

$$G^{(n)} = d|n\delta \le |d| < (n+1)\delta \tag{3}$$

where δ =100 and n $\in$ N when N is set of natural numbers. d is set of collected flight data and $|d|$ is the number of the aircrafts. From this equation, we have the probability function $p(n)$ regarding the number of passenger aircrafts in the sky by:

$$p(n) = \frac{\sum_{\forall d \in G^{(n)}} |d|}{\sum_{\forall G^{(m)}} \sum_{\forall d \in G^{(m)}} |d|} \tag{4}$$

where m, $n \in$ N. Meanwhile, we plot $p(n)$ as shown in Fig. 3. It shows that at least/most 2500/5500 passenger aircrafts are flying in the sky. Therefore, we only consider $25 \le n \le 55$ for $p(n)$. Therefore, we can say that there are about $4079.85 (\approx 100 \sum_{n=25}^{55} np(n))$ passenger aircrafts in the sky on average.

Next, we investigate what distribution is the most suitable for the distances between the closest passenger aircrafts. For this purpose, we classify the obtained statistics depending on the average number of passenger aircrafts in the sky. Then, we rearrange them according to the distances between the closest passenger aircrafts.

Fig. 4 shows an exemplary histogram painted gray for the statistics at a randomly chosen day, i.e., GMT+7 0:00~12:00, 7th June 2014. In this figure, x- and y-axes represent the distance in units of kilometers between the closest passenger aircrafts and the probability value, respectively. In this figure, we can see four distributions of which parameters are selected to best represent the distribution of the collected data. Especially, the parameters of Beta, Lognormal, and Pareto distributions are chosen with the algorithms provided by Python's SciPy package [11] while those of Gamma-Pareto distribution are selected with the method detailed later. Consequently, we choose Gamma-Pareto since it is the most suitable for the distribution of the collected data as detailed later.

In this figure, we can observe that the distribution of the collected data has long tail decaying slowly when the x-range is much greater than '0.' The distribution having this feature, named *Gamma-Pareto* distribution, was derived by the authors of [12]. We assume Gamma and Pareto distributions are given by $\frac{1}{\beta^{\alpha}\Gamma(\alpha)}t^{\alpha-1}e^{-t/\beta}$ and $\frac{k\theta^{k}}{u^{k+1}}$ respectively, where $\alpha, \beta, k, \theta > 0$ and $u \geq \theta$ for random variables t and u. In this case, the *Gamma-Pareto* distribution is given by [12]:

$$f(x;\alpha,c,\theta) = \frac{k^{\alpha}}{x\Gamma(\alpha)\beta^{\alpha}}(\frac{\theta}{x})(\log(\frac{x}{\theta}))^{\alpha-1} \quad (5)$$

where c = β/k and $x > 0$. Therefore, we need to know what similarity exist between the statistics and *Gamma-Pareto* distribution. In order to find best-fit probability density function (pdf) of *Gamma-Pareto* distribution well-matched by the statistics, we adopt *exhaustive search* algorithm to find minimum mean square error provided by:

$$\arg min_{\alpha,c,\theta}(\frac{1}{M}\sum_{m=0}^{M-1}(\hat{f}(m) - f(m;\alpha,c,\theta))^2) \quad (6)$$

where M = 400. $\hat{f}(m)$ represents the probability values regarding the statistics. In this figure, the red line indicates the pdf of the found *Gamma-Pareto* distribution when α = 71.2, c = 0.11, and θ = 0.0145. Herein, we recognize that the Gamma-Pareto is well matched by the statistics with the very small value (=0.0006) of mean square error.

By adopting the exhaustive search algorithm, we find the parameters for the best-fit pdfs depending on the number of passenger aircrafts in the sky. Fig. 5 shows the obtained parameters exist within small ranges. Now, we have expected values for the parameters with Eq. (4) and the equation below:

$$(\bar{\alpha},\bar{c},\bar{\theta}) = \sum_{n=25}^{55}(\alpha^{(n)},c^{(n)},\theta^{(n)}) \times p(n) \quad (7)$$

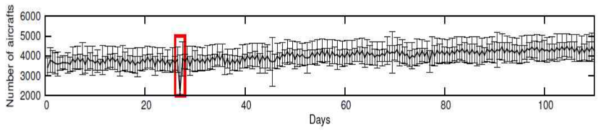

Figure 1: The number of passenger aircrafts according to the data collecting period.

Fig. 1 The number of passenger aircrafts accroding to the data collecting period

From this equation, we have the expectations by:

$$\overline{\alpha}=80.4759,\ \overline{c}=0.101546,\ \overline{\theta}=0.0122556 \qquad (8)$$

We need to evaluate how much we can rely on $\overline{\alpha}, \overline{c}$ and $\overline{\theta}$ for realistic environments. For this purpose, we have root mean square errors with standard deviations for the flight data obtained every minute shown in Fig. 6. In this figure, the root mean square errors are smaller than 0.002. It implies that we can employ $\overline{\alpha}, \overline{c}$ and $\overline{\theta}$ for the Gamma-Pareto distribution to estimate the distances between the closest aircrafts.

Finally, it is necessary to evaluate how much the statistics of the collected data is consistent with an assumed distribution. For this purpose, we conduct Kolmogorov-Smirnov (K-S) test to verify the goodness-of-fit since it is usually used to decide if a sample comes a population with a specific distribution [13]. For our K-S test, Gamma-Pareto, Gamma and Lognormal distributions are considered as reference distributions and the best distribution parameters are selected for the collected data. In the goodness-of-fit test, *p-value*, where $0 \leq$ *p-value* ≤ 1, is used as a metric representing the consistency. We can say that collected experimental statistics follow the assumed distribution if the *p-value* of the statistics is greater than a threshold, typically 0.05 or 0.01.

For Gamma-Pareto distribution, the *p-value* is 0.9964 with the obtained parameters in Eq. (8), and hence, it is acceptable the hypothesis that the closest distances follow Gamma-Pareto distribution. Similarly, the p-value is 0.4040 for Lognormal distribution. It implies that Lognormal distribution is acceptable either. Therefore, we compare root mean square errors when Gamma-Pareto and Lognormal distributions are applied to the

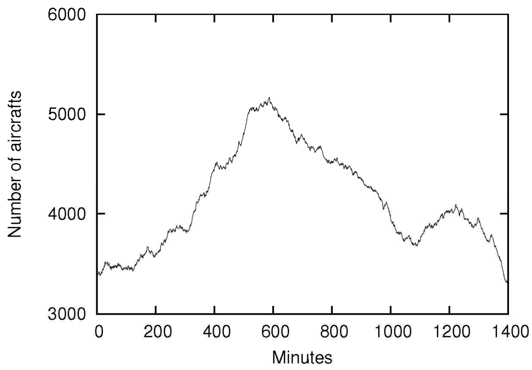

Fig. 2 The number variation of the passenger aircrafts on a global scale for a day.

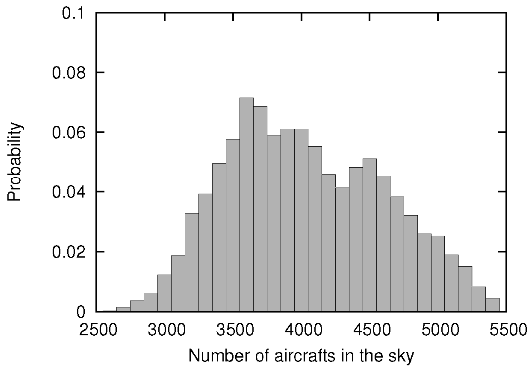

Fig. 3 Probability function $p(n)$ representing the number of passenger aircrafts in the sky.

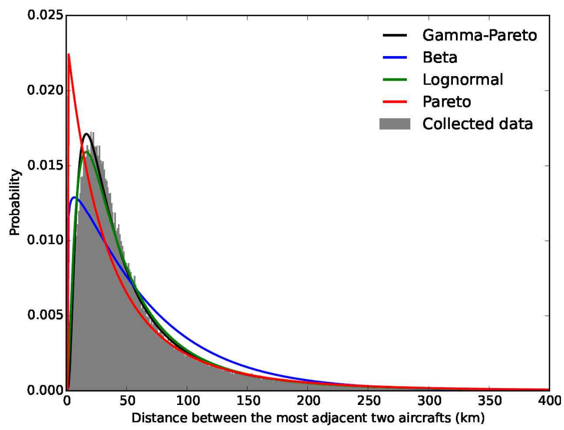

Fig. 4 Best-fit probability density function for collected data.

distributions of every half-day's flight data with the best parameters. Then, we obtain the average root mean square errors for Gamma-Pareto and Lognormal distributions by 2.30e-5 and 6.10e-5, respectively. Their standard deviations are 2.561e-6 and 6.10e-5, respectively. For this reason, we claim that Gamma-Pareto distribution is more appropriate for the distances while Lognormal distribution is also acceptable.

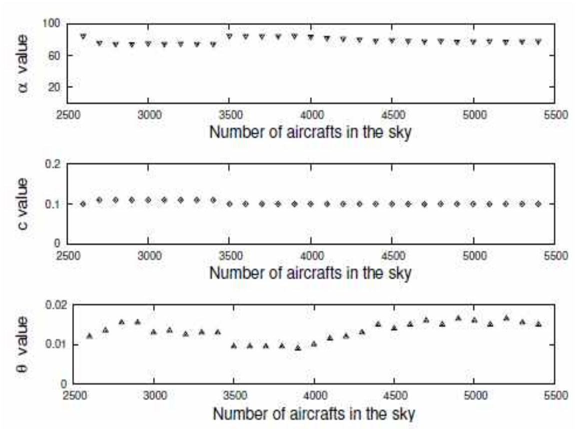

Fig. 5 Parameters depending on the number of passenger aircrafts in the sky.

4. Conclusion and Future Work

In this paper, we reveal that it is possible to approximate the distribution for the distances between the closest passenger aircrafts with Gamma-Pareto distribution. This work is helpful for the research on aeronautical ad hoc networks. From the finding, we further study what frequency bands are appropriate for the ad hoc networks, how we can develop interference models for the frequency bands, and what modulation policy we need to adopt, sequentially. Additionally, we need to consider that aircrafts' antenna may not cover all angles so that neighboring aircrafts may lose

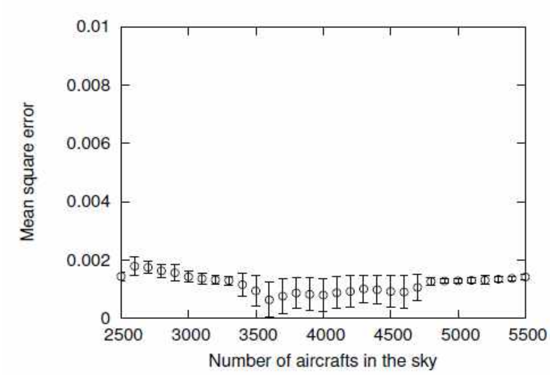

Fig. 6 Average mean square errors with standard deviations when $\overline{\alpha}, \overline{c}$ and $\overline{\theta}$ are used.

wireless connectivity in reality. Therefore, we will study the issue either. This work is also helpful for other studies including analytical modelings and simulations of aeronautical ad hoc networks. As stated in [14,15], the investments on the network services may improve the benefits of information technology companies.

References

[1] http://www.gogoair.com.

[2] http://www.aeromobile.net.

[3] http://www.onair.aero.

[4] Point-to-Point Broadband Wireless for Service Providers, White Paper. Motorola, 2007.

[5] http://www.exelisinc.com/solutions/TALON/Pages/default.aspx.

[6] Flickenger, R., Okay S., Pietrosemoli E., Zennaro M., and Fonda C., "Very Long Distance Wi-Fi Networks," in Proc. ACM SIGCOMM NSDR'08, 2008.

[7] Li H., Yang B., Chen C., and Guan X., "Connectivity of Aeronautical Ad Hoc Networks," in Proc. IEEE GLOBECOM'10, December 2010.

[8] Medina D., Hoffmann F., Ayaz S., and Rokitansky C.-H., “Feasibility of an Aeronautical Mobile Ad Hoc Network over the North Atlantic Corridor,” in Proc. IEEE SECON’08, June 2008.

[9] http://www.flightradar24.com.

[10] Junkins J. L. and Shuster M. D., “The Geometry of the Euler Angles,” JAS, Vol. 41, No. 4, 1993.

[11] http://www.scipy.org.

[12] Alzaatreh A., Famoye F., and Lee C., “Gamma-Pareto Distribution and Its Applications,” JMASM, Vol. 11, No. 1, 2012.

[13] I. M. Chakravarti, R. G. Laha, and J. Roy, Handbook of Methods of Applied Statistics, Volume I. McGraw-Hill, 1997.

[14] Park, Jin-hyung, Ok, Seok-jae, and Choi, Seung-hoon, “The Impacts of Network Communication Technologies on the Firm’s Competitive Advantage,” the Journal of Internet Electronic Commerce Research, Vol. 7, No. 4, pp. 243-261, 2007.

[15] Cho, Seonghan and Park. Dongjin, “Study on the Analysis Regarding National Profit of Communication Industry,” the Journal of Internet Electronic Commerce Research, Vol. 16, No. 1, 2007.

진 성 근 (Jin Sunggeun)

• 정회원
• 경북대학교 전자공학과 공학사
• 경북대학교 전자공학과 공학석사
• 서울대학교 전자정보공학과 공학박사
• 대구대학교 정보통신대학 컴퓨터소프트웨어학과 조교수
• 관심분야 : 무선네트워크

김 진 경 (Kim Jinkyeong)

• 정회원
• 인하대학교 전자공학과 공학사
• 인하대학교 전자공학과 공학석사
• 충남대학교 정보통신공학과 공학박사
• 한국전자통신연구원
• 관심분야 : 무선네트워크