*Applied Probability Trust* (7 July 2024)

# MASS FUNCTION OF THE NUMBER OF COIN FLIPS TO GET $M$ CONSECUTIVE HEADS

INSTITUT NATIONAL DES SCIENCES APPLIQUÉES,* *Jan Aalmoes*

**Abstract**

In this paper, we study the odds of streaks: how long must we wait before observing a string of consecutive repetitions? This problem has significant sociological and industrial applications. In baseball, consecutively hitting home runs in multiple games can be crucial for a player's career. In reliability engineering, consecutive failures of redundant parts can result in system failure. In networking, particularly in the Internet of Things, consecutive packet losses can have critical repercussions. Contrary to previous work, we compute the exact probability distribution of the waiting time before a streak of arbitrary length occurs. This allows practitioners and system designers to assess risks and choose optimal parameters accordingly.

*Keywords:* redundancy; reliability; networking; communication

2020 Mathematics Subject Classification: Primary 60G07

Secondary 65C20

## 1. Introduction

We call **streak** the repetition without interruption of an event in a stream of events. For instance, a streak of $m$ heads means that in an experiment of repeatedly throwing a coin, at some point, the coin lands on head $m$ times consecutively. Streaks and their occurrences is of interest in many fields: sports, physics, or networking. In baseball, successfully repeating an exceptional action such as hitting a home run can be decisive for a player's career [1].

In reliability engineering the probability of system failure is studied through statistical analysis. Redundancy is then used to reduce this probability [5]. If $m$ is the

* Postal address: 20, avenue Albert Einstein - 69621 Villeurbanne CEDEX, France

* Email address: jan.aalmoes@protonmail.com

number of redundant parts, an occurring streak of size $m$ means that the system has failed. Especially in spacecrafts when designing redundancy, it is important to know with a high probability, how long the system can perform before every redundant part has failed [12]. This analysis allows system designers to choose the optimal number of redundant parts: reducing cost while assuring with a controlled probability level that the system will not fail for the duration of its mission.

In networking, the wildly used Transmission Control Protocol (TCP) uses retransmission to manage packet loss. Consecutive packet loss perturbs the retransmission scheme [10]. A similar problem is also present in Internet of Things network protocols (e.g. LoRaWAN), characterized by high packet loss rates [9]. Th LoRaWAN protocol coupled with a sequential pseudonyms scheme is highly sensitive to packet loss [11]. In this case, consecutive packet losses results in device desynchronization: a costly process when power is sparse. Let us say that $m$ packets lost implies device desynchronization. It is especially important for network administrators to know with a high probability when the first streak of size $m$ will occur because after desynchronization the device must rejoin the network. This operation makes it so that the streak probability becomes independent of the packets sent before desynchronization, hence the probability of loosing connection is once again the probability of the first streak of size $m$.

From those examples, we remark that the probability law of the number of tries before a streak of size $m$ occurs can help practitioners to adjust adequately various parameters. Hence we propose in this work to study this probability law. With our work, network administrator can, in function of network parameters such as packet loss rate and packet forwarding rate, assess for how long a network holds with arbitrary probability. For instance a network administrator can say "We are sure with a probability of 99% that the network stays up for $x$ years". Our theoretical study aims at giving quantitative and exact answers to questions such as :

- What are the odds of consecutive failure of redundant part?
- What are the odds of consecutive packet loss after some time?
- What are the odds of streaks?

Answering those questions is key to building reliable systems in both software and hardware.

There are many theoretical problems considering streaks, for instance the number of streaks among $n$ tries, or the waiting time in-between streaks. In this paper we choose to study the number of tries before a streak of size $m$ because of its many industrial applications. This problem is studied via two methods. First, numerical simulations have been used [4] but suffer from significant computing time and eventual cost of the simulation if real-world experiments have to be made. Second, by using the inclusion exclusion principle and the Bonferroni inequality to find an approximation [5].

Contrary to the last approaches, our work gives a theoretical analysis of the number of tries it takes to get a streak of size $m$. We are the first to compute the mass function and the cumulative distribution function of the number of tries before a streak.

In addition, we have developed a comprehensive python package which, based on our theoretical work, allows practitioners and systems experts to assess risks of failure and make rational choices of parameters (`https://jaalmoes.com/doc/spistats/`). Our work provides a way to assess the risk of having a streak of size $m$ in reasonable time. The formula's computation relies on two key algorithms: root finding [8] and matrix inversion. As the degree of the polynomial and the rank of the matrix is $m$, the complexity of the numerical method depends on $m$.

## 2. Problem definition

To make this model more general and easier to apply to real world problems, we use the example of repeated coin flips. We consider a coin with a probability $p \in [0, 1]$ to land heads up. We set $q = 1 - p$ for readability. Let $X_0, \cdots, X_{n-1}$ be $n$ random variables independent and identically distributed that follow a Bernoulli law of probability $p \in [0, 1]$. Each $X_i$ models one coin flip and takes the value 1 for head and 0 for tail. Let $m \in \{1, \cdots, n\}$. We are interested in computing the probability of $m$ consecutive heads. We model those events using the following vector of random variables:

$$\forall i \in \{0, \cdots, n - m + 1\} \quad Y_i = \prod_{j=i}^{i+m-1} X_i$$

With this model, the $i$th collection of $m$ consecutive coin flips results in only heads if and only if $Y_i = 1$. On the contrary, if $Y_i = 0$ at least one the $m$ coin flips ends up on tail. This setup is represented in Figure 1.

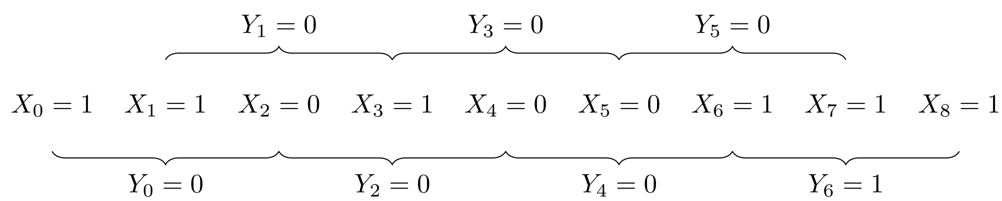


FIGURE 1: Repeating coin flips until $m = 3$ heads in a row.

We remark that even if $(X_i)$ is mutually independent, $(Y_i)$ is not. It means that computing the probability of events such as $\{Y_0 = 0\} \cap \cdots \cap \{Y_n = 0\}$ is not immediate (e.g. a product of $P(Y_i = 0)$). As we see in further developments, this kind of events is key for the problems at hand.

## 3. Mass function

The probability to obtain $m$ heads in a row after flipping the coin $n$ times, and not before, is the probability of the following event: "The first collection of $m$ flips to give only heads is the $(n - m + 1)$th". Let $f(n)$ be the probability of this event. We start by computing $f(n)$ for $n \leq 2m$.

$$f(n) = \begin{cases} 0 & \text{if} \quad n < m \\ p^m & \text{if} \quad n = m \\ qp^m & \text{if} \quad m < n \leq 2m \end{cases}$$

We now study the case where $n > 2m$.

$$P\left(Y_0 = 0 \cap \cdots \cap Y_{n-m-1} = 0 \cap Y_{n-m} = 1\right) \tag{3.1}$$

$$= P\left(\bigcap_{i=0}^{n-m-1} Y_i = 0 \cap \bigcap_{i=0}^{m-1} X_{n-m+i} = 1\right) \tag{3.2}$$

$$= P\left(\bigcap_{i=0}^{n-2m-1} Y_i = 0 \cap X_{n-m-1} = 0 \cap \bigcap_{i=0}^{m-1} X_{n-m+i} = 1\right) \tag{3.3}$$

$$= P\left(\bigcap_{i=0}^{n-2m-1} Y_i = 0\right) P\left(X_{n-m-1} = 0 \cap \bigcap_{i=0}^{m-1} X_{n-m+i} = 1\right) \tag{3.4}$$

$$= P\left(\bigcap_{i=0}^{n-2m-1} Y_i = 0\right) (1-p)p^m \tag{3.5}$$

We give here more precision regarding the step to go from Equation 3.2 to Equation 3.3. We remark that

$$\{Y_{n-m-1}=0\}\cap\bigcap_{i=0}^{m-1}\{X_{n-m+i}=1\}=\{X_{n-m-1}=0\}\cap\bigcap_{i=0}^{m-1}\{X_{n-m+i}=1\}$$

We use the event $\{X_{n-m-1}=0\}$ to simplify the intersection $\bigcap_{i=n-2m}^{n-m-1}\{Y_i=0\}$ because:

$$\forall i\in\{n-2m,\cdots,n-m-1\}\quad\{X_{n-m-1}=0\}\subset\{Y_i=0\}$$

Hence, we have that

$$\bigcap_{i=n-2m}^{n-m-1}\{Y_i=0\}\cap\{X_{n-m-1}=0\}=\{X_{n-m-1}=0\}$$

We continue our development from Equation 3.5. We study the sequence $u_n = P\left(\bigcap_{i=0}^{n}Y_i=0\right)$.

**Case 1** $n=0$**.** $u_0=P(Y_0=0)=1-P(X_0=1\cap\cdots\cap X_{n+m-1}=1)$. Since $(X_n)_{n\in\mathbb{N}}$ is mutually independent, $u_0=1-p^m$.

**Case 2** $n\leq m$**.**

$$u_n=\left(1-P\left(Y_n=1\quad|\quad\bigcap_{i=0}^{n-1}Y_i=0\right)\right)u_{n-1}=u_{n-1}-(1-p)p^m$$

In this case we can solve the sequence :

$$u_n=1-p^m-n(1-p)p^m$$

**Case 3** $n>m$ According it Bayes' formula:

$$u_n=P\left(\bigcap_{i=0}^{n}Y_i=0\right)=P\left(Y_n=0\quad|\quad\bigcap_{i=0}^{n-1}Y_i=0\right)u_{n-1}$$

Since $Y_n$ takes values in $\{0,1\}$,

$$u_n=\left(1-P\left(Y_n=1\quad|\quad\bigcap_{i=0}^{n-1}Y_i=0\right)\right)u_{n-1}\tag{3.6}$$

To compute $P\left(Y_n = 1 \quad | \quad \bigcap_{i=0}^{n-1} Y_i = 0\right)$ we use Bayes' formula once more and obtain:

$$\frac{P\left(Y_n = 1 \cap \bigcap_{i=0}^{n-1} Y_i = 0\right)}{u_{n-1}}$$

By using the same method as in Equations 3.1 to 3.5 and by substitution in Equation 3.6

$$u_n = u_{n-1} - (1-p)p^m u_{n-m-1}$$

We solve the sequence in case 3 by using the initial conditions described in cases 1 and 2. We start by finding the roots of the characteristic polynomial $Q(\lambda) = \lambda^{m+1} - \lambda^m + (1-p)p^m$ for $\lambda \in \mathbb{C}$. Let those roots be $\lambda_0, \cdots, \lambda_m$. Then, $u_n = \sum_{i=0}^{m} c_i \lambda_i^n$ where $(c_i)_{i\in\{0,\cdots,m\}} \in \mathbb{C}^{m+1}$ are found by using the initial conditions of cases 1 and 2. Namely :

$$\begin{pmatrix} c_0 \\ \vdots \\ c_m \end{pmatrix} = \Lambda^{-1} \begin{pmatrix} u_0 \\ \vdots \\ u_m \end{pmatrix} \quad \text{with } \Lambda = \begin{pmatrix} \lambda_0^0 & \cdots & \lambda_m^0 \\ \vdots & \vdots & \vdots \\ \lambda_0^m & \cdots & \lambda_m^m \end{pmatrix}$$

Finding $\Lambda$ and inverting it can be done on a case by case basis numerically. We use Numpy's implementations for numerical matrix inversion and root finding [7]. Hence, for all $n \in \mathbb{N}^*$, we have

$$f(n) = \begin{cases} 0 & \text{if} \quad n < m \\ p^m & \text{if} \quad n = m \\ (1-p)p^m & \text{if} \quad m < n \leq 2m \\ u_{n-2m-1}(1-p)p^m & \text{if} \quad n \geq 2m+1 \end{cases}$$

We display $f$ for $m = 5$, $p = 0.9$ and $m = 20$, $p = 0.2$ on Figure 2.

To demonstrate the utility of our formula when coupled with numerical methods, we interpret this figure using the example of the Internet of Things. Sequential pseudonyms for LoRaWAN are pregenerated and allow for some loss packets before desynchronization [11]. Here the chosen values of $m$ and $p$ correspond to a network where each device has a list of 5 and 20 pseudonyms pregenerated with a packet loss rate of 0.9 and 0.2. It represents two extreme cases: in the first one we expect frequent desynchronization contrary to the second case. Comparing the two graphs, we indeed

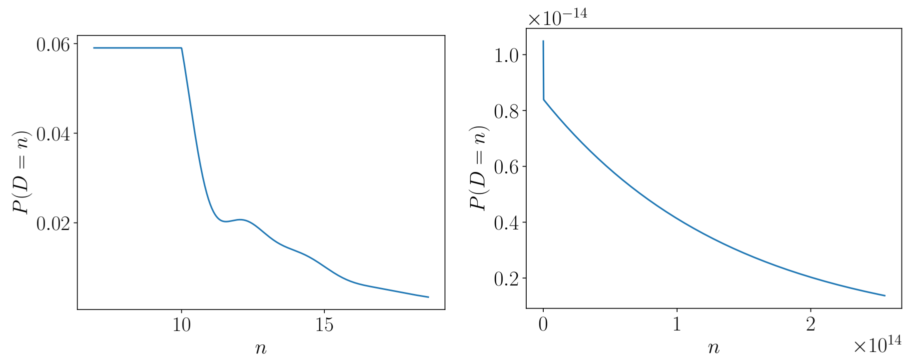


FIGURE 2: Mass function for $p = 0.9$, $m = 5$ on the left and $p = 0.2$, $m = 20$ on the right.

observe that the first instance has a high chance to desynchronize around 10 packets sent.

## 4. Cumulative distribution function

While the mass function gives the exact probability of the streak occurring exactly after $n$ tries, the Cumulative Distribution Function (CDF) provides a more visual interpretation. For instance, we find an interval such that the streak occurs in this interval with a probability of 99%.

We build the random variable $D$ such that $\forall n \in \mathbb{N}\ P(D_n) = f(n)$. Let $F$ be the CDF of $D$: for all $N \in \mathbb{N}^*$, $F(N) = P(D \leq N)$. Hence,

$$F(N) = \sum_{n=1}^{N} f(n) = \begin{cases} 0 & \text{if} \quad N < m \\ p^m & \text{if} \quad N = m \\ p^m + (N-m)(1-p)p^m & \text{if} \quad m < N \leq 2m \\ p^m + m(1-p)p^m + (1-p)p^m \sum_{n=1}^{N-2m-1} u_n & \text{if} \quad N < m \end{cases} \tag{4.1}$$

We compute the sum of $u_n$. Let $n \in \mathbb{N}$,

$$\sum_{i=0}^{n} u_i = \sum_{i=0}^{n} \sum_{j=0}^{m} c_j \lambda_j^i = \sum_{j=0}^{m} c_j \sum_{i=0}^{n} \lambda_j^i = \sum_{j=0}^{m} c_j \frac{1-\lambda_j^{n+1}}{1-\lambda_j} \tag{4.2}$$

We compute $\sum_{j=0}^{m} c_j \frac{1-\lambda_j^{n+1}}{1-\lambda_j}$ numerically because, where $N$ can be large (we go up to $10^{14}$ in Figure 3), $m$ stays small and manageable in practical applications. Indeed, $m$ can be the number of redundant part, the number of pregenerated pseudonyms, etc.

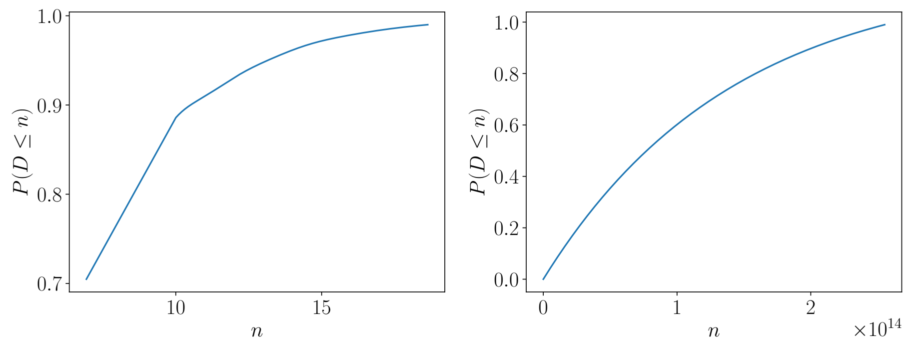


FIGURE 3: CDF for $p = 0.9$, $m = 5$ on the left and $p = 0.2$, $m = 20$ on the right.

Finally, by substitution of the sum in the last line of equation 4.1 we obtain a general formula for the CDF:

$$F(N) = \sum_{n=1}^{N} f(n) = \begin{cases} 0 & \text{if} \quad N < m \\ p^m & \text{if} \quad N = m \\ p^m + (n-m)(1-p)p^m & \text{if} \quad m < N \leq 2m \\ p^m + m(1-p)p^m + (1-p)p^m \sum_{j=0}^{m} c_j \frac{1-\lambda_j^{N-2m}}{1-\lambda_j} & \text{if} \quad N < m \end{cases}$$

We display $F$: the CDF of $D$, in Figure 3. To analyse those CDF, we keep the example of sequential pseudonyms [11]. With the CDF formula, we can state that in the first instance desynchronization happens before 19 packets sent with a probability of 99% which makes this network configuration unfit for production. In the second instance where desynchronization is less likely, it happens after $10^{12}$ packets sent with a probability of 99%. In LoRaWAN, the average packet forwarding rate is around one packet per hour [11]. It means that this network configuration does not desynchronize before $10^8$ years making it extremely stable.

## 5. Average number of coin flips to get $m$ heads in a row

Finally, we use this development to compute the expectancy to allow us to answer question such as: "On average, how many tries are required to obtain a streak of size $m$?"

We compute the expected value of $D$, noted $E(D)$.

$$E(D) = \sum_{i=m}^{+\infty} iP(D = i) = mp^m + (1-p)p^m \sum_{i=m+1}^{2m} i + (1-p)p^m \sum_{i=2m+1}^{+\infty} iu_{i-2m-1} \quad (5.1)$$

We study the partial sums of the last equation. Let $A \in \mathbb{N}$ such that $A > 2m+1$.

$$\sum_{i=2m+1}^{A} iu_{i-2m-1} = \sum_{i=0}^{A-2m-1} (i+2m+1)u_i = \sum_{i=0}^{A-2m-1} iu_i + (2m+1)\sum_{i=0}^{A-2m-1} u_i$$

We study separately the two previous sums.

On one hand: according to Equation 4.2:

$$(2m+1)\sum_{i=0}^{A-2m-1} u_i = (2m+1)\sum_{j=0}^{m} c_j \frac{1-\lambda_j^{A-2m}}{1-\lambda_j}$$

Hence:

$$\lim_{A\to+\infty}(2m+1)\sum_{i=0}^{A-2m-1} u_i = (2m+1)\sum_{j=0}^{m} \frac{c_j}{1-\lambda_j} \tag{5.2}$$

On the other hand:

$$\begin{aligned}
&\sum_{i=0}^{A-2m-1} iu_i \\
=&\sum_{i=0}^{A-2m-1} i\sum_{j=0}^{m} c_j\lambda_j^i \\
=&\sum_{j=0}^{m} c_j \sum_{i=0}^{A-2m-1} i\lambda_j^i \\
=&\sum_{j=0}^{m} c_j \frac{1}{1-\lambda_j}\left(\frac{\lambda_j-\lambda_j^{A-2m}}{1-\lambda_j} - (A-2m-1)\lambda_j^{A-2m}\right)
\end{aligned}$$

Hence:

$$\lim_{A\to+\infty}\sum_{i=0}^{A-2m-1} iu_i = \sum_{j=0}^{m}\frac{c_j\lambda_j}{(1-\lambda_j)^2} \tag{5.3}$$

By substitution of the results of Equations 5.2 and 5.3 in Equation 5.4, we obtain:

$$E(D) = mp^m + (1-p)p^m\frac{3m^2+m}{2} + (1-p)p^m\left(\sum_{j=0}^{m}\frac{c_j\lambda_j}{(1-\lambda_j)^2} + (2m+1)\sum_{j=0}^{m}\frac{c_j}{1-\lambda_j}\right) \tag{5.4}$$

Using this equation for the expected value, we get insight on the variations of the average of coin flips needed to achieve $m$ heads in a row. We do so using two plots. The first one in function of $m$: the streak size, the second one in function of $p$: the probability to get head. We present those two plots in Figure 4. We observe that increasing $m$ results in an exponential growth in the average number of coin flips.

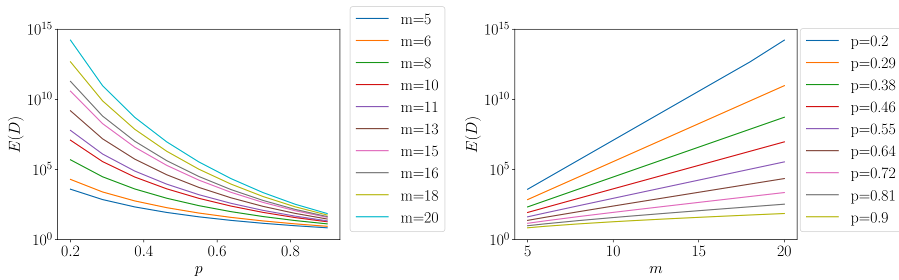


FIGURE 4: Average number of coin flips to get $m$ head in a row where the probability to get head is $p$. On the left, average in function of $p$, on the right, average in function of $m$. Both figures are in log-scale for the y-axis.

## 6. Related work

Despite its numerous applications [12, 5, 10, 1] this problem has seldom been theoretically studied. Using Markov chains gives that the expected number of coin flips until $m$ heads in a row is $\frac{p^{-m}-1}{1-p}$ [6]. This formula is easier to work with than the expression we found for $E(D)$ in Equation 5.4 because it does not need eigenvalues finding. But [6] does not provide an expression for the mass function nor the CDF. Those two functions give more information on the law of $D$ than just the average and can be used to find quantile, moments or even generate sample using inverse transform sampling [3].

## 7. Conclusion

We have successfully computed an expression of the mass function and the CDF of the number of coin flips required before getting $m$ heads in a row. Thanks to our theory and with our implementation, practitioners can now compute confidence intervals, statistical testing and random sampling of the number of tries to get $m$ streaks. Domain experts use those pieces of information to select optimal parameters of the system they are designing (e.g. the number of redundant parts, size of identifier list etc).

## References

[1] Albright, S. C. (1993). A statistical analysis of hitting streaks in baseball. *Journal of the American Statistical Association* **88,** 1175–1183.

[2] Altmann, E. M. and Burns, B. D. (2005). Streak biases in decision making: Data and a memory model. *Cognitive Systems Research* **6,** 5–16.

[3] Collins, R. Intro to Sampling Methods. https://www.cse.psu.edu/ rtc12/CSE586/lectures/cse586samplingPreMCMC.pdf 2005. [Online; accessed 2024-01-23].

[4] de Reffye, J. (2017). Numerical simulation of system dynamical reliability. In *2017 2nd International Conference on System Reliability and Safety (ICSRS)*. pp. 20–30.

[5] Duroeulx, M. (2020). Évaluation de la fiabilité des systèmes modélisés par arbres de défaillances grâce aux techniques de satisfiabilité. *PhD thesis*. Université de Lorraine.

[6] Ginsparg, P. How many coin flips on average does it take to get n consecutive heads? https://www.cs.cornell.edu/ ginsparg/physics/INFO295/mh.pdf 2005. [Online; accessed 2024-01-17].

[7] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C. and Oliphant, T. E. (2020). Array programming with NumPy. *Nature* **585,** 357–362.

[8] Horn, R. A. and Johnson, C. R. (2012). Matrix analysis, 2nd ed.

[9] Liu, Q., Mu, Y., Zhao, J., Feng, J. and Wang, B. (2020). Characterizing packet loss in city-scale lorawan deployment: Analysis and implications. In *2020 IFIP Networking Conference (Networking)*. pp. 704–712.

[10] Nishikawa, Y., Nogami, K. and Takahashi, E. (2016). A method for reducing tcp latency due to consecutive packet losses in mobile networks. In *2016 18th Asia-Pacific Network Operations and Management Symposium (APNOMS)*. pp. 1–4.

[11] Pélissier, S., Aalmoes, J., Mishra, A. K., Cunche, M., Roca, V. and Donsez, D. (2024). Privacy-preserving pseudonyms for LoRaWAN. In *WiSec 2024 - 17th ACM Conference on Security and Privacy in Wireless and Mobile Networks*. Association for Computer Machinery (ACM). ACM, Seoul, South Korea. pp. 1–6.

[12] Tafazoli, M. (2009). A study of on-orbit spacecraft failures. *Acta Astronautica* **64,** 195–205.