Journal of Libyan Academy Bani Walid **2026**
مجلة الأكاديمية الليبية بني وليد
e-ISSN: 3104-3860
Volume 2, Issue 1, 2026, Pages: 368-377
Website: https://journals.labjournal.ly/index.php/Jlabw/index

Information-Theoretic Criteria for Bluetooth Signal Detection

Saleh A. AJOUAT [1*], Ismail S. Almarimi [2]
[1] Libyan Center for Engineering Research and Information Technology L.C.E.R.I.T
Bani Walid/ Libya
[2] Department of Communication Engineering, College of Electronic Technology, Bani Walid, Libya
*Email: saleh.abulgasem@gmail.com

معايير نظرية المعلومات لكشف إشارات البلوتوث

صالح أبو القاسم خليفة [1*]، إسماعيل سالم المريمي [2]
[1] المركز الليبي للبحوث الهندسية وتقنية المعلومات (*L.C.E.R.I.T*)، بني وليد، ليبيا
[2] قسم هندسة الاتصالات، كلية التقنية الإلكترونية، بني وليد، ليبيا

Received: 27-12-2025	**Accepted:** 20-02-2026	**Published**: 04-03-2026

Abstract

Detecting the onset point of Bluetooth signals is a foundational step in wireless communication analysis and signal processing applications. This study introduces an information-theoretic approach named as Minimum Description Length (MDL) principle as a technique for identifying signal transition point based on statistical complexity. MDL works by splitting the signal into chunks and picking the point where the model description is most tight, as a result getting rid of redundancy. The MDL affords a flexible and decision-making process to detect transformation without making threshold. The technique is introduced in a comparative way with Akaike Information Criterion AIC and examined using real Bluetooth signal dataset, showing its ability to recognize the true signal effectively. The suggested method uses the embedded statistical features of the signal, offering a clear system for device weak-up moment recognition in RF signal processing.

Keywords: Minimum Description Length, MDL Onset, Transient Detection, RF Signal.

الملخص

يُعدّ تحديد نقطة بداية إشارات البلوتوث خطوة أساسية في تحليل الاتصالات اللاسلكية وتطبيقات معالجة الإشارات. تقدم هذه الدراسة نهجًا قائمًا على نظرية المعلومات يُعرف بمبدأ الحد الأدنى لطول الوصف (Minimum Description Length - MDL) كطريقة لتحديد نقطة انتقال الإشارة اعتمادًا على التعقيد

الإحصائي. يعمل مبدأ MDL على تقسيم الإشارة إلى أجزاء واختيار النقطة التي يكون عندها وصف النموذج أكثر إحكامًا، مما يؤدي إلى تقليل التكرار. يوفّر هذا المبدأ عملية مرنة قائمة على اتخاذ القرار لاكتشاف التحولات دون الحاجة إلى تحديد عتبة مسبقة. تم تقديم التقنية بشكل مقارن مع معيار أكايكي للمعلومات (Akaike Information Criterion - AIC)، وتم اختبارها باستخدام بيانات حقيقية لإشارات البلوتوث، حيث أظهرت قدرتها على التعرف على الإشارة الحقيقية بكفاءة. تعتمد الطريقة المقترحة على الخصائص الإحصائية الكامنة في الإشارة، مما يوفر نظامًا واضحًا لتحديد لحظة استيقاظ الجهاز في معالجة إشارات الترددات الراديوية (RF)

الكلمات المفتاحية: الحد الأدنى لطول الوصف، بداية MDL، كشف الانتقال، إشارات الترددات الراديوية.

Introduction

The detection of Bluetooth signal onset points plays a critical role in applications such as wireless communication monitoring, spectrum management, and IoT device synchronization. Precise and automated detection ways qualify better signal processing, boosted system performance, and offer more reliable data transmission. Several techniques, especially envelope with threshold-based energy detection [1] and statistical methods such as well-known Akaike Information Criterion method [2], have been widely used.

This article introduces the Minimum Description Length (MDL) method as a technique for estimating the turned-on point of Bluetooth devices. Based on information theory; MDL gives a statistical concept to balance a model's simplicity against its accuracy [3], and it automatically picks the best model for the signal. By modeling Bluetooth signals with a parametric time-series model, the MDL criterion spots change in signal behavior without needing to preset threshold, making it dynamic to signal variation conditions.

The studied method exhibits some benefits, boosts computational efficiency, and increases reliance on automatic parameter adaptation. This paper summaries the theoretical foundation of the MDL method, depicts its application to Bluetooth signal detection, and displays simulation findings that highlight its precision and strength. The outcomes propose that MDL-based detector can be used for real-time RF signal detection and analysis.

Related Work

1. Bluetooth Signal Detection

RF signal detection is well-studied problem in field of wireless communication and security. Common suggested methods for detecting Bluetooth signals include a threshold-based detection and matched filter. In threshold-based detector; the signal's amplitude is compared to a preset level, making it simple but suffers from degraded performance in noisy channels [4]. Matched filtering, the second approach, tries to deduct the received signal that has been attenuated by noise with a known template, it offers better execution but demands prior knowledge of the transmitted signal [5]. Despite their widespread use, these methods frequently encounter problems in precision and adaptability for security-critical tasks, such as device intrusion detection and fingerprinting.

Many researches in Bluetooth signal detection have concentrated on machine learning-based approaches. For example, neural networks and support vector machines have been used to classify Bluetooth signals and detect outliers [6]. However, these techniques frequently need large amount of training data and may lack interpretability, restricting their capabilities in real-world situations. Furthermore, their performance can degrade in dynamic environments with varying signal characteristics or under adversarial conditions [7], [8].

2. Minimum Description Length (MDL) Principle

The Minimum Description Length MDL is an information-theory for model selection that optimizes the trade-off between model complexity and goodness of fit [9], making it particularly effective for tasks like signal segmentation and change-point detection [10]. MDL has been successfully tested in various domains from bioinformatics for DNA sequence analysis [11], to machine learning for model selection [12]. Despite its usefulness, the application of MDL to Bluetooth signal detection is still uninvestigated.

3. Security Applications of Signal Detection

In the field of cybersecurity, detecting Bluetooth signals is essential for identifying specific devices, spotting intruders, and finding security deficiencies. Device fingerprinting is the process of finding the unique digital label that every Bluetooth device has, even if two devices are the exact same model, tiny dissimilarities in their hardware and radio signals make them distinct. By identifying these patterns, a system can tell devices apart for security checks or tracking [13], [1]. Intrusion Detection System (IDS) depends on exact onset detection to catch malicious activities, such as impersonating or man-in-the-middle attacks [14]. Current detection frameworks, however, frequently struggle with the non-stationary noise typical of the 2.4 GHz band. They miss the resolution to isolate the subtle hardware transients that often serve as the first indicators of a spoofing attempt or unauthorized device entry.

Data Gathering & Preparation

For the experimental portion of this research, we sourced the raw Bluetooth signal captures from the public repository provided by [15]. Although the source dataset includes high-frequency captures at 5, 10, and 20 Gsps, our facility's processing constraints led us to select the 250 Msps sampling rate. This specific subset provided a robust sample size—roughly 5,000 individual Bluetooth records. To maintain a diverse hardware baseline, we utilized signals from five distinct manufacturers: IPhone, Samsung, Xiaomi, Sony, and LG. To refine the raw data for processing, we implemented a filtration stage designed to suppress out-of-band noise and stabilize the signal envelope. This optimization ensures the input is sufficiently clean for the MDL algorithm's requirements. To validate the practical utility of this approach, we tested the system against the same real-world RF datasets used in our previous work on AIC-based Bluetooth detection [2]. Utilizing these datasets which contain pre-verified onset points, allowed for a direct performance comparison between our previous AIC results and this new MDL-based framework.

Akaike Information Criteria AIC

The segmentation logic follows the classic AIC algorithm developed by Maeda [16], which serves to partition a signal into two distinct statistical intervals. For a given window of N samples, the algorithm iteratively calculates a value for each index k to detect the transition between the background noise and the signal arisen. The relationship is defined as:

$$AIC(k) = k \times \log\Big(var\big(x(1,k)\big)\Big) + (N-k-1) \times \Big(\log\big(var(x(1+k,N))\big)\Big) \quad \ldots 1$$

In equation (1); the index k acts as a sliding separator. The first term stands for the variance of what supposed to be noise, while the second term captures the variance of the coming signal. By finding the global minimum of the overall curve, algorithm can then pinpoint the exact sample where the signal's statistical properties come up from the noise floor, identifying the optimal onset point.

Minimum Description Length MDL

The Minimum Description Length (MDL) principle approaches signal detection through the lens of data compression. At its core, MDL functions mathematically seeking the most efficient representation of a signal by optimizing the model's goodness-of-fit against its complexity. By minimizing the total description length, then it can be effectively filtered out the overfitting common in noisy environments, ensuring that the detected onset reflects a true hardware transition rather than a stochastic anomaly. The MDL serves as primary criterion for model selection, grounded in the theory that the best model is the one that leads to the greatest compression of the observed data [9].

The MDL formula usually consists of two main components; **Model description cost** which is the cost of describing the model (its parameters, structure, etc.), and the cost of describing the data once the model is specified.

In practice, the MDL principle is applied in tasks like Choosing between different candidate models (e.g., linear regression, decision trees, etc.), identifying points where the statistical properties of a dataset change (often used in time series), and grouping similar data points into clusters by minimizing the total description length of the model and the data.

Applying MDL to Time Series Segmentation

The MDL objective function has been structured around two primary costs: the model description cost and the data description cost. The former represents the complexity of the signal model—specifically its parameters and structural partitions—while the latter accounts for the residual error after the model is applied to the raw samples. In this implementation, we leverage this dual-cost framework to solve three specific challenges in Bluetooth security:

Model Selection: Determining the optimal order of the signal model to prevent overfitting to the noise floor.

Change-Point Detection: Pinpointing the exact temporal index where the statistical properties of the channel transition from ambient noise to an active Bluetooth burst.

Data Partitioning: Segmenting the signal into stationary and non-stationary components by minimizing the total description length, ensuring a precise capture of the hardware's transient.

The MDL criterion is given by equation [17]:

$$MDL(k) = k \times \log\Big(var\big(x(1,k)\big)\Big) + (N - k - 1) \times (\log(var(x(k+1, N)))) + (\frac{k}{2}) \times \log\,(N) \quad \ldots 2$$

In a **particular implementation** of the MDL principle in time series segmentation eq. (2), where k is the partition point in the data, the **first part** of the formula represents the cost of describing the first segment of data, the **second** represents the cost of describing the second segment, the **third part** adds a penalty for the number of parameters (or complexity) involved in the segmentation itself. The goal is to find the partition k that minimizes this total description length, balancing the fit (variance of each segment) and complexity (the split point k).

MDL VS AIC

While both the Minimum Description Length (MDL) and the Akaike Information Criterion (AIC) function as mathematical "optimizing acts" between model fit and complexity. The

AIC—backgrounded in maximum likelihood estimation—is designed primarily to minimize prediction error. In contrast, the MDL framework adopts an information-theoretic approach, treating the detection problem as one of optimal data compression. Crucially, the MDL penalty term—specifically the k/2 log(N) component—provides a defense against overfitting. Unlike AIC, which uses a more static penalty, this MDL term scales with the number of data points (k) relative to the total window (N), it effectively penalizes the "cost" of partitioning the signal, allowing us to distinguish the true hardware-induced onset from the stochastic interference prevalent in the 2.4 GHz band.

Results & Discussion

A visual inspection of the point distributions reveals a consistent lead for the MDL-based approach. For the majority of the samples, the MDL detections (plotted in red) sit noticeably lower on the y-axis than the AIC detections (blue). This suggests that the MDL framework is consistently triggering on the leading edge of the burst earlier than its AIC counterpart. This trend is further supported by the MDL boxplot, which shows a tighter, more compressed range of values.

What stands out, however, are the two outlier clusters where this relationship flips. In these specific signal groups, the red MDL points actually shift above the blue AIC marks. We attribute this localized behavior to unique transient profiles where the MDL's description cost is higher, forcing a later segmentation point to maintain statistical fidelity. This dual-pattern—a general downward shift punctuated by two distinct upward clusters—proves that while MDL is typically more aggressive in finding the onset, it remains sensitive enough to adapt to non-standard signal structures that AIC might overlook.

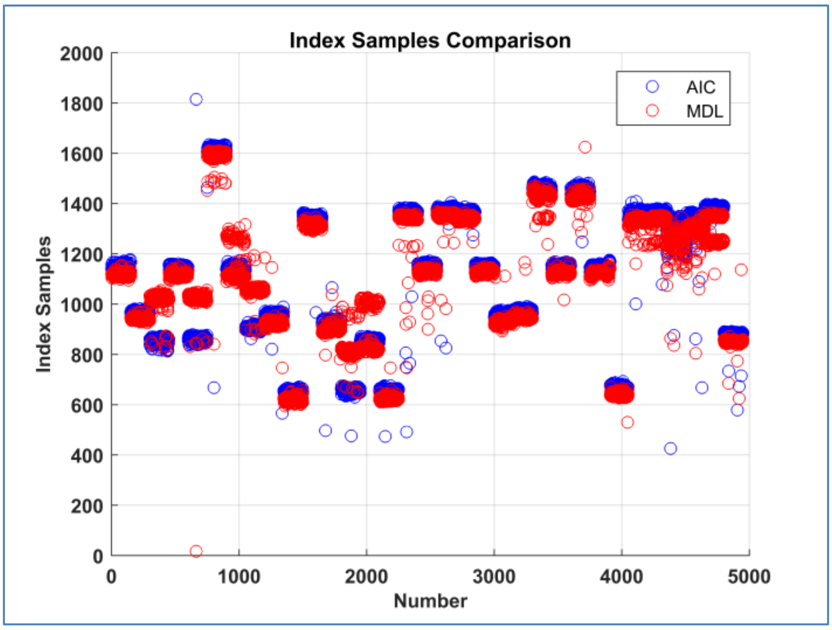

Figure 1 AIC & MDL Index Scatter.

The scatter plot in Figure (1) shows that both the AIC and MDL methods produce nearly identical detection results across the 5,000 signal records for most of the data, the red circles (MDL) and blue circles (AIC) are completely overlapped, forming solid horizontal bands. This indicates that both criteria are selecting the same onset points for the vast majority of the Bluetooth bursts.

However, there are a few noticeable differences in specific areas. In almost all samples MDL algorithm detects onset point earlier than AIC except for some samples (two models), particularly in the 0 to 1000 range and near the 500 mark in x-axis, the blue AIC points appear lower compared to the red MDL points. Most notably, a single red MDL outlier appears near the bottom of the graph (close to index 0) around sample number 700. This rare event suggests a specific case where the MDL's penalty for model complexity triggered a much earlier detection than the AIC.

MDL (red), on the other hand, exhibits a much tighter clustering effect. The dispersion is noticeably lower, which points to a more disciplined response to the signal's statistical shifts. Interestingly, while MDL is generally more stable, we observed a lone extreme outlier at the very foot of the signal. This rare early trigger likely occurs when the model compression cost drops abruptly at the first sign of a noise-to-signal transition, considering the MDL to fail identifying the correct detection point.

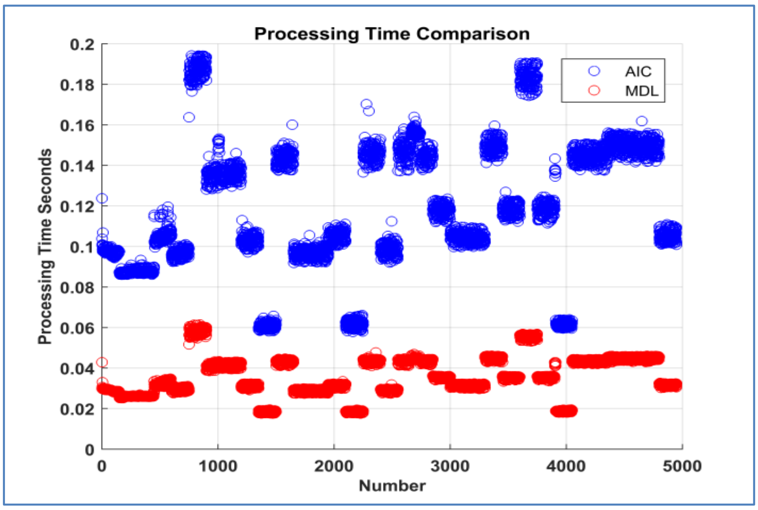

Figure 2AIC & MDL Processing Time Scatter.

The processing time comparison in Figure 2 reveals a dramatic performance gap between the two methods. While the AIC (blue) processing times are scattered across a higher range—mostly between 0.1 and 0.2 seconds—the MDL (red) times remain consistently lower, primarily clustering between 0.02 and 0.06 seconds. This clear separation shows that MDL is significantly faster at identifying signal onsets than AIC.

This speed advantage is likely due to how the MDL penalty term handles complexity, allowing the algorithm to stabilize and reach a decision much quicker than the AIC framework. By cutting the processing time by nearly 60-70% on average, MDL proves to be much more efficient for real-time Bluetooth detection tasks. The fact that the red MDL cluster stays so tight and low across all 5,000 samples indicates that this efficiency is highly stable and not just a result of a few lucky samples.

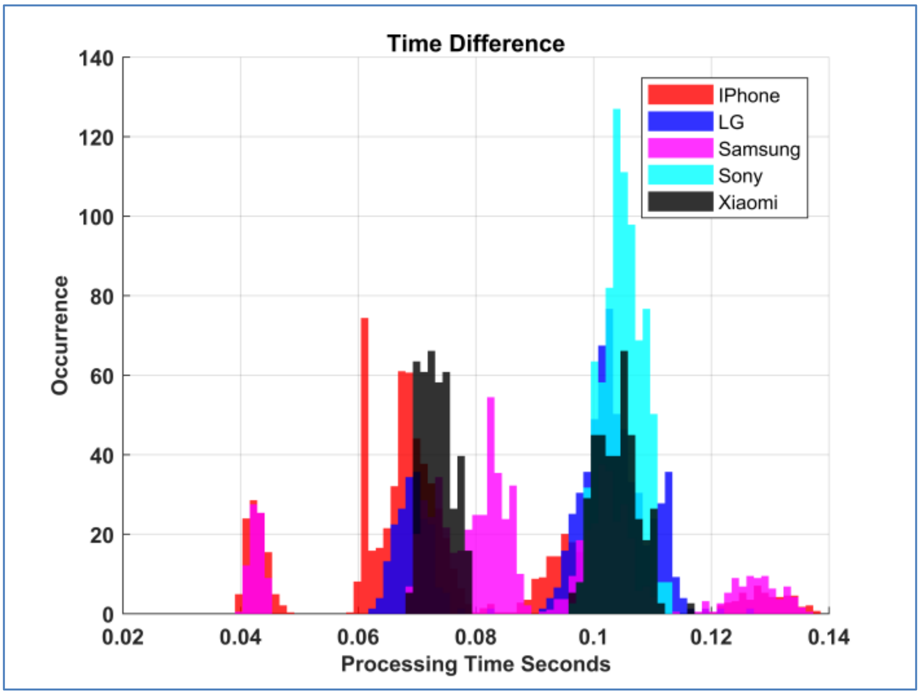

Figure 3 Time Distribution Across Smartphone Brands.

The distribution of processing times across different brands, as shown in Figure 3, discloses unique timing signatures for each manufacturer. While the total execution time for all devices stays within a window of 0.039 to 0.14 seconds, the internal clustering behavior varies significantly by brand:

Sony: Exhibits the most stable performance, with a massive density peak of 127 occurrences in a single cluster. This suggests a highly predictable computational load for Sony hardware.

Xiaomi & IPhone: Both show moderate consistency. IPhone results are more spread out with a peak of 75 occurrences, while Xiaomi splits into two distinct groups peaking at 66, indicating some internal variance in how their Bluetooth transients are processed.

LG & Samsung: These manufacturers show the most fragmentation. LG is split across three clusters, and Samsung shows the widest dispersion of all, with numerous small groups peaking at only 55 occurrences.

These results illustrate that while the underlying MDL algorithm is consistent, the device-specific hardware characteristics introduce distinct variations in computational behavior.

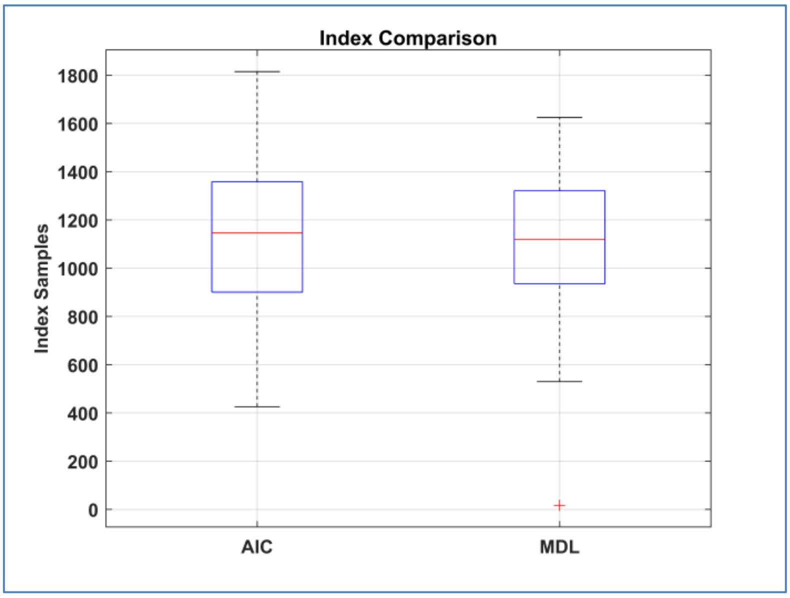

Fig. 4 AIC & MDL Index Comparison Boxplot**.**

The statistical distribution of detection points, captured in Figure 4, confirms that while AIC and MDL operate in a similar window, their stability profiles differ. The AIC results show more stretch in the data, with values reaching an upper limit of 1815 and a wider interquartile range (901 to 1358). This suggests that AIC is more prone to shifting deep into the signal body when faced with complex transients.

In contrast, the MDL framework provides a tighter, more concentrated response. Its distribution is more compact, with a median of 1120 and an upper range that stops at 1625. This reduced spread indicates that MDL is more resistant to the fluctuation that affects AIC. Interestingly, the boxplot highlights a single extreme MDL outlier at index 17. This isolated event represents as a fauilure where the model compression shifted suddenly at the very start of the sample. Overall, these metrics show that MDL offers a more consistent and reliable detection.

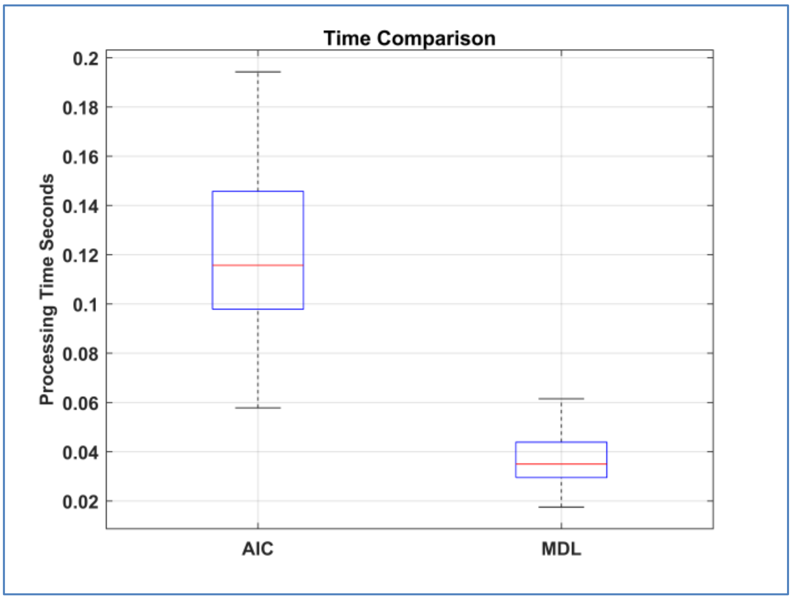

Figure 5 AIC & MDL Processing Time

The processing-time boxplot in Figure 5 visually confirms a substantial performance gap, with the MDL method operating significantly faster. While the AIC execution times are broadly distributed—ranging from a lower adjacent of 0.057 sec to an upper peak of 0.194 sec—the MDL distribution is remarkably compact.

In fact, the entire MDL range (peaking at 0.061 sec) barely reaches the starting point of the AIC's distribution. With a median time of 0.035 sec compared to AIC's 0.115 sec, the MDL approach achieves a roughly 70% reduction in computational latency. The compressed interquartile width of the MDL plot further demonstrates its stability; it isn't just faster on average, but consistently faster with less than half the variability of the AIC.

These results provide clear evidence that MDL serves as a high-efficiency alternative for real-time signal processing where timing is critical.

REFERENCES

[1] Ali, Aysha & Kara, Ali & Uzundurukan, Emre. (2022). Assessment of Features and Classifiers for Bluetooth RF Fingerprinting. IEEE Access. 7. 50524-50535. 10.1109/ACCESS.2019.291145.

[2] AJOUAT S. A., TEZEL N. S., "Radio frequency transient segment detection based on Akaike Information Criterion", Politeknik Dergisi, 25(4): 1681-1686, (2022). https://doi.org/10.2339/politeknik.967341

[3] Rissanen, J. (1989). Stochastic Complexity in Statistical Inquiry.

[4] Y. Yuan, Z. Huang, H. Wu, and X. Wang, "Specific emitter identification based on Hilbert-Huang transform-based timefrequency-energy distribution features," IET Communications, vol. 8, no. 13, pp. 2404–2412, 2014.

[5] F. Salahdine, H. El Ghazi, N. Kaabouch, and W. F. Fihri, "Matched filter detection with dynamic threshold for cognitive radio networks," in Proc. IEEE Int. Conf. Wireless Networks and Mobile Communications (WINCOM), 2015.

[6] Uzundurukan, Emre & Ali, Aysha & Dalveren, Yaser & Kara, Ali. "Performance Analysis of Modular RF Front End for RF Fingerprinting of Bluetooth Devices", (2020). Wireless Personal Communications. 112. 10.1007/s11277-020-07162-z.

[7] Adesina, D., Hsieh, C. C., Sagduyu, Y. E., & Qian, L. (2022). Adversarial Machine Learning in Wireless Communications: Challenges and Opportunities. Wireless Communications and Mobile Computing, 2022, 1-14.

[8] Z. Bao, Y. Lin, S. Zhang, Z. Li and S. Mao, "Threat of Adversarial Attacks on DL-Based IoT Device Identification," in IEEE Internet of Things Journal, vol. 9, no. 11, pp. 9012-9024, 1 June1, 2022, doi: 10.1109/JIOT.2021.3120197.

[9] Grünwald, Peter. "The Minimum Description Length Principle", (2007). 10.7551/mitpress/4643.001.0001.

[10] Davis, Richard & Lee, Thomas & Rodriguez-Yam, Gabriel. (2006). Structural Break Estimation for Nonstationary Time Series Models. Journal of the American Statistical Association. 101. 223-239. 10.1198/016214505000000745.

[11] S. Evans, S. Markham, A. Torres, A. Kourtidis and D. Conklin, "An Improved Minimum Description Length Learning Algorithm for Nucleotide Sequence Analysis," 2006 Fortieth Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 2006, pp. 1843-1850, doi: 10.1109/ACSSC.2006.355081.

[12] Hinton, G. E., & van Camp, D. Keeping the neural networks simple by minimizing the description length of the weights, (1993). Proceedings of the 6th Annual Conference on Computational Learning Theory (COLT '93).

[13] O. Ureten and N. Serinken, “Wireless security through RF fingerprinting,” in Canadian Journal of Electrical and Computer Engineering, Winter 2007. https://doi.org/10.1109/CJECE.2007.364330

[14] Hall, Jeyanthi et al. “Detecting rogue devices in bluetooth networks using radio frequency fingerprinting.” CCN (2006).

[15] Uzundurukan, E. & Dalveren, Y. & Kara, A. “A Database for the Radio Frequency Fingerprinting of Bluetooth Devices”, (2020). https://doi.org/10.3390/data5020055

[16] N. Maeda, “A method for reading and checking phase time in auto-processing system of seismic wave data,” Zisin (Journal of the Seismological Society of Japan. 2nd ser.), vol. 38, (1985).

[17] Wax, Mati & Kailath, Thomas “Detection of signals by information theoretic criteria. Acoustics”, (1985)., Speech and Signal Processing, IEEE Transactions on. 33. 387 - 392. 10.1109/TASSP.1985.1164557.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.