Challenges of the Age of Information Paradigm for Metrology in Cyberphysical Ecosystems

Alberto Zancanaro*, Giulia Cisotto†, and Leonardo Badia*
* Dept. of Information Engineering (DEI), University of Padova, Italy
† Dept. of Informatics, Systems and Communication (DISCo), University of Milano-Bicocca, Italy
email: alberto.zancanaro.1@phd.unipd.it, giulia.cisotto@unimib.it, leonardo.badia@unipd.it

***Abstract*—We are facing a transition towards interconnection of computing systems, people, and things, where boundaries are disappearing and new challenges are emerging. This trend also applies to smart living environments, which are becoming a cyberphysical ecosystem of devices and individuals. Generally, meta-descriptors such as age of information are exploited to obtain efficient content representation and semantic characterization, with the advantage of better data handling. However, the strong relevance of living support in the involved applications imposes to rethink of this approach whenever it is important to factor the human-in-the-loop. In this paper, we discuss how the investigations related to age of information, in particular aimed at statistical descriptions and/or network operation modeling, can be influenced in such scenarios, for what concerns overarching machine learning for data classification and its impact on the sensing frequency, as well as the presence of data correlation that allows for a parsimonious handling of the updates.**

***Index Terms*—Age of Information; Internet of Things; Data acquisition; Networks; Modeling.**

I. Introduction

The last decade has seen an unprecedented development in smart environments due to the technological advancements in the Internet of Things (IoT), sensors, and artificial intelligence. There is a wide gamma of applications for these innovations in smart living environment, from smart houses to assisted living, especially for elderly people [1], [2]. Also, IoT techniques contribute to achieve better sustainable energy consumption [3] and the introduction of these solutions for sensing, data analysis, and active system control, enables the creation of smart cyberphysical ecosystems, where machine and people are interconnected [4]. Unfortunately, using such new technologies also leads to a tremendous increase in the amount of data produced and consequently hinders their management [5].

Novel approaches are needed to address these problems from a holistic perspective. A possible solution comes from the concept of Age of Information (AoI). In the context of network analytics, AoI [6] is often proposed as a new metric to quantify freshness of data coming from real-time monitoring of status updates or control [7], [8]. Thus, instead of just exchanging data, we also track their degree of innovation they bring to the historical description of the cyberphysical ecosystem.

Over the years, different approaches have been proposed to optimize various network features with an eye to AoI as a key performance indicator. For example, [9] optimizes transmission and sampling cost in a wireless network under AoI constraints through Lyapunov optimization theory. In [10], game theory is used to minimize the AoI from two different competing sources. Another area where AoI is becoming increasingly important is energy optimization like in [11]–[13] where the problem of assessing the impact of energy harvesting on AoI is analyzed.

Nonetheless, various challenges remain open in the field of smart living environment. Despite its popularity, AoI is a simple metric that does not codify the intrinsic value of information in a possibly complex scenario. From the perspective of AoI, an anomaly has the same importance of a normal status update, which is clearly undesirable whenever the scenario is supposed to provide some application in a smart living context.

Another important factor that can be considered to optimize the data exchange by keeping AoI into account is source correlation [14], [15], i.e., the fact that an update from a source could also contain information related to other sources. This is especially the case when we have multiple uncoordinated sources that monitor different but related metrics, or are just redundant.

Also, while AoI can be very adequate to increase data representation and explainability, it can also introduce a bias and be subject to error propagation [16]. Indeed, AoI assumes an underlying binary information. In reality, information coming from sensors, especially tracking smart living applications, can be multistructured, and an interpretation is often required. We can think of applying machine learning (ML) to manage multidimensional data and extract meaningful information that can be handled in the updates [17]–[20]. In light of the aforementioned points, this approach can offer both pros and cons for smart living ecosystems. First of all, ML applied to multiple measurements from different sources can leverage the correlation among the updates and therefore reduce the redundancy and eventually limit the signaling and consumption of the remote sensors. On the other hand, the superimposition of ML techniques can make data handling less robust, because of error propagation and the critical impact of misclassification in the learning procedure.

In this paper, we are touching the aforementioned points, by giving some numerical results that highlight and possibly quantify the impact of these challenges, and we suggest possible developments for future applications of the AoI paradigms in smart cyberphysical ecosystems.

The rest of this paper is organized as follows. In Section II, we discuss our main scenario, what are the typical updating epochs in a smart environment, also in relationship to power consumption of the battery and how to exploit possible correlation between the multiple sensors. In Section III, we discuss the application of ML and show how this could lead to different possible results. In Section IV, we present some numerical results. Lastly, Section V presents the conclusions and future work.

II. Scenario and Methodology

Consider a smart living environment monitored by a network of N sensors, i.e., belonging to set $\mathcal{N} = \{1, 2, \dots, N\}$, that samples information and sends it to a central server, where they are processed and analyzed. Time is assumed to be discrete, i.e. $t \in \mathbb{Z}^+$, and in each slot a sensor can decide to sample new information and send an update to the central server.

Modern sensors can have a high sampling rate, up to the order of seconds, but this is not always useful in smart living scenarios, where the application to track has slow dynamics. For example, sensors that monitor the temperature inside a room would waste resources even by sending updates every 10 seconds, since the changes in temperature of a living environment are usually slower. More in general, a dense sampling of information from the sensor can present two major drawbacks. Firstly, from a network perspective, sending all updates from every sensor will cause an overload, together with all the related problems, like collisions and reduced throughput and available bandwidth [21]. Secondly, from the individual sensor standpoint, sending frequent updates will consume a high amount of energy, which goes against principles of efficiency and reducing carbon emissions. In addition to the general requirement for low energy consumption, depending on the specific kind of ecosystems, nodes can be connected to the grid or they can be fully wireless and rely on batteries, which implies different approaches to the power management [12].

It is therefore necessary to find an adequate sampling rate, that allows to optimize the energy consumption and at the same time to maximize the information received [22]. From a theoretical perspective, reducing the frequency of sampling of the sensors can be related to Nyquist theorem [23]. Suppose for example that a sensor that can send an update every minute. If the sensors are tracking a process that changes once per hour (frequency f = 1 hour^{-1}) then you may want to sample as frequently as twice that (i.e., two times an hour) and not every minute. In this scenario, AoI represents an important metric to reach our objective since it allows to quantify freshness of information, and together with the knowledge of the process monitored, it can be used to make the update process efficient.

We can add the further condition that the measurements of a generic sensor i in the smart living environment are at least partially correlated with those of other surrounding sensors, i.e., its *neighbors*, belonging to a neighborhood n_i. The exact nature of this neighborhood can be physical, in particular whenever there is spatial redundancy of the metrics tracked in the living environment (e.g., the temperature in two different positions of the same room), or logical, which happens if the nodes in the neighborhood n_i are those measuring metrics with strong correlations [24]. Whatever the reason for the correlation within a neighborhood, it is clear that, when a node sends data, its update can be also useful to its neighbors. Thus, we can exploit it for either decrease unnecessary updates, and also in the context of applying ML to extrapolate meaningful information from aggregation of multiple measurements.

In the following, we will show that even a basic scenario like this presents challenges in the application of ML techniques, and, more generally, non-negligible practical consequences, which are at the same time enabling a more efficient data handling but also prone to misdetection and error propagation.‘

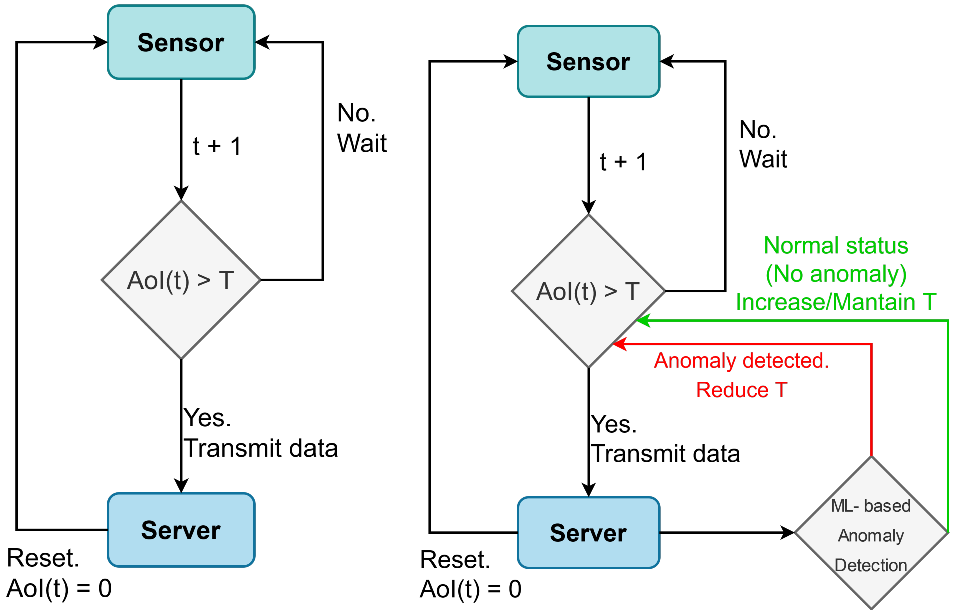
III. Machine Learning for AoI

AoI often assumes a simple view regarding the data, since it just describes how recent the last update was regardless of its content [7]. In real world scenarios, however, information coming from sensors can be multi-structured and data can have different importance levels for the end user. The application of ML offers a powerful instrument to integrate this aspect and extend the concept of AoI in the more general concept of “value of information,” where also the semantic aspects of the data themselves are important to evaluate when to transmit them or not.

Consider the scenario presented in Section II, where N sensors monitor a living environment and send their updates to a server. Instead of simply adjusting update rates based on their age, an ML algorithm can further analyze the data and classify each update as *normal status* or *anomaly*. This adds a further processing step to the system and might lead to different possible results.

For example, the update can have no important information (*normal status*), so the AoI for that process can be updated less frequently (i.e., to save power and leave bandwidth free for other sensors). Alternatively, an alarm needs to be raised (*anomaly* detected), and AoI must be kept very low in the short-term, so the update rate is increased, neglecting possible higher energy consumptions (i.e., avoid power saving modes) at least for the time being. Finally, the update can be inconclusive. This happens when the content of the update is not clear, so old data are kept being used, with an AoI value that is increased by 1.

The main drawback of this method resides in classification errors. For example, there can be an apparently valuable update (*anomaly*), which is actually a false positive. This error has a weak impact on the system since its main consequence is to force a fresh update from a process that carries little relevance. Still, energy is wasted, which may lead to inefficiency at the ecosystem level. On the other hand, if no valuable update (*normal status*) is reported, but this is actually a false negative, the problem is even more acute [16]. This error is harmful since an anomaly that can damage the system is undetected, but the sensor that monitors the process has no reason to maintain its AoI low, and thus, no frequent updates are performed to keep the process monitored.



(a) Baseline updating scheme

(b) ML-empowered updating scheme

Fig. 1: The role of ML in the AoI evaluations. A baseline scheme without ML (a) is compared with an ecosystem with ML in the loop (b), with a dynamic adjustment of the AoI policies (updating threshold).

Another option to increase the robustness of analysis is to use ML to aggregate different measures taken over time instead of simply classifying each update. This process could potentially localize anomalies that require multiple measures to be found and provide in the end more accurate classification based also on historical records [20].

A possible insertion of ML techniques in AoI procedures, which will be explored further, concerns the adjustment of AoI operating policies according to the actual content of the updates [19]. For example, one can think of comparing a baseline scheme, where an update is sent whenever AoI is greater than a predefined threshold T, with a scheme extended through ML-based classification of the updates into anomalies or normal data, so that the value of T is updated accordingly, e.g., to give higher priority to signaling anomalies. A logical scheme of this comparison is shown in Fig. 1.

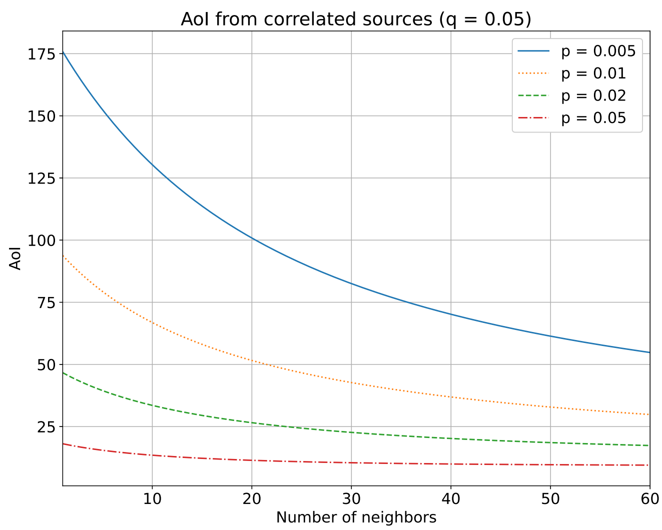


Fig. 2: AoI vs. number of neighbors in a loosely correlated scenario ($q = 0.05$).

IV. Implementation, Results, and Discussion

We now present some numerical results obtained through our simulations for the previously described scenario. We consider a network of multiple sensors that are used to monitor a common environment. The time is slotted and in each slot a sensor $i \in \mathcal{N}$ can send an update with probability p_i and reset its AoI, or not sending an update, with probability $1-p_i$ and increasing its AoI of 1. Also, we assume the cardinality of neighborhoods to be n nodes, so that if in each time slot, at least one of the n neighbors of the sensor sends an update this will also reset the AoI of the sensor itself with probability q_i. For the sake of simplicity, we consider a symmetric scenario where $p_i = p$ and $q_i = q$, for all $i \in \mathcal{N}$.

A. Exploiting data correlation

We investigate how the presence of correlation in the monitored data of the smart living environment can improve the AoI. Specifically, every update from a sensor, happening with probability p, does not only benefit the specific AoI value it refers to, but can also benefit the AoI of another node in the neighborhood (and thus reset its AoI to 0) with probability q. We investigate how this is impacted by the numerical values of p, q, and the size of the neighborhood n.

The results are reported in Figs. 2 and 3, which show the average AoI of a node versus the number of neighbors n for

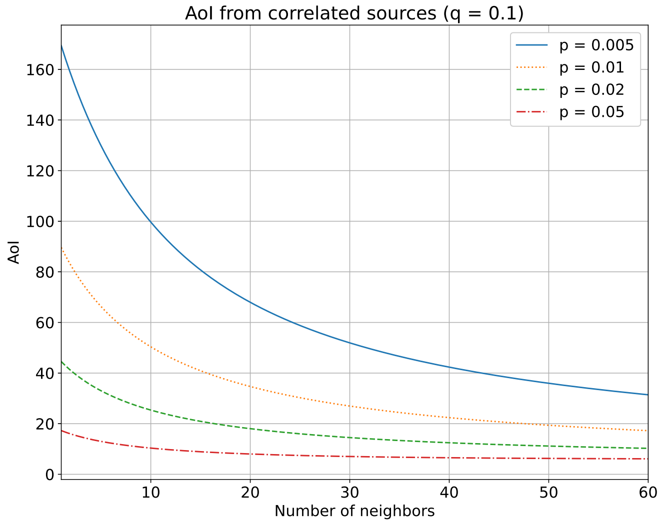


Fig. 3: AoI vs. number of neighbors in a strongly correlated scenario ($q = 0.1$).

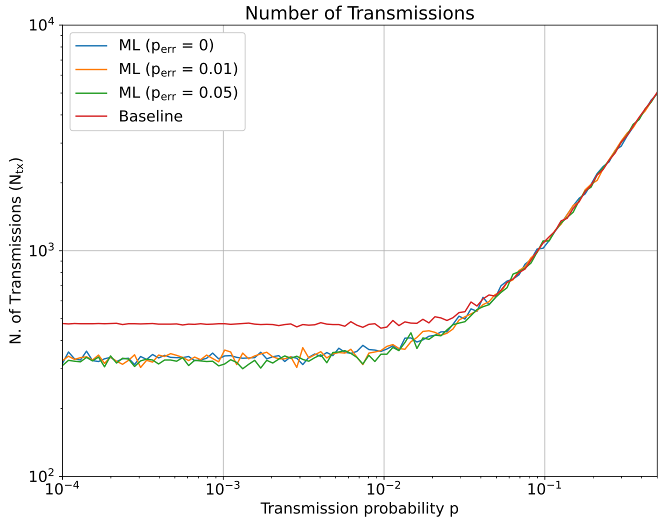


Fig. 4: Total number of transmissions at the end of the simulation.

different values of the transmission probability p [25]. The two figures report different values of the probability q, i.e., that a neighbor transmission is useful; specifically, they show $q = 0.05$ and $q = 0.1$, respectively.

In both figures, it is possible to note how the average AoI decreases if the number of helpful neighbors increases. One interesting fact is that the decrease is more evident for a lower probability of transmission (blue continuous line). This is likely due to the fact that, when each node updates more often, the contributes of its neighbors are marginal. Instead, for lower p, the neighbor updates gain more weight.

This can be leveraged whenever we want to reduce the energy consumption of the sensors without significantly affecting the AoI. In fact, based on this simulation, with enough neighbors we can have a small p (i.e., sparse updates) but also keep the AoI low. The main advantage will be to extend the battery life of the sensors, because fewer updates means fewer transmissions and, thus lower energy consumption. At the same time, fewer transmissions means lower network overload and this will reduce the likelihood of collisions and the consequent loss of data.

B. Testing the AoI of ML-empowered sensing

For assessing the impact of ML, we simulate the behavior of a single sensor tracking the average AoI and the total number of transmissions. We simulate two scenarios, one without ML (henceforth referred to as the *baseline* case) and one with a ML for classifying the received updates. We do not actually consider a specific ML scheme, but we just account for the misclassification events and the possibility of aggregating and leveraging information from neighbor nodes.

The simulation considers a discrete time axis, with 10000 time epochs. The status of a single sensor and its AoI are tracked at each time step, where 4 outcomes are possible: 1) The sensor sends an update with probability p. Therefore, AoI is reset to 0 and the number of transmissions is increased by 1; 2) At least one of the N neighbours sends a useful update with probability q, therefore the AoI of the sensor is reset but the number of transmissions is not increased; 3) The AoI exceeds the predefined value T (set at the beginning of the simulation to some quantity T_0) and the sensor is forced to send an update, so that once again the AoI is reset to 0 and the number of transmissions is increased by 1; 4) None of the previous cases, so no update is performed, in which case the AoI is increased by 1 but the number of transmissions from that sensor is kept the same.

In case of an ML-empowered scheme, we also consider another option. Each update is classified through a ML algorithm into a binary outcome ("normal status" or "anomaly"), with a symmetric probability of misclassification equal to $p_{\rm err}$. According to our previous description (see Fig. 1), we modify the AoI threshold according to how the ML procedure classifies the update. In particular, the initial threshold is set to $T := T_0$; then, whenever an anomaly is detected, the threshold is set to $\max(1, T/2)$ to force the system to sending more frequent updates (ideally, every slot if the anomaly persists). Otherwise, the threshold is increased by 1, so $T := \min(T + 1, T_0)$.

We used $N = 30$, $q = 0.05$, $e = 0.05$ and $T_0 = 20$. Then, we simulate for different values of p and $p_{\rm err}$. At the end of each simulation, we save the total number of transmissions, namely $N_{\rm tx}$, and the average AoI. The results are reported in Fig. 4 and 5, respectively, where the two values are plotted against the probability of transmission p.

As can be observed from both figures, the effect of ML is more evident for lower p. For lower transmission probabilities, the baseline scenario obtains an average AoI and a number of transmissions that are only influenced by T_0, since the only way that the AoI is reset to 0 is when the sensor is forced to update from hitting T_0. In this same situation, the impact of an ML-empowered tracking is to decrease the number of transmissions, since it allows to exploit the redundancy present from the network structure, but also consequently implying a

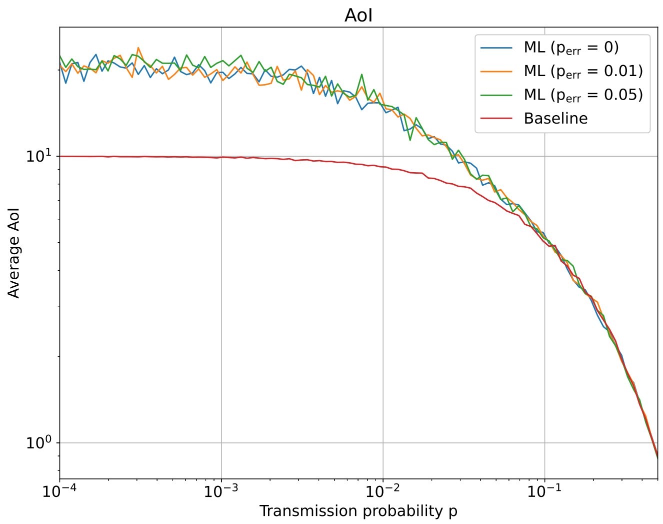


Fig. 5: Average AoI at the end of the simulation.

slight increase in the average AoI. This effect vanishes, as is reasonable, with the increase of the transmission probability. No relevant difference was noted for the p_{err} tested, thereby implying that a limited error rate can be recovered thanks to subsequent correct updates.

We can conclude that for a scenario with sporadic updates, ML-empowered techniques can be useful to reduce the number of transmissions, and therefore energy consumption as well as network congestion. On other hand, the ML caused an increase in the average AoI and makes the system possibly prone to missing critical updates when anomalies are misclassified. Yet, collecting data over multiple sensors and/or time instances may obtain a richer description to avoid this problem. Future tests in more extended setups, and possibly in real world scenarios, will be needed to find the adequate trade-off between reducing the number of transmissions and the need for precisely tracking the metrics of interests.

V. Conclusions and Future Works

In resource-constrained environments, the availability of fresh information is an important challenge that can be addressed through AoI. In this paper, we present possible applications of AoI for monitoring smart living environments and we show the importance of applying ML-empowered classifications of the state of the ecosystem, to increase data significance and avoid errors. Also, we show how considering the neighbors of each sensors can be useful to further optimize the network and the energy consumption. Future extensions of this work will include more results regarding the implementation and impact of ML in this type of systems.

Acknowledgment

This work was supported by REPAC, a project funded by the University of Padova under the initiative SID-Networking 2019. Part of this work was also supported by the Italian Ministry of University and Research under the initiative *"Departments of Excellence"* (Law 232/2016).

References

[1] A.-L. Kor, M. Yanovsky, C. Pattinson, and V. Kharchenko, "SMART-ITEM: IoT-enabled smart living," in *Proc. Future Techn. Conf.*, 2016, pp. 739–749.

[2] G. Cisotto and S. Pupolin, "Evolution of ICT for the improvement of quality of life," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 33, no. 5-6, pp. 6–12, 2018.

[3] S. Ali and D.-H. Kim, "Simulation and energy management in smart environment using ensemble of GA and PSO," *Wireless Personal Communications*, vol. 114, 09 2020.

[4] S. L. Ullo and G. R. Sinha, "Advances in smart environment monitoring systems using iot and sensors," *Sensors*, vol. 20, no. 11, p. 3113, 2020.

[5] M. Ma, P. Wang, and C.-H. Chu, "Data management for internet of things: Challenges, approaches and opportunities," in *Proc. IEEE GreenCom*, 2013, pp. 1144–1151.

[6] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. IEEE Infocom*, 2012, pp. 2731–2735.

[7] R. D. Yates and S. K. Kaul, "Status updates over unreliable multiaccess channels," in *Proc. IEEE ISIT*, 2017, pp. 331–335.

[8] I. Kadota, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Minimizing the age of information in broadcast wireless networks," in *Proc. Allerton Conference, IEEE*, 2016, pp. 844–851.

[9] E. Fountoulakis, N. Pappas, M. Codreanu, and A. Ephremides, "Optimal sampling cost in wireless networks with age of information constraints," in *Proc. IEEE Infocom Wkshps*, 2020, pp. 918–923.

[10] L. Badia, "Age of information from two strategic sources analyzed via game theory," in *Proc. IEEE CAMAD*, 2021, pp. 1–6.

[11] B. T. Bacinoglu, E. T. Ceran, and E. Uysal-Biyikoglu, "Age of information under energy replenishment constraints," in *Proc. IEEE ITA*, 2015, pp. 25–31.

[12] L. Crosara and L. Badia, "A stochastic model for age-of-information efficiency in ARQ systems with energy harvesting," in *Proc. European Wireless*, 2021, pp. 1–6.

[13] B. T. Bacinoglu, Y. Sun, E. Uysal-Bivikoglu, and V. Mutlu, "Achieving the age-energy tradeoff with a finite-battery energy harvesting source," in *Proc. IEEE ISIT*, 2018, pp. 876–880.

[14] B. Zhou and W. Saad, "On the age of information in Internet of Things systems with correlated devices," in *Proc. IEEE Globecom*, 2020, pp. 1–6.

[15] L. Badia, "On the impact of correlated arrivals and errors on ARQ delay terms," *IEEE Trans. Commun.*, vol. 57, no. 2, Feb. 2009.

[16] ——, "On the effect of feedback errors in Markov models for SR ARQ packet delays," in *Proc. IEEE Globecom*, Dec. 2009, pp. 1–6.

[17] A. Elgabli, H. Khan, M. Krouka, and M. Bennis, "Reinforcement learning based scheduling algorithm for optimizing age of information in ultra reliable low latency networks," in *Proc. IEEE ISCC*, 2019, pp. 1–6.

[18] M. Samir, C. Assi, S. Sharafeddine, D. Ebrahimi, and A. Ghrayeb, "Age of information aware trajectory planning of UAVs in intelligent transportation systems: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 12 382–12 395, 2020.

[19] E. Uysal, O. Kaya, S. Baghaee, and H. B. Beytur, *Age of information in practice*, ser. Age of Information. Cambridge Univ. Press, 2022.

[20] M. Braei and S. Wagner, "Anomaly detection in univariate time-series: A survey on the state-of-the-art," 2020. [Online]. Available: https://arxiv.org/abs/2004.00433

[21] S. S. I. Samuel, "A review of connectivity challenges in IoT-smart home," in *Proc. ICBDSC*, 2016, pp. 1–4.

[22] E. Gindullina, L. Badia, and X. Vilajosana, "Energy modeling and adaptive sampling algorithms for energy-harvesting powered nodes with sampling rate limitations," *Trans. Emerg. Telecommun. Technol.*, vol. 31, no. 3, p. e3754, 2020.

[23] C. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, jan 1949.

[24] G. Cisotto, A. V. Guglielmi, L. Badia, and A. Zanella, "Joint compression of EEG and EMG signals for wireless biometrics," in *Proc. IEEE Globecom*, 2018, pp. 1–6.

[25] A. Zancanaro, G. Cisotto, and L. Badia, "Modeling value of information in remote sensing from correlated sources," in *Proc. IEEE MedComNet*, 2022.