

Using Correlated Information to Extend Device Lifetime

Jernej Hribar^{*}, Maice Costa^{*}, Nicholas Kaminski[†], and Luiz A. DaSilva^{*}

^{*}CONNECT- Research Centre at Trinity College Dublin, Ireland

[†]Institute for Defense Analyses, Alexandria Virginia USA

E-mail: jhribar@tcd.ie.

Abstract—The massive device deployments in the Internet of Things (IoT) generate immense amounts of data that can be leveraged to improve overall network performance. This work outlines how data gathered from correlated sensor nodes can be used to improve the timeliness of updates of another sensor node in the network. We consider a system of two correlated information sources, i.e. sensor nodes, which periodically send updates to a gateway, regarding the observed physical phenomenon distributed in space and evolving in time. The optimal use of updates in such a system greatly depends on the correlation between the two sources, and to explore this effect we investigate three different models of the covariance between independently obtained observations of the phenomenon of the interest. We extract values for the parameters in the covariance models from data coming from a real sensor network, to provide the reader with a realistic feel for scaling parameters values and the applicability of our analysis in a real scenario. We demonstrate that using correlated information results in a significant increase in device lifetime and compare our approach to others proposed in the literature.

Keywords—*Internet of Things (IoT), energy efficiency, Age of Information, status updates*

I. INTRODUCTION

The Internet of Things (IoT) connects the physical world with a digital one using ubiquitous computing, various communication technologies, and embedded devices with sensors and actuators [1]. Many of the connected devices will be battery-powered sensing devices, with the primary function to observe phenomena (e.g. weather, vehicular traffic, etc.) to provide a system with information regarding the physical world. A gargantuan number of sensing devices inevitably leads to high correlation in time and space between different observations of the same phenomenon. By leveraging the correlation between sensors, i.e. information sources, it is possible to improve network performance. In this paper, we focus on devices with limited available energy within an IoT network and show that correlation can be used to improve the lifetime of those devices.

In a network with correlated information sources, it is possible to save energy by controlling the rate at which the sources transmit messages to update the information available about the observed phenomenon. These messages, called status updates, contain a measurement of the observed process, and a timestamp to indicate when the observation was made. The

information gathered at a given time or location can be used to predict the phenomenon at a different position and at a future time instant. Therefore, by taking advantage of correlation between different observations of the same phenomenon, it is possible to decrease the rate at which information sources send status updates to a gateway. In our work, we demonstrate the device lifetime gains that can be obtained by controlling the frequency of status updates from two correlated information sources.

The timeliness of a status update depends on the time elapsed since the last update from the information source and can be characterised through a metric called the Age of Information (AoI) [2]. AoI is relevant in a network in which the receiver is interested in fresh information, but information sources are limited in how often they can send updates, e.g. due to energy constraints or excessive contention for the transmission channel. A network of battery-powered devices is an example of such a network. By considering correlation between observations made at different times or locations, it is possible to use up-to-date information collected from a correlated source instead of using outdated information gathered at the point of interest. In our work, we examine when fresher information from a correlated source can replace information from the source of interest.

Our system model consists of one information source of interest and one secondary correlated source. In such a case, it is possible to use the updates from the secondary source to reduce the frequency of updates from the source of interest. The secondary source may exist because it collects information on its own or it may have been placed in the network with the sole intent to assist the source of interest. An example of the first scenario is vehicular traffic monitoring sensors installed at selected sites in a city. While the main purpose of each sensor is to collect information from its own site, the correlation between the information collected by all sensors can be leveraged to reduce their update rates. An example of the second scenario would be when the source of interest is battery-powered, and the desire to extend its lifetime may limit its update rate. In such a case, it may be possible to deploy a secondary source capable of observing a correlated phenomenon, but without the same stringent energy constraints. Whenever a correlated source is present in the network, it is possible to utilise it to improve the timeliness of information regarding the phenomenon of interest.

In this paper, we analyse the timeliness of updates in a system comprising of a pair of correlated sources and determine how the scheduling of updates from the secondary source may be of use to the source of interest. We show that the time difference between consecutive updates from the sources is crucial. We derive the optimal time difference between the two sources' updates, for which the presence of correlated sources will bring the highest benefits to the timeliness of the information collected. Due to system dependence on the correlation between the sources, we model the correlation using three different covariance models, one separable and two non-separable [3], [4], to illustrate how various covariance models impact the timeliness of updates. In the last section, we extract values for the parameters in the covariance models from real IoT data and demonstrate that a system can extend the lifetime of a source by relying on fresher information from a correlated source.

II. RELATED WORK

The idea of using the correlation between neighbouring sensor nodes that transmit information regarding the same phenomenon has its roots in the area of wireless sensor networks (WSNs) [5]. The initial intention was to use only a fraction of all available sensor nodes to reconstruct the observed physical phenomenon. This objective bears similarities to approaches proposed in estimation and detection theory, where the primary aim was to determine the necessary density to detect the observed physical phenomenon correctly [6]. Our work is related to both of these concepts, but differs in one crucial aspect: we are not reconstructing the observed physical phenomenon. Instead, our main contribution is analysing the timeliness of updates, i.e. measuring how outdated the information is, and then use correlated measurements to reduce the rate at which sources transmit their updates.

Several data-driven techniques that utilise the correlation between sensor nodes for energy conservation have been proposed in the past [7]. Of these, the one most related to our work is the data prediction technique, in which information sources use mathematical models to estimate the value of the observed physical phenomenon at a given time and place. In some techniques, the sink estimates the value of the observed phenomenon by aggregating information from multiple sources. Sources transmit updates only when the information available at the sink results in an error greater than a certain threshold. The authors in [8] show that, by using this approach, the sources can save energy by sending information only when necessary. However, the approach requires that each of the sources calculates the estimated value of the observed phenomenon, which may present a challenge for sensors with limited processing capability. Another option is a model-based active sampling technique, which relies on models used only at the sink. In this case, the sink has to control when sources send an update. As demonstrated in [9], the energy savings are two-fold: fewer transmissions and less energy needed to capture the data. However, the downside of this approach is that the source has to constantly listen to discover when to send an update. In our work, we do not rely on any model

at the source or the sink, but rather focus on analysing the timeliness of updates to improve the energy efficiency of the system.

Recently a few approaches relying on correlated information to preserve energy were proposed. In [10] and [11], the authors propose updating mechanisms, developed independently of, but closely related to, our work. In [10], the authors propose an algorithm that clusters sources based on their spatial and temporal correlation. Each time an update is required a different source in a cluster will transmit it, thus conserving energy and simultaneously maintaining real-time reporting. Similarly, in [11], the authors propose a scheduling algorithm that groups sources based on their physical locations to conserve energy. In Table I we present key characteristics of the above-mentioned updating mechanisms and compare them to our own. The most noticeable difference is that we do not rely on a scheduling algorithm. Instead, we calculate the optimal rate at which each source should produce updates. Moreover, our approach is data-driven, making it ready to utilise machine learning methods. The trade-off is the same for all cases: the greater the accuracy required in the data gathering, the more energy sources will use. Additionally, our work provides new insights for a network of correlated sources, using a metric called the AoI.

TABLE I: Qualitative comparison with the state of the art

	Our approach	[10]	[11]
Mechanism	Calculating update rates	Scheduling algorithm	Scheduling algorithm
Main advantage	Data-driven	Real-time reporting	Adaptiveness
Main drawback	Requires data	Limited to mesh networks	Limited to ZigBee Technology
Trade-off	Accuracy vs energy	Accuracy vs energy	Accuracy vs energy

The concept of AoI was introduced in [12] to quantify the freshness of information as a metric of interest in vehicular networks. With age, it is possible to determine how the outdatedness of information impacts the decision-making process of services or applications using that information. Various age-based metrics have been proposed to measure the impact of outdated information on the system. Analysing a system using even a basic metric such as the average age can lead to interesting results. The work in [13] considers a pair of sources with a first-come-first-serve M/M/1 queue and analyses the average age. They establish the equilibrium in which both sources achieve the optimal average age, given the update rate of the other source. In [14], the authors demonstrate that when considering age, the status updates should be processed on a last-come-first-served basis because it is expected that the packet at the end of the queue is the most up to date. The peak age, the maximum value of age before the next update from the source is received, was defined in [15]. The peak age provides a better measure for a system as individual updates may be very outdated, thus greatly impacting the decision-

making process, while the average age may be low due to a small number of fresh updates. An alternative is to introduce a penalty function to represent dissatisfaction with out-of-date updates, and propose a status update policy using an algorithm which minimises this dissatisfaction as shown in [16]. The AoI is relevant in any system in which sources are not able to constantly provide fresh information, usually due to contention for the transmission channel, and yet the system has to act based on the available information.

In a system of correlated sources, the system can act on the outdated information from a dedicated source or use fresher information from a correlated source. In [17], we analysed a pair of correlated sources using a separable covariance model and established both the minimal and optimal time shifts between sources' updates, indicating when it is beneficial for the system to use an update from the correlated source. Moreover, we analytically calculated the maximum gain from using correlated information and showed the trade-off between reduced error gain and update rate, which must be considered when setting the optimal updating strategy. In this paper, we expand on our work by analysing the system with two additional non-separable covariance models to show the impact of different models of correlation on the optimal scheduling of updates from correlated sources. Furthermore, we analyse data, collected in a real sensor network, to provide intuition regarding the use of covariance models, and compare the energy savings achieved by our solution to related work in the recent literature.

III. SYSTEM MODEL

A. System model

Our system consists of two sensor nodes periodically sending updates to the gateway, regarding the observed physical phenomenon distributed in space and evolving in time. We represent the observed phenomenon at a position \mathbf{x} in 3-dimensional space, at a given time instant t , with $Z(\mathbf{x}, t)$. The observations are made available to the gateway by means of status update messages, which carry the observation and a time stamp indicating when the observation was made. We quantify the timeliness of the received updates using the AoI, denoted as $\Delta(t)$.

The AoI is defined as the time that has elapsed at the gateway since the last update from a given source [15], meaning that in a system where sources update deterministically, the age will also be deterministic. For the i -th source the age is as follows:

$$\Delta_i(t) := t - t_i, \quad (1)$$

where $i \in \{1, 2\}$, and t_i represents the time of the last update generated by source S_i . We assume that each sensor node sends updates periodically at a rate λ_i , i.e. the interval between consecutive updates is $1/\lambda_i$. In such a case, the AoI is also deterministic and varies between 0 and $1/\lambda_i$.

The system can estimate the value of the observed physical phenomenon at the desired location using aged information originating from one of the two available information sources.

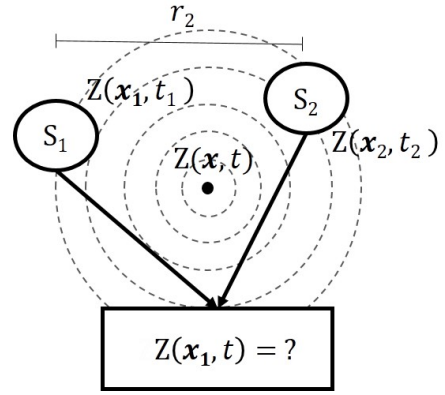


Fig. 1: Two correlated sources observing the same physical phenomenon and sending the information back to the sink.

The accuracy of estimation also depends on the distance between the desired location of estimation and the location of the information source. With r_i we denote the Euclidean distance between the S_i source and the exact location at which the system estimates the value of observed physical phenomena. Our system as represented in Figure 1 is interested in estimating the value of $Z(\mathbf{x}_1, t)$ at the location of information source S_1 , meaning that updates sent by S_1 are perfectly spatially correlated to the desired value of $Z(\mathbf{x}_1, t)$, i.e. $r_1 = 0$. The only effect to be taken into account, in this case, is the age of the available information, i.e. $\Delta_1(t)$. The information from the correlated source S_2 may be outdated, i.e. $\Delta_2(t) \geq 0$, and separated in space, i.e. $r_2 > 0$. As a result, the information sent by S_2 is subject to the effects of spatial correlation and aging. The system may use outdated information $Z(\mathbf{x}_1, t_1)$ from source S_1 or outdated and spatially correlated information $Z(\mathbf{x}_2, t_2)$ from source S_2 , when estimating the value of $Z(\mathbf{x}_1, t)$. Given the most recent observation available, the estimator which minimises the mean square error is the conditional expectation:

$$\hat{Z}_i = \mathbf{E}[Z(\mathbf{x}_1, t) | Z(\mathbf{x}_i, t_i)], \quad i \in \{1, 2\}. \quad (2)$$

The conditional variance is proportional to $1 - \rho_i^2$, where ρ_i is the correlation coefficient, defined assuming unit variance as

$$\rho_i := \text{cov}(Z(\mathbf{x}_1, t), Z(\mathbf{x}_i, t_i)), \quad (3)$$

which is a function of the age and the spatial separation between the selected sources.

B. Covariance models

A covariance model can be used to describe how information from two correlated sources observing the same physical phenomena jointly varies over time and/or space. We consider that the observed phenomenon can be represented by a stationary Gaussian process. Therefore, the observations from the two sources $Z(\mathbf{x}_1, t_1)$ and $Z(\mathbf{x}_2, t_2)$ are assumed to be jointly

Gaussian. The covariance indicates how likely it is that if the value of the observed phenomenon changed at the distant correlated source, i.e. S_2 , the value at the point of interest, i.e. S_1 , has also changed, and vice versa. In our analysis, we consider a separable covariance model [4]:

$$\rho_i^I(r_i, t_i) = \exp(-br_i - a\Delta_i(t)), \quad (4)$$

and two non-separable models [3]:

$$\rho_i^{II}(r_i, t_i) = \frac{a\Delta_i(t) + 1}{\left((a\Delta_i(t) + 1)^2 + b^2r_i^2 \right)^{(d+1)/2}}, \quad (5)$$

$$\rho_i^{III}(r_i, t_i) = \exp(-b^2r_i^2 - a\Delta_i(t) - c\Delta_i(t)r_i^2). \quad (6)$$

Note that d in (5) refers to dimensions: we consider 3-dimensional space, i.e. $d = 3$. In the models above, a , b , and c are the scaling parameters.

Covariance functions provide tractable mathematical modelling for the spatial-temporal variation of the observed physical phenomenon. However, there is a significant difference between the separable and the non-separable class regarding mathematical tractability. With a separable covariance class, it is possible to formulate the covariance as a separate product of only the spatial covariance and only the temporal covariance, while for non-separable classes the spatial and temporal components are inseparable. In general, the non-separable class models space-time interactions and is therefore applicable to a broader range of data gathering scenarios in comparison to the separable class. However, as we illustrate in Section VI, all three covariance models defined above can be applicable in realistic scenarios

With the correlation between the two sources established, it is possible to determine the estimation error resulting from the use of spatially and temporally correlated information.

C. System estimation error

Whenever the system uses outdated information to estimate the current value of the physical phenomenon, it incurs an estimation error, which depends on the age and the location of the used information. We consider that the observed phenomenon can be represented by a stationary Gaussian process. Therefore, status updates sent from two information sources are assumed to be jointly Gaussian. We can then define estimation error as:

$$\varepsilon_i(r_i, t) = 1 - \rho_i^2(r_i, t_i). \quad (7)$$

Note that ρ_i represents one of the covariance models defined in previous subsection, i.e. ρ_i^I , ρ_i^{II} , or ρ_i^{III} . With ε_i we calculate the estimation error for the last update received from the i -th source. The higher the error, the lower the value of the information in the status update, and the less likely it is that the system will use that update.

The system always uses the information from the source which results in the lower estimation error, meaning that the system estimation error is:

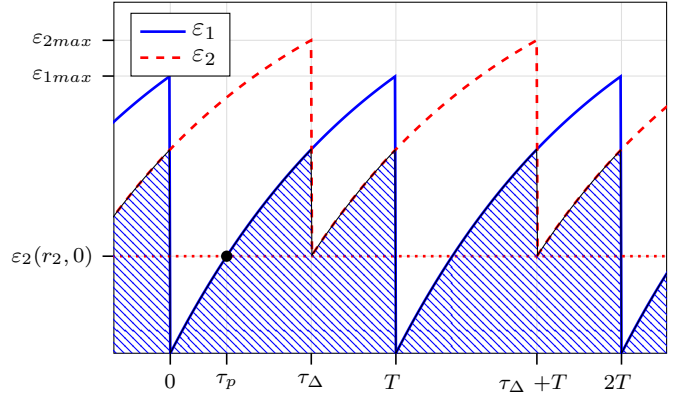


Fig. 2: Estimation error change over two status updates from S_1 , for two correlated sources with equal update rates. At $t = 0$, S_1 sends an update, followed by an update from S_2 with time shift τ_Δ . The dashed area is the system estimation error. The covariance model adopted here is the one in Equation (4).

$$\varepsilon_{sys}(r, t) = \min(\varepsilon_1(0, t), \varepsilon_2(r_2, t)). \quad (8)$$

Figure 2 illustrates how the estimation error changes over time when both sources have the same update rate, with (4) describing the correlation between sources. Let T_1 denote the period of updates from source S_1 , i.e. $T_1 = 1/\lambda_1$. Every time that S_1 transmits an update, ε_1 drops to zero. As information becomes more outdated, the error slowly rises according to the expression in (9). We denote T_2 as the status update period for S_2 . Updates for S_2 are transmitted at time instants $\tau_\Delta + jT_2, j \in \{1, 2, \dots\}$, at which times $\Delta_2(t) = 0$, and the error based on information received from S_2 is $\varepsilon_2(r_2, 0) = 1 - \exp(-2br_2)$ and then slowly rises according to the expression in (10). In Figure 2, the system estimation error is zero every time S_1 sends an update and then increases along with expression (9) until an update from S_2 is received, after which the system estimation error equals the error corresponding to the information from S_2 , until S_1 updates again and repeats the cycle.

$$\varepsilon_1(0, t) = 1 - \exp(-2a\Delta_1(t)) \quad (9)$$

$$\varepsilon_2(r_2, t) = 1 - \exp(-2br_2 - 2a\Delta_2(t)). \quad (10)$$

In the next sections, we will analyse the proposed system and determine when sources should send updates such that the system estimation error will be minimal.

IV. EQUAL UPDATE RATES

In this section, we consider a system in which both sources update with equal update rates, i.e. $\lambda_1 = \lambda_2$, meaning that $T_1 = T_2 = T$. In such a case, the system estimation error depends only on the time shift between the sources' updates. We denote this time shift as τ_Δ , which may take values in

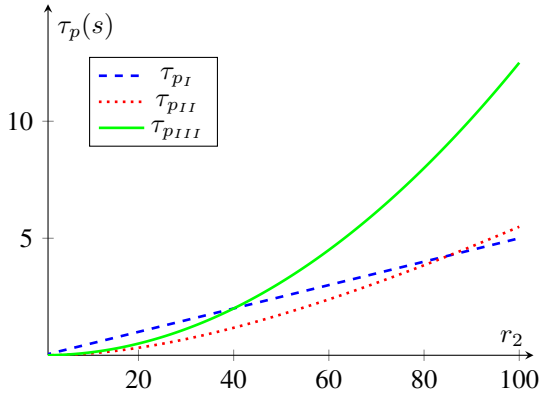


Fig. 3: The graph shows the increase of τ_p with the distance r_2 between information sources, for all three covariance models. The graph was obtained using scaling parameters $a = 0.5$, $b = 0.025$, and $c = 0.0125$.

$[0, T)$. In Figure 2, when S_2 sends its first update, the age of information from S_1 equals τ_Δ , i.e. $\Delta_1(t) = \tau_\Delta$. Whenever S_2 produces an update, this update may incur a higher or lower estimation error than that from the latest update from S_1 , depending on τ_Δ and the correlation between the two sources. As seen in Figure 2, the error at the time of an update from S_2 equals $\varepsilon_2(r_2, 0)$.

An update from the correlated source can replace only sufficiently aged updates from the source of interest. We refer to τ_p as the minimal age an update from S_1 needs to have so that an update from S_2 can reduce the system estimation error. When S_2 updates with time shift τ_p , i.e. $\tau_\Delta = \tau_p$, the estimation error from S_2 at $\Delta_2 = 0$ equals the error from source S_1 at time τ_p . Hence, we can calculate τ_p by setting:

$$\varepsilon_2(r_2, 0) = \varepsilon_1(0, \tau_p). \quad (11)$$

τ_p results for ρ^I , ρ^{II} , and ρ^{III} , respectively, are:

$$\tau_{pI} = \frac{br_2}{a}, \quad (12)$$

$$\tau_{pII} = \frac{(1 + b^2 r_2^2)^{\frac{2}{3}} - 1}{a}, \quad (13)$$

$$\tau_{pIII} = \frac{b^2 r_2^2}{a}. \quad (14)$$

τ_p is independent of update rates and depends only on the scaling parameters for the covariance model and the spatial separation between sources. Furthermore, it imposes a condition on the system, that for the information from the correlated source to be beneficial in reducing the system estimation error, the update period for the two sources has to be higher than τ_p , i.e. $T > \tau_p$.

With an increase in the distance between the information sources, τ_p always increases, regardless of which covariance model describes the correlation between the sources, as shown

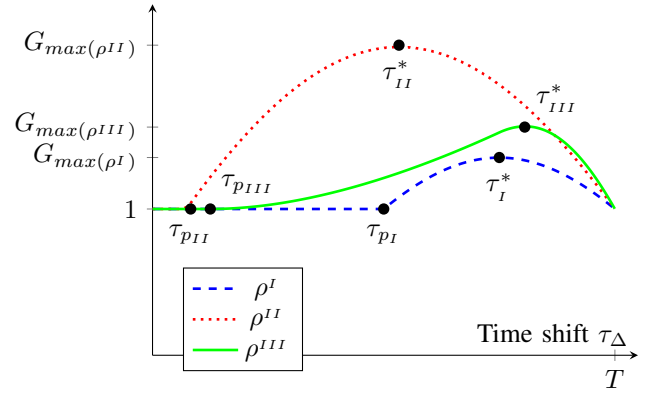


Fig. 4: Gain change for two correlated sources with equal update rates, as the time shift between consecutive updates from the two sources varies from zero to T , for three different covariance classes. The graph was obtained using scaling parameters $a = 0.5$, $b = 0.025$, $c = 0.0125$, and $T = 1$, with distance $r_2 = 10$.

in Figure 3. When the correlation can be described by the separable model ρ^I , τ_p increases linearly, while with non-separable covariance models ρ^{II} and ρ^{III} the increase is exponential. In the case depicted in Figure 3, the scaling parameter a is larger than parameter b , meaning that sensors are more temporally than spatially correlated. The scaling parameters have a direct impact on the τ_p value. The more spatially correlated sources are, i.e. $b \gg 0$, the higher the value of τ_p . However, regardless of the scaling parameter values, the curve shape remains the same. The increase of τ_p with distance illustrates how the shorter the update period, the closer together the sources must be for the updates from S_2 to be advantageous in reducing the system estimation error.

To calculate the system estimation error when two correlated sources update with equal update rate, it is necessary to integrate $\varepsilon_{sys}(r, t)$ over one period T , which can be expressed as:

$$\bar{\varepsilon}_{sys} = \frac{1}{T} \left(\int_0^{\tau_\Delta} \varepsilon_1(0, t) dt + \int_{\tau_\Delta}^T \varepsilon_2(r_2, t) dt \right), \quad (15)$$

where, for $0 \leq t \leq T$:

$$\Delta_1(t) = t \text{ and } \Delta_2(t) = t - \tau_\Delta,$$

and with the condition:

$$\tau_\Delta \geq \tau_p.$$

The system estimation error varies with the time shift τ_Δ between the updates from the two information sources. To illustrate this effect, we define the *error gain*, which captures the impact that the updates from S_2 have on reducing the system estimation error. We define the gain as the ratio between the average error if the system relied solely on updates from S_1 to the system estimation error when updates from S_2 are also considered:

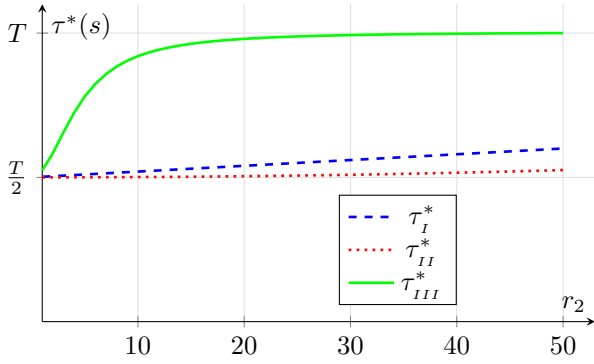


Fig. 5: The graph shows the increase of τ^* with the distance r_2 between the two information sources, for all three covariance models. The graph was obtained using scaling parameters $a = 0.5$, $b = 0.1$, $c = 0.0125$, and $T = 50$.

$$G(r, t) = \frac{\bar{\varepsilon}_1(0, t)}{\bar{\varepsilon}_{sys}(r, t)}. \quad (16)$$

Figure 4 depicts the change in the gain for different τ_Δ values. For $\tau_\Delta < \tau_p$, the gain is one, and the graph exhibits a plateau, meaning that an update from S_2 does not reduce the system estimation error. When $\tau_\Delta > \tau_p$, the update from S_2 reduces the error and the gain increases. At $\tau_\Delta = \tau^*$, the system generates the lowest system estimation error and the highest gain. When the time shift for the update from S_2 is greater than τ^* , the gain decreases again. The reason for such behaviour lies in the fact that the error caused by outdated information from S_1 is no longer effectively compensated by an update from S_2 . Furthermore, the scaling parameters influence the values of τ_p and τ^* ; however, they do not affect the shape of the curve. Higher dependence on spatial separation in the second non-separable covariance model ρ^{III} causes non-symmetrical behaviour as seen in Figure 4. For all three covariance models, we can establish that there is always an optimal time shift, i.e. $\tau_\Delta = \tau^*$, for which receiving an update from a correlated source is most beneficial.

To determine τ^* we calculate the derivative of (15), i.e. $d\bar{\varepsilon}_{sys}/d\tau_\Delta$. For separable model ρ^I and non-separable model ρ^{III} , we are able to derive closed-form expressions for the optimal time shift τ^* (equations (17) and (18)). For the second non-separable model ρ^{II} we cannot produce a closed form expression, and instead used a numerical solution to obtain Figure 5.

$$\tau_I^* = \frac{r_2 b + T a}{2a}, \quad (17)$$

$$\tau_{III}^* = \frac{(c r_2^2 + a) T + b^2 r_2^2}{c r_2^2 + 2a}. \quad (18)$$

The optimal time shift depends on the update period, the distance between sources r_2 , and the value of the scaling parameters. With an increase in distance the value of τ^* increases, but the value is always relative to T , as illustrated in

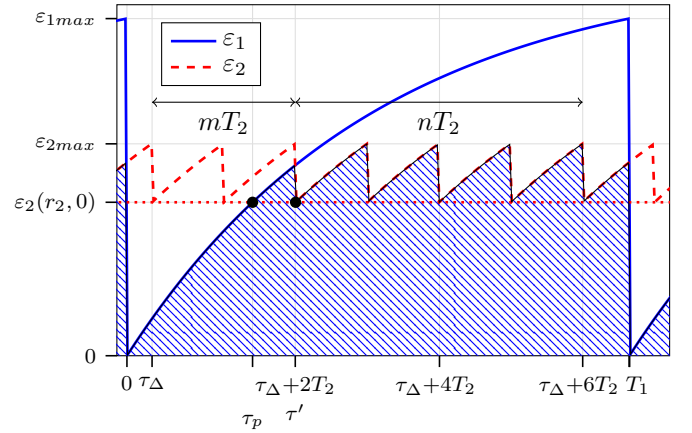


Fig. 6: Estimation error change over one period of T_1 for two sources when S_1 updates less often than S_2 , i.e. $\lambda_1 < \lambda_2$. The dashed area represents the system estimation error. The covariance model adopted here is the one in Equation (4).

Figure 5. For the separable model (4), it is clear from Equation (17) that τ^* is always larger than $T/2$. The same applies for non-separable model (5), which we determined numerically. The increase of τ^* is linear with distance for both models. The second non-separable model (6) has no such limitation in the value of τ^* and as the distance between the sources increases τ^* approaches T .

In summary, the observed system has two notable points of interest, τ_p and τ^* , and a desirable condition: $\tau_\Delta > \tau_p$. τ_p represents the minimal time shift the update from the correlated source should have to reduce the system estimation error. Furthermore, τ_p is independent of update rates. τ^* denotes the time shift at which a received update from the correlated source would result in the minimal system estimation error. The system dynamics change when sources update with different update rates, as we explore in the next section.

V. DIFFERENT UPDATE RATES

Information sources observing the same physical phenomenon in an IoT network may, for many reasons, update with different update rates. For example, one of the sources may transmit more frequently because the cost of its transmission is lower than the cost of transmission of the other (e.g. it may rely on a more stable energy source). In this section, we analyse two possible cases, namely when S_1 updates more or less frequently than S_2 .

A. S_1 updates less frequently than S_2

When S_1 updates less frequently than S_2 , i.e. $\lambda_1 < \lambda_2$, the system estimation error changes depending on the update rate of S_2 and the time shift τ_Δ . After S_1 sends an update, there is a time period during which any update sent from S_2 is irrelevant, because it does not reduce the error. As illustrated in Figure 6, the first m updates from S_2 do not influence the average error value. The time period during which updates from S_2

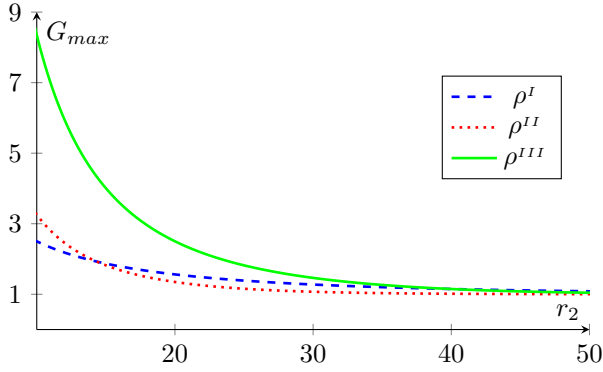


Fig. 7: The graph shows the decrease of G_{max} with the distance r_2 between information sources, for all three covariance models. The graph was obtained using scaling parameters $a = 0.5$, $b = 0.025$, $c = 0.0125$, and $T_1 = 50$.

are irrelevant is described as τ_p . Note that τ_p is independent of update rates and is calculated as shown in (11).

τ' represents the time shift between an update from S_1 and the first update from S_2 which reduces the system estimation error (see Figure 6). Before that update, the first m updates from S_2 are irrelevant as they do not reduce the error. The number n represents the number of full transmissions from S_2 between τ_p and the next update from S_1 . For example, in Figure 6, $n = 4$. It is necessary to integrate $\varepsilon_2(r_2, t)$ n times over one period of T_2 , as every time S_2 sends an update $\Delta_2(t) = 0$. What remains is only part of an S_2 update which reduces the system estimation error only partially because an update from S_1 intervenes.

For the source of interest the benefit of using updates from a correlated source is limited, as an update from correlated source can reduce estimation only to a value of $\varepsilon_2(r_2, 0)$. This limit can be calculated by increasing k to infinity. The result is the maximum possible reduction in the system estimation error, as presented in Equations (20), (21), and (22), for each covariance model. In all cases, the maximum gain falls rapidly with distance as presented in Figure 7. G_{max} depends only on the scaling parameters and the distance between the sources.

$$G_{max}(r, t) = \frac{\bar{\varepsilon}_1(0, t)}{\lim_{\lambda_2 \rightarrow \infty} \bar{\varepsilon}(r, t)}, \quad (19)$$

and the limits when sources are correlated according to models ρ^I , ρ^{II} , and ρ^{III} , respectively, are:

$$\lim_{\substack{\lambda_2 \rightarrow \infty \\ T_2 \rightarrow 0}} \bar{\varepsilon}_I(r, t) = \frac{1}{T_1} \left(\tau_{pI} + \frac{\exp(-2a\tau_p) - 1}{2a} + (T_1 - \tau_{pI})(1 - \exp(-2r_2b)) \right), \quad (20)$$

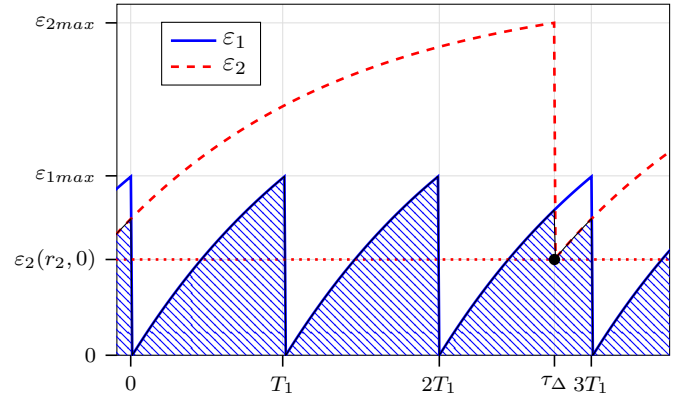


Fig. 8: Estimation error change over one period of T_2 for two sources when S_1 updates more often than S_2 , i.e. $\lambda_1 > \lambda_2$. The dashed area represents the system estimation error. The covariance model adopted here is the one in Equation (4).

$$\lim_{\substack{\lambda_2 \rightarrow \infty \\ T_2 \rightarrow 0}} \bar{\varepsilon}_{II}(r, t) = \frac{1}{T_1} \left(\tau_{pII} + \frac{1}{5a(a\tau_{pII} + 1)^5} - \frac{1}{5a} + (T_1 - \tau_{pII}) \left(1 - \frac{1}{(1 + 2b^2r_2^2)^4} \right) \right), \quad (21)$$

$$\lim_{\substack{\lambda_2 \rightarrow \infty \\ T_2 \rightarrow 0}} \bar{\varepsilon}_{III}(r, t) = \frac{1}{T_1} \left(\tau_{pIII} + \frac{\exp(-2a\tau_p) - 1}{2a} + (T_1 - \tau_{pIII}) (1 - \exp(-2b^2r_2^2)) \right). \quad (22)$$

Whenever the correlated source updates more often, only updates which effectively reduce the system estimation error are beneficial to the system. In general, the more updates S_2 sends, the greater is the gain of using updates from the correlated source. However, due to information being collected at a different location, there is a limit to the gain, but as demonstrated in this subsection, using multiple updates from the correlated source can be useful to the overall performance of the system.

B. S_1 updates more frequently than S_2

When S_1 updates more frequently than S_2 , i.e. $\lambda_1 > \lambda_2$, using correlated information becomes less beneficial. In such a case, an update from S_2 reduces the system estimation error only every few updates from S_1 , as illustrated in Figure 8, where updates from S_2 reduce the system estimation error every third update from S_1 . Furthermore, the update from the S_2 has to arrive when the error, due to outdated information from S_1 , is higher than $\varepsilon_2(r_2, 0)$, i.e. $\varepsilon_1(\Delta_1(t), 0) > \varepsilon_2(r_2, 0)$ with $\Delta_1(t) = \tau_\Delta$. In this case, the system estimation error depends mostly on updates from S_1 .

An update from S_2 can reduce only a fraction of the overall system estimation error as depicted in Figure 8. The higher the

update rate for S_1 , the lower the benefit of using updates from S_2 is. Furthermore, the timing of updates from the correlated source plays a crucial part whenever the source of interest updates more often than the correlated source. It may happen that updates from S_2 will not reduce the system estimation error; in such a case, an update from S_2 arrives with delay smaller than τ_p after S_1 sends update, i.e. $\tau_\Delta < \tau_p$. The higher the frequency of updates from the source of interest is, the lower the chance that an update from the correlated source will arrive at the right time. However, updates from the correlated source may still be beneficial to the system, provided they arrive at the right moment.

VI. EVALUATION

In this section, we bridge the gap between the theoretical analysis presented in the previous two sections and the real world. In the first part of this section, we analyse data provided by the Intel Berkeley Research laboratory [18] to extract scaling parameters for the covariance models described earlier. In the second part, we use the covariance model with the extracted scaling parameters to demonstrate how the lifespan of the source of interest, S_1 , can be extended by using updates from a correlated source, S_2 . We conclude the section by comparing our approach to a clustering algorithm presented in the literature.

A. Data analysis

In our analysis, we use temperature and humidity measurements collected from nine sensors deployed at Intel Berkeley Research laboratory [18], University of California. The data was gathered in March 2004. We focus on a subset of available sensors residing in one large open space office. Sensors deployed in the same room are subject to roughly the same environmental factors, e.g. air circulation, thus ensuring that causation for humidity and temperature changes is the same for every sensor we have selected in our analysis. Sensors transmit an update twice in every minute. To ensure time domain consistency, we average multiple temperature and humidity readings reported by each sensor over an hour.

We calculate empirical space correlation for hourly averages of measurements from the data set using Pearson's correlation coefficient formula for samples. We start by calculating the spatial correlation for time lag zero. At time lag zero, we compare measurements taken by each sensor at the same hour. Each point in Figure 9(a) and 9(c) represents the empirical spatial correlation coefficient for one pair of sensors separated by distance r . Nine sensors yield 45 different pairs; of those, nine pairs are a sensor paired with itself. For such a pair, correlation at time lag zero is one. We repeat this process for multiple time lags, i.e. compare measurements taken from different sensors hours apart, to calculate temporal correlation.

To obtain temporal correlation, we calculate the average of empirical spatial correlation coefficients for a specific time-lag. For example, we obtain the temporal correlation at time zero in Figure 9(b) by calculating the average of spatial correlation coefficient points in Figure 9(a). As can be seen in Figure 9(b) and 9(d), the higher the time lag between the data points, the

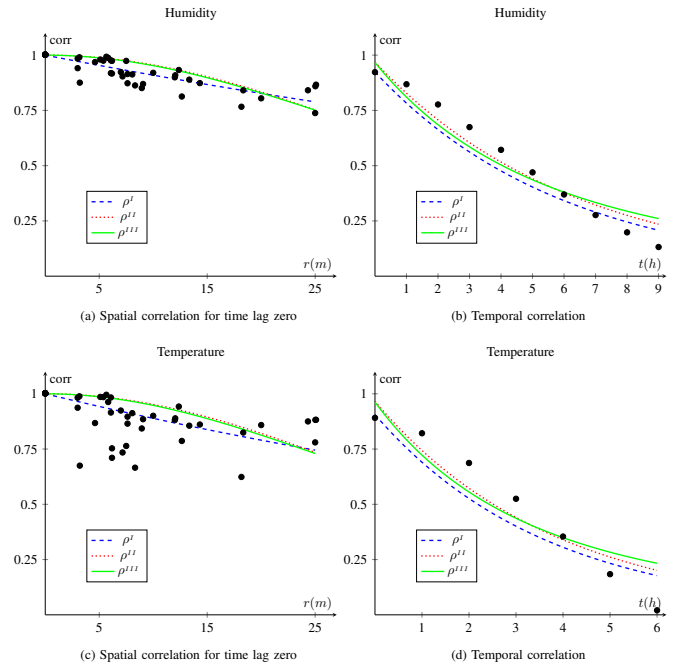


Fig. 9: Empirical space and time correlation points plotted alongside covariance models with scaling parameters extracted from the data.

lower the empirical correlation. The empirical graphs show better temporal correlation for humidity in comparison to temperature. One noticeable difference is that humidity data is correlated even when time lag is nine hours, in contrast to temperature, which has negligible correlation after around six hours.

With correlation coefficient points determined, the last step in our data analysis is to fit models to calculated points. All three covariance models (4), (5), and (6) provide a reasonable fit to the data. In Figure 9 we plot all three covariance models, with scaling parameters extracted from the data, alongside the empirical space-time correlation points. For example, the separable covariance model (4) should have scaling parameter values $a = 0.272$ and $b = 0.012$ to describe temperature spatial-temporal correlation. The pattern repeats in non-separable covariance models where the temporal scaling parameter, a , is also larger than the spatial scaling parameter, b . The same is true for humidity, where the non-separable model (5) scaling parameters are $b = 0.016$ and $a = 0.062$ to best describe spatial-temporal correlation. In general, scaling parameters describing humidity correlation have lower values in comparison to scaling parameters describing temperature correlation. In the next subsection, we adopt the separable covariance model, i.e (4), for temperature and the first non-separable model, i.e. (5), for humidity, with scaling parameters values mentioned above.

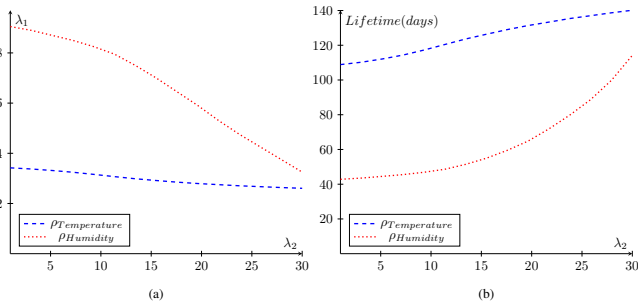


Fig. 10: (Left) The decrease in update rate for S_1 when S_2 update rate increases while the system estimation error is fixed. (Right) The increase in expected device lifetime by taking advantage of correlation as a function λ_2 , as the system estimation error remains fixed. Both graphs were obtained by using distance $r_2 = 4m$.

B. Extending lifespan

Sensing devices consume energy to create information, i.e. measure the observed physical phenomena. In particular, battery-powered devices are of great concern as every update brings the device closer to expiration. By using updates from the correlated source, the source of interest, S_1 , can reduce its update rate and consequently extend its lifespan at the expense of the correlated source. The lifespan of a source is measured from the moment the device is deployed in the network, to the moment the source is considered non-functional. We consider a source non-functional when the source depletes all its available energy and ceases to transmit updates. We model source lifetime \mathcal{L} as in [19]:

$$\mathbb{E}[\mathcal{L}] = \frac{\mathbb{E}_0}{P_c + \lambda \mathbb{E}[E_r]}. \quad (23)$$

\mathbb{E}_0 represents the initial total non-rechargeable energy. P_c is the continuous power consumption that the information source uses for sensing and basic operations. $\mathbb{E}[E_r]$ denotes the expected energy needed to transmit a status update, and λ represents the source update rate. In our case, we selected the lifetime model parameters to reflect the power consumption in a battery-powered device. We base our power consumption parameters on the Mica2Dot board, a sensor board used in Intel Berkeley Research laboratory deployment [18]. We estimate that the sensor requires five milliseconds to produce the measurements and that the sensor's radio is active for one millisecond to transmit the update. Using information from the datasheet we determine that $P_c = 45\mu W$ and $0.2mJ$ of energy is required to obtain and transmit the measurement, i.e. $\mathbb{E}[E_r] = 0.2mJ$. We consider that the sensor would be powered by a $620mAh$ Lithium coin battery suitable for the desired application and device type. Note that λ in our case refers to the number of updates by a sensor transmitted in an hour.

In our experiment, we focus on a system with a pair of correlated sources as described in the previous sections. The

system has a primary interest in information gathered from one source, S_1 , and can use updates from a correlated source, S_2 , to prolong the lifetime of the source of interest. An example of such a system would be a battery-powered sensor S_1 deployed directly at the center of the observed physical phenomenon, with S_2 deployed as a backup at a more distant location but with a connection to the power grid. In such a case, the system can use updates from S_2 to reduce the number of updates sent by S_1 and prolong its lifetime while simultaneously maintaining the required accuracy of information, i.e. keep the system estimation error below a certain threshold. In other words, the source of interest S_1 can reduce its update rate because information received from correlated source S_2 is reducing the system estimation error. We set the system estimation error value to 7.5% for observing temperature and 2% for humidity. To obtain results presented in Figure 10 we calculated the required update rate from source S_1 , λ_1 , to satisfy the system estimation error requirement depending on the S_2 update rate, λ_2 .

The higher the rate of updates from the correlated source S_2 is, the lower the required rate of updates from the source of interest and the longer its expected lifetime. Figure 10 shows the change of update rates and the lifetime increase of the sensors by taking advantage of correlated updates, for cases of temperature and humidity. In both cases, significant improvement in the lifetime of sensors is achievable. For humidity, benefits are more significant because humidity exhibits better spatial-temporal correlation in comparison to temperature. As noted in the previous subsection, covariance scaling parameters for humidity were significantly smaller in comparison to temperature, indicating that humidity variation over time and space is much lower. Consequently, the gain of using correlated information is much higher for humidity.

Translating our experiment into a real world deployment where it is expected that information for temperature and humidity are transmitted simultaneously, the correlated source S_2 should transmit 30 updates, while battery-powered source S_1 would transmit only three updates in a hour. Such a system is capable of gathering information accurately for three months. Our results show that even a small decrease in the device update rate has a notable impact on the device lifetime. The gain depends on the variation of the observed physical phenomenon across time and space. In our example, the temperature data exhibit high variation, yet sensor lifetime increase is 20%. Whenever data is more highly correlated, e.g. humidity, relying on more frequent updates from a correlated sensor can more than double the source lifetime.

C. Energy Saving Comparison

In this subsection, we compare our energy savings approach with the approach presented in [11]. Both approaches have to meet the same accuracy constraint, meaning that the estimation error threshold is equal in both cases. We set the accuracy constraint to 95%, in other words, we set the maximum estimation error to 5%. Note that we calculate estimation error using Equation (7). For our approach, we calculate the update rate such that each source satisfies the set accuracy constraint.

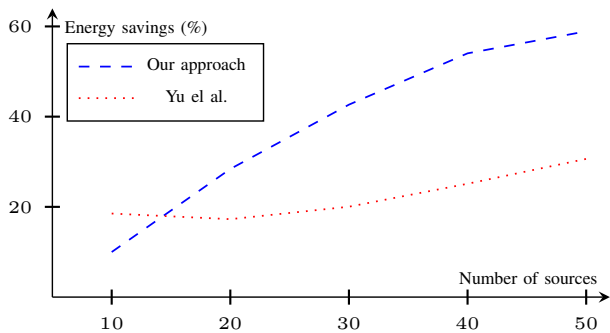


Fig. 11: Average energy saved per source by using information from correlated information sources in comparison to energy saved used by clustering and scheduling algorithm proposed in the literature.

In our implementation of the approach presented in [11] we try to reproduce their scheduling and clustering algorithm as it would work in our simulated sensor network. We achieve that by clustering sources based on their proximity to cluster heads; each source joins the cluster group of the cluster head with the shortest distance to it. Once the clusters are formed, we calculate the super cycle. A supercycle is a period within which every source in a cluster sends one update to the sink, while still meeting the information accuracy requirements. We schedule the transmission of source updates evenly throughout the super cycle. This means that sources' update rate is the reciprocal of the time of the supercycle, i.e. $\lambda_c = 1/T_{sc}$. The first source to transmit in the cycle is the cluster head, followed by the nearest source, followed by second nearest and so forth. With such an implementation, we effectively mimic the proposed approach.

The source's energy consumption is directly related to its update rate in both approaches. We calculate the energy savings (P_{es}) as the ratio between P_0 and P_n as follows:

$$P_{es} = \frac{P_0 - P_n}{P_0} \times 100\%. \quad (24)$$

P_0 represents the energy a source uses when transmitting with an update rate that enables the source to meet the required accuracy constraint on its own. Furthermore, P_0 is the same for every source in both cases. P_n is the energy sources use when taking advantage of correlated sources or when using the approach in [11]. We calculate P_n using an update rate we obtain using simulation. We calculate both P_0 and P_n as $P_c + \lambda \mathbb{E}[E_r]$, with the values for P_c and $\mathbb{E}[E_r]$ as we used in the previous subsection.

Sources in our simulation are uniformly randomly distributed in a room of size 15×15 m. Such a room size is comparable to the room size in the Intel lab deployment where the sensor data was collected. Furthermore, we use non-separable covariance model (5) with scaling parameters extracted from humidity data. In [11] the authors showed that three clusters yield a good result. Therefore we decided to cluster sources into three groups. In our comparison, we use

the average energy consumption of sources in the deployment. In both cases, we assumed no contention for the transmission channel and instantaneous transmission of updates.

In Figure 11 we show the increase in energy savings as the number of sources in the room increases. When the number of sources is low, the clustering approach is very efficient due to the evenly scheduled updates within the supercycle. However, as the number of sources increases the same scheduling becomes a handicap and reduces the efficiency of the clustering approach. As the number of sources increases our approach of taking advantage of correlated information proves to be better. Our result indicates that energy savings can be improved much more than what the current state of the art is offering.

VII. CONCLUDING REMARKS AND FUTURE WORK

In this paper, we analysed two correlated information sources and showed how the different spatial-temporal variation of the observed physical phenomenon influences the optimal use of updates from the two sources. The system employs information from the correlated information source to minimise the system estimation error when estimating the value of an observed physical phenomenon. We established that there exists a minimal time shift between the two sources' updates, for all three tested covariance models, so that the arrival of an update from a correlated source will reduce the system estimation error. We also derived the optimal time shift between the two sources, for all three tested covariance models, for which the system estimation error is minimal. We examined some real data and extracted the covariance model parameters, to provide the reader with a realistic feel for scaling parameters values and the applicability of our analysis in a real scenario. We demonstrated that using correlated information results in a significant increase of device lifetime.

When billions of devices will be connected to one network, as the IoT vision promises, using available correlated information will become essential to improve the performance of energy constrained devices. In our work we focused on extending device lifespan; another gain of our approach is reduced contention for the transmission channel. By relying on the concept of AoI we offered an alternative take on the problem of accuracy, decision-making, and lifespan of devices in a network of sensors. A more accurate decision requires more frequent updates, and more updates lead to a shorter lifespan for battery-powered devices. However, by sharing information and making use of correlated information, benefits such as more accurate decisions, or an increase in devices' lifespan, or reduction of contention for the transmission channel can be obtained.

ACKNOWLEDGMENTS

This work was funded in part by the European Regional Development Fund through the SFI Research Centres Programme under Grant No. 13/RC/2077 SFI CONNECT. We also acknowledge support from the Science Foundation Ireland under Grant No. 10/IN.1/I3007.

REFERENCES

- [1] E. Borgia, "The internet of things vision: Key features, applications and open issues," *Computer Communications*, vol. 54, pp. 1–31, 2014.
- [2] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *IEEE INFOCOM, 2012*, pp. 2731–2735.
- [3] N. Cressie and H.-C. Huang, "Classes of nonseparable, spatio-temporal stationary covariance functions," *Journal of the American Statistical Association*, vol. 94, no. 448, pp. 1330–1339, 1999.
- [4] M. L. Stein, "Space-time covariance functions," *Journal of the American Statistical Association*, vol. 100, no. 469, pp. 310–321, 2005.
- [5] M. C. Vuran, Ö. B. Akan, and I. F. Akyildiz, "Spatio-temporal correlation: theory and applications for wireless sensor networks," *Computer Networks*, vol. 45, no. 3, pp. 245–259, 2004.
- [6] J.-F. Chamberland and V. V. Veeravalli, "How dense should a sensor network be for detection with correlated observations?" *Transactions on Information Theory*, vol. 52, no. 11, pp. 5099–5106, 2006.
- [7] G. Anastasi, M. Conti, M. Di Francesco, and A. Passarella, "Energy conservation in wireless sensor networks: A survey," *Ad hoc networks*, vol. 7, no. 3, pp. 537–568, 2009.
- [8] D. Chu, A. Deshpande, J. M. Hellerstein, and W. Hong, "Approximate data collection in sensor networks using probabilistic models," in *IEEE ICDE, 2006*.
- [9] A. Deshpande, C. Guestrin, S. R. Madden, J. M. Hellerstein, and W. Hong, "Model-driven data acquisition in sensor networks," in *VLDB 2004*, pp. 588–599.
- [10] L. A. Villas, A. Boukerche, D. L. Guidoni, H. A. De Oliveira, R. B. De Araujo, and A. A. Loureiro, "An energy-aware spatio-temporal correlation mechanism to perform efficient data collection in wireless sensor networks," *Computer Communications*, vol. 36, no. 9, pp. 1054–1066, 2013.
- [11] T. Yu, A. M. Akhtar, A. Shami, and X. Wang, "Energy-efficient scheduling mechanism for indoor wireless sensor networks," in *81st IEEE VT Conference, 2015*.
- [12] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *IEEE SECON, 2011*.
- [13] R. D. Yates and S. Kaul, "Real-time status updating: Multiple sources," in *IEEE ISIT, 2012*.
- [14] S. K. Kaul, R. D. Yates, and M. Gruteser, "Status updates through queues," in *IEEE CISS, 2012*.
- [15] M. Costa, M. Codreanu, and A. Ephremides, "On the age of information in status update systems with packet management," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 1897–1910, 2016.
- [16] Y. Sun, E. Uysal-Biyikoglu, R. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," in *IEEE INFOCOM, 2016*.
- [17] J. Hribar, M. Costa, N. Kaminski, and L. A. DaSilva, "Updating strategies in the internet of things by taking advantage of correlated sources," in *IEEE Globecom, 2017*.
- [18] P. Bodik, W. Hong, C. Guestrin, S. Madden, M. Paskin, and R. Thibaux, "Intel lab data," *Online dataset*, 2004.
- [19] Y. Chen and Q. Zhao, "On the lifetime of wireless sensor networks," *Communications letters*, vol. 9, no. 11, pp. 976–978, 2005.

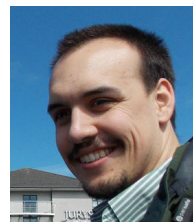


Jernej Hribar is currently a PhD student at the CONNECT Telecommunications Research Centre at Trinity College Dublin (Ireland). He graduated in 2013 at the University of Ljubljana, Faculty of Electrical Engineering. His research focuses on the sustainability of the Internet of Things, particularly on using data analytics to improve networks performance.



Maice Costa received her Ph.D. degree in Electrical Engineering from the University of Maryland, College Park, in 2015. She is the recipient of a Distinguished Graduate Fellowship from the Clark School of Engineering at the University of Maryland, and also the recipient of a RISE Professional Scholarship for training in Germany, from the Deutscher Akademischer Austausch Dienst (DAAD). She has been a visitor researcher at the University of Oulu, Oulu, Finland, an intern at Bell Labs, Alcatel-Lucent, Stuttgart, Germany, and a Postdoctoral Fellow at

Trinity College Dublin, Ireland. She is currently a Research Scientist at Intelligent Automation, Inc. Her research interests are broadly related to wireless networks, including network control and resource management. She is a member of IEEE.



Nicholas Kaminski received the master's degree in electrical engineering and the Ph.D. degree from Virginia Tech, USA, in 2012 and 2014, respectively. He conducts research focused on extending the bounds of wireless technology by deploying targeted intelligence to act in harmony with flexible radio systems. He advances distributed radio intelligence through experimentation-based research and furthers wireless communications by incorporating techniques for understanding communications from complex systems science. During this time he was funded as a Bradley Fellow with the Bradley Department of Electrical and Computer Engineering, Virginia Tech.



Luiz A. DaSilva holds the chair of Telecommunications at Trinity College Dublin, where he is the Director of CONNECT, a research centre on Future Networks funded by the Science Foundation Ireland. Prior to joining Trinity College, Prof DaSilva was with the Bradley Department of Electrical and Computer Engineering at Virginia Tech. His research focuses on distributed and adaptive resource management in wireless networks, and in particular radio resource sharing and the application of game theory to wireless networks. Prof DaSilva is a Fellow of

Trinity College Dublin, an IEEE Communications Society Distinguished Lecturer and a Fellow of the IEEE, for contributions to cognitive networks and to resource management in wireless networks.