Quantifying the variability of power and energy consumption for IoT edge nodes

Salma Tofaily*, Issam Raïs*, Otto Anshus*
Department of Computer Science, UiT The Arctic University of Norway, Tromso, Norway*
Corresponding authors: salma.tofaily, issam.rais@uit.no

***Abstract*—For IoT and edge systems, measuring, predicting and optimizing energy consumption is an open field. It is important to accurately and precisely characterize power and energy consumption of edge nodes, as energy can be a scarce and key resource. However, there are no fine-grain studies that aim at understanding the potential variability of power and energy consumption of edge nodes. Existing research works give minor or no significance to this potential variability.**

This paper addresses this problem by quantifying the variability of power and energy consumption on a single edge node, and among multiple homogeneous edge nodes, for three scenarios: Idle, CPU intensive "matrix product" and RAM intensive "flip". These scenarios are found in edge applications. Identical controlled experiments are repeated thoroughly, for each scenario.

Results show that power and energy variability exist for all studied scenarios. On a single node, power and energy variability measurements are relatively low. On multiple homogeneous nodes, the variability can be significant. For example, for CPU intensive "matrix product", the variability in energy is equivalent to an idle up-time, in a month, of 7 hours and 52 hours, when considering single and multiple homogeneous nodes, respectively.

***Index Terms*—IoT, edge computing, power, energy, consumption, measurement, variability, analysis**

I. Introduction

Systems of small connected nodes are built to solve various problems. They can help for example in monitoring environment [1], [2], health care [3] and crowd-sensing [4]. These systems are part of the "Internet of Things (IoT)". Combined, edge nodes can form "distributed edge systems". The number of IoT systems is massively increasing [5], leading to an inevitable overall increase in energy consumption by edge nodes and systems.

Edge nodes can be resource-constrained, powered by batteries, or with a limited energy budget. As the number of IoT nodes keeps rising, it becomes challenging to be energy efficient and scale (e.g with maintenance and the need of potential batteries replacement). With extreme scenarios, nodes can also be not accessible for several months [1]. Consequently, systems should be frugal when it comes to energy consumption, to limit energy consumption and sustain long operational lifetime. A first step to achieve these goals is to be able to accurately characterize and predict energy consumption for nodes.

Understanding and evaluating possible energy-saving techniques at the scale of small edge nodes being part of a large distributed system is an open field [6]. However, literature gives no or minor importance to potential power and energy consumption variability on single edge node or among homogeneous edge nodes.

Understanding the accuracy and precision of power and energy consumption allows for detailed and accurate studies. It becomes especially critical when: (i) making and using models to predict energy consumption of nodes; (ii) doing comparisons with related works for results that differ within the not considered variability; (iii) the system is deployed in extreme conditions where nodes, not accessible during several months, need to accurately predict their limited lifetime; (iv) using monitoring results to calibrate simulator inputs. Thus, it is crucial to understand the accuracy of power and energy consumption measurements.

In this paper, we study power and energy consumption variability on single and multiple homogeneous edge nodes, for specific scenarios. Experiments are conducted on a remote open access testbed, Fit IoT-LAB [7], for its external and ultra precise energy consumption monitoring. External monitoring is very important for the energy variability study, as it does not add any workload to the nodes.

We conduct our experiments using Raspberry Pies, all 3 B available in the testbed, as these nodes are highly studied in literature for prototyping [8] and deployments of IoT and edge systems [9]. However, fine-grain studies of power and energy variability are not conducted on these types of nodes.

We quantify the variability of idle state, RAM and CPU components. We uniformly stress one specific component several times, which results in power and energy consumption measurements for the same scenario, under controlled conditions. We evaluate the variability of power and energy consumption, for these states and components used in every application. This study is a first step towards understanding potential variability in edge applications.

The contributions of this paper are:

- A thorough and fine-grain study of power and energy variability on a single edge node
- A thorough and fine-grain study of power and energy variability on multiple homogeneous edge nodes

This paper is organized as follows: Section II presents related work. Section III presents experimental setup. Section IV presents results and observations. Section V presents discussions. Finally, section VI presents conclusions and future work.

II. Related Work

Power and energy consumption metrics are used in multiple papers to study specific characteristics of IoT and edge systems [6], [10]–[14]. These papers use values from literature [10], local unique [9] or average [12], [13] measurements. Simulations can be an alternative for experiments [10], [15].

Papers from large-scale infrastructures show that nodes, homogeneous in hardware, can be heterogeneous in power [16]. However, papers generally use average power measurements from a single node [17].

In [18], authors analyze and evaluate variability of energy consumption among different external and internal power monitoring devices, tested with several benchmarks, on two types of nodes: server and desktop. Energy consumption of idle state is subtracted from energy consumption of benchmark runs. However, potential energy variability on a single node is not considered and only one run per experiment is conducted. These papers shows that energy variability has been studied in an established field: large-scale infrastructures. It highlights the need to address this concern in the edge context.

In [19], authors conduct experiments with OpenCV based benchmarks on a Raspberry Pi 3B. Energy consumption is measured using external monitoring. The goal is to conclude optimal configuration, from energy consumption perspective, per benchmark. For some benchmarks (e.g circle detection), measurements from different configurations (e.g single and multi thread) are close to each other (by 2.9412%). Experiments are neither repeated on the node nor conducted on multiple homogeneous nodes.

In [9], authors characterize Raspberry Pi 3B nodes, that are part of an agricultural monitoring wireless sensor network, from a power perspective during different workloads: idle, sensing, transmitting, and file logging. The goal is to estimate minimum and maximum lifespan of the battery powering sensor nodes. Experiments are conducted with 4 nodes, running workloads for 15 seconds and measuring power with external monitoring. Power varies from one node to another. Reported measurements are used in linear regressions to build models. To evaluate the model, authors monitor the remaining percentage of node batteries, once per hour. Linear regression with these measurements is used to build another model, predicting remaining up-time of the sensor nodes. This metric is used as a base to compute the accuracy of the power model, reporting an accuracy of 80%. Power measurements are taken by conducting only one run per workload, on each node. Reported accuracy, and consequently expected lifetimes, needs to be tested on larger sets of experiments.

In [20], authors present a framework to analyze energy consumption on Raspberry Pi 3B+. The framework runs benchmarks stressing several components on a node: processor, memory, storage, connectivity, and network usage. Energy measurements are carried out by external monitoring. System performance metrics are collected. All experiments are made of 3 iterations, 15 minutes each. Average energy consumption is reported and standard deviation is graphically presented.

Energy measurements are analyzed or used in linear regressions to build a model that predicts energy consumption of an application. When analyzing an experiment for one component, estimated energy consumption of other components is subtracted from total energy measurement. The model is claimed to be reusable for other types of nodes after recomputing its constants. To evaluate it, authors run 2 experiments in 2 different scenarios. 2 nodes are used per experiment. Corresponding 4 energy measurements, obtained from different workloads, are compared to the model predictions, to compute 4 different accuracy levels.

Reported model average accuracy, 95%, does not represent all benchmarks. Potential delta in energy consumption, from each component, can influence computed remaining energy consumption. Idle energy measurement, which has the lowest accuracy level (91%), is subtracted from all experiments measurements used to build the model. Conducted experiments are made of only 3 runs. The assumption for having 15 minutes per iteration, with only 3 runs, on a unique node is not validated. Furthermore, the model is calibrated with only one node, evaluated on only two homogeneous nodes, and assumed to be reusable on all nodes of the same hardware. This previous enumeration underlines the fact that the model does not account for possible heterogeneity in energy consumption among multiple homogeneous or single nodes.

In [21], authors propose a methodology to characterize variability of performance and power, on a single edge node and across multiple homogeneous edge nodes, for both NVIDIA Jetson AGX and Nano AI edge platforms. The study is conducted for a set of CNN and Rodinia benchmarks. The experimental protocol tries to limit variability sources (e.g by fixing CPU frequency). The paper focuses on power, without energy. Reported power variability varies from one benchmark to another. As the study is conducted for high level and complex benchmarks, it does not explore the variability for separate node components (e.g CPU and RAM). A fine-grain study, for individual components, can help in understanding different degrees of power variability across benchmarks.

As a summary, characteristics of energy consumption is studied at the edge without a detailed attention to its potential variability, on a single node or multiple homogeneous nodes. This can be noticed in experimental protocols, simulators calibrations, and contributions comparisons. Few papers measure power and energy variability for multiple homogeneous edge nodes, but with only one single run per node, or with few nodes. Paper [21] is the closest to our contribution, as it thoroughly studies power variability. However, it focuses on high level benchmarks. Its results highlight the need to have low level power variability analysis, as a first step for understanding power and energy consumption variability on edge nodes. Furthermore, research work aiming at conducting fine-grain power analysis on a single node do not thoroughly investigate or consider observed variability of energy consumption.

On top of our knowledge, no paper with edge as a context thoroughly and accurately quantifies power or energy variability for specific node components, on a single edge node or homogeneous edge nodes. This paper starts to address these problems by quantifying the variability of power and energy consumption on edge nodes, for fine-grain scenarios stressing specific components of edge nodes.

III. Experimental Setup

This section presents the used testbed, nodes, experimental protocol, stress scenarios, and evaluation metrics.

A. Testbed

Testbeds are important part of literature that facilitate research around IoT and edge systems [7], [22]. We conduct our experiments using a remote large-scale open access testbed, FIT IoT-LAB [7], to conduct automated real experiments with external power monitoring devices. We use the testbed for its ultra precise external power monitoring setup, that does not add overhead to nodes.

The testbed is composed of more than 1500 edge nodes deployed across 6 sites in France. It allows access to monitoring tools and provides SSH connections to nodes, through a front-end. An API allows interactions with the testbed. In our experiments, we use all available Raspberry Pi 3B, 5 nodes at Grenoble, widely used in edge and IoT research [8], [9].

IoT-LAB nodes hardware setup[1]: An IoT-LAB node is an *edge node* connected to a *gateway* and to a *control node*, by its serial ports. The *gateway*[2] is a small Linux computer that is responsible for reprogramming the *edge node* and linking it to the front-end. It also has a *control node* connected, an autonomous on-board system that measures the power of the *edge node* and controls it. This setup makes the edge nodes available for experimentation, from the front-end.

Power measurements: The described set-up guarantees real-time execution of power measurements. The measurements are for the *Raspberry Pi 3 B* and the *gateway's USB port*. The external monitoring setup does not add workload on the node. It does not add variable overhead to power measurements. This is very important for our energy variability study. The power monitoring device is the *INA226*[3]. It is ultra precise: a maximum of 10 µm*V offset* and a maximum of *0.1% Gain Error*. We retrieve one power measurement *every 0.2 second*.

Raspberry Pi 3B nodes: A Raspberry Pi 3B is a single board computer with a quad core 1.2 GHz CPU and a memory of 1 GB RAM. In the testbed, a Raspberry Pi 3B Ethernet interface is used for both power supply and LAN connectivity. Thus, network connection cannot be avoided, as shutting down Ethernet port will cut power. The nodes' supply voltage is measured to be very close to 4.8 V.

There are no other peripherals connected to the nodes. We specifically asked, for our experiments, to remove additional hardware [4] that are usually connected to the nodes.

[1] https://www.iot-lab.info/docs/getting-started/design/
[2] https://github.com/iot-lab/iot-lab/wiki/Hardware-Iotlab-gateway
[3] https://www.ti.com/lit/ds/symlink/ina226.pdf
[4] https://www.iot-lab.info/docs/boards/raspberry-pi-3/

The nodes operating system is a Linux distribution built by the Yocto Project[5], an open-source project that allows the creation of embedded Linux distributions. The nodes' Linux distribution is based on "Poky" reference distribution.

B. Experiments

We define an *experiment* as an extensive uniform 100 runs of one scenario, on an edge node. We conduct the experiments on *five identical* raspberry Pi 3B edge nodes. Each experiment is *repeated 10 times* on each node.

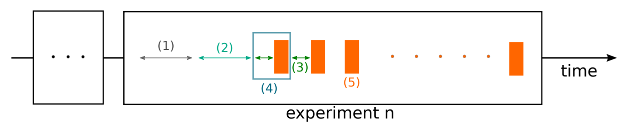


Fig. 1. Experimental Protocol: (1) node reservation and setup, (2) initial cooling, (3) cooling before stress, (4) iteration and (5) stress. An iteration (4) is made of a cooling (3) and a stress (5) period.

Experimental protocol: Figure 1 presents the experimental protocol for stressing an edge node. Before an experiment, the node is turned off. An experiment starts by reserving a node, turning it on, installing the OS, and starting power monitoring.

After setting up the node, it is idle for 3 minutes to cool down. Then, consecutive *100 stress iterations* for the scenario are run. Each iteration is *one minute* of *stress* preceded by *one minute* of idle for *cooling down*.

To start an iteration, we SSH into the node, start the iteration and SSH out. During a stress, there is no open SSH connection on the node and only the stress command from our experimental setup is running. We take *timestamps* before the stress starts and after it ends to retrieve power measurements for the stress, from timestamped power measurement logs recorded by the testbed.

We focus on quantifying variability of power and energy related metrics. Several potential sources of variability on the node are *eliminated*: (i) WiFi interface is shutdown, (ii) Bluetooth is disabled, (iii) LEDs are turned off, (iv) and CPU frequency is set to a fixed *performance* mode, as default *powersave* mode can cause sudden frequency changes that influence performance, and thus consumed energy [23].

Therefore, our experimental protocol aims at (i) reducing its impact on power and energy consumption on the node (ii) and eliminating existing variability, to reduce possible noise, for measuring variability from the stresses.

Stress scenarios:

Linux *stress-ng* version 0.11.17 is used to generate stresses on edge nodes. *stress-ng* is selected because it contains several workloads to stress specific components of a node, independently. It is popularly used in literature [18], [20]. Each stress scenario is selected to stress the node in a specific state or specific component, used in every application: idle, CPU and RAM. This is a first step towards understanding potential variability in edge applications.

[5] https://www.yoctoproject.org/

Idle: a node is in the *idle state* when it is *ON* but it is not doing any useful work [17]. To reach this state, the experimental protocol (Figure 1) has an initial cooling down period. It is followed by a *"sleep 60"* command to wait for the end of the supposed "stress" period, while the node is not doing any useful work.

CPU intensive, matrix product: *"stress-ng −−cpu 4 −−cpu-method matrixprod -t 60"* command is used to stress the CPU. 4 workers stress the 4 cores available on a node. Each worker continuously do matrix product of two 128 × 128 matrices of long doubles.

RAM intensive, flip: *"stress-ng -m 4 −−vm-method flip -t 60"* command is used to stress the RAM. 4 workers are started in this workload, in order to avoid thread migration between cores. A worker initially sets its subset of memory to a random pattern. Then, it works sequentially through memory 8 times. At each time, only one bit of a $byte$ is flipped, effectively inverting each $byte$ in 8 passes.

C. Evaluation Metrics

The metrics time, power, and energy are in seconds (s), watts (W), and joules (J), respectively.

Power: The *stress iteration average power*, noted $AvrgPow(stress)$, is computed from the power measured during the stress of an iteration (noted 5 in figure 1). The *Experiment Average Power*, noted $AvrgPow(exp)$, is the average of 100 $AvrgPow(stress)$. For each stress scenario, there are 10 $AvrgPow(exp)$.

The Percentage of Change [24] in $AvrgPow(exp)$ measurements for a scenario, noted $\%\Delta AvrgPow(exp)$, shows the percentage that the maximum delta(Δ) in $AvrgPow(exp)$, noted $\Delta AvrgPow(exp)$, represents with regards to the minimum $AvrgPow(exp)$. It is defined as:

$$\Delta AvrgPow(exp) = max(AvrgPow(exp)) - min(AvrgPow(exp)) \quad (1)$$

$$\%\Delta AvrgPow(exp) = \frac{\Delta AvrgPow(exp)}{min(AvrgPow(exp))} \times 100 \quad (2)$$

$AvrgPow(stress)$ and $AvrgPow(exp)$ are used to make results comparable with related works using average power (e.g in calibrations).

Energy: The *stress iteration energy*, noted $E(stress)$, is the energy consumed during the stress of an iteration. It is computed using the trapezoidal rule, as follows:

$$\int_0^N P(t)dt \quad (3)$$

where N is the stress duration of one iteration. P is measured power at a specific time t.

For an experiment, variability of energy consumption is presented using the following metrics: minimum (noted $min(E(exp))$), lower quartile ($Q1$), second quartile (median), upper quartile ($Q3$), maximum (noted $max(E(exp))$), and average of $E(stress)$ measurements. *Experiment Median Energy* (noted $MdnE(exp)$) is the median of 100 $E(stress)$ measurements in an experiment. Each stress scenario has 10 $MdnE(exp)$.

For a stress scenario, the *Percentage of Change in* $MdnE(exp)$ *measurements* (noted $\%\Delta MdnE(exp)$) represents the maximum Δ in $MdnE(exp)$ (noted $\Delta MdnE(exp)$) to the minimum $MdnE(exp)$. It is defined as:

$$\Delta MdnE(exp) = max(MdnE(exp)) - min(MdnE(exp)) \quad (4)$$

$$\%\Delta MdnE(exp) = \frac{\Delta MdnE(exp)}{min(MdnE(exp))} \times 100 \quad (5)$$

$equiv\Delta E_{month}$ is a rough estimation of measured $\Delta MdnE(exp)$, for a month. We make the assumption that stress duration is increased to a month (30 days, each of 24 hours, each of 60 minutes, each of 60 seconds). Thus, $equiv\Delta T_{month}$ is a rough estimation of variability, translated in idle up-time, for a month. They are defined as:

$$equiv\Delta E_{month} = \frac{\Delta MdnE(exp) \times 30 \times 24 \times 60 \times 60}{duration(stress)} \quad (6)$$

$$equiv\Delta T_{month} = \frac{equiv\Delta E_{month}}{IdlePow(exp)} \quad (7)$$

where $duration(stress)$ is stress duration in an iteration. $IdlePow(exp)$ is calibrated using $AvrgPow(exp)$ from conducted experiments, for single node and multiple homogeneous nodes analysis, separately. It can be max or min measured values, to compute $min(equiv\Delta T_{month})$ and $max(equiv\Delta T_{month})$, respectively.

Translating power and energy related metrics to time makes it easier to represent and understand variability. It helps the reader perceive how far expectations of remaining lifetime for edge nodes can be, when variability is not considered.

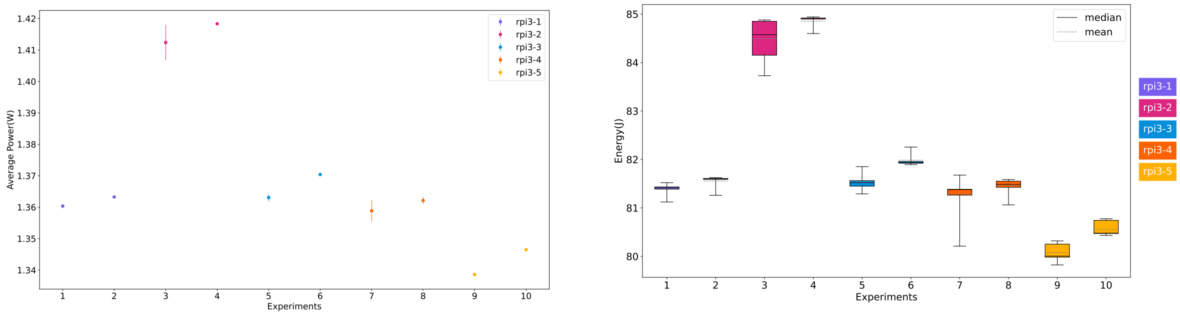
The median, $MdnE(exp)$, is chosen as in most experiment results, $E(stress)$ distributions are skewed. Median metric is therefore more representative than average [25]. When developing the metrics, we analyzed the impact of choosing median or average, for our experiments. We measured a low Δ between these two metrics. In idle scenario experiments, it is a maximum of 0.1407 J, with an equivalence of only 0.8674 h as an idle up-time, for a month.

IV. Results and Observations

For single node analysis, variability is quantified from experiments conducted on one node. For multiple homogeneous nodes analysis, for each scenario, experiments with the minimum and maximum average power are selected, per node.

A. Idle observations

Power observations: Figure 2(a) presents 10 $AvrgPow(exp)$ measurements for 5 homogeneous nodes. From each node, 2 experiments are selected out of 10 uniform



(a) Average power and standard deviation. Experiments with min and max average power, out of 10 experiments per node.

(b) Energy measurements. Experiments with min and max average power, out of 10 experiments per node. Boxes represent interquartile-range. Error bars show min and max measurements for an experiment.

Fig. 2. Idle scenario experiments on 5 homogeneous raspberry pi 3b nodes. An experiment has 100 iterations. Identical experimental protocol.

experiments: the experiments having $min(AvrgPow(exp))$ and $max((AvrgPow(exp)))$.

$AvrgPow(exp)$ from multiple similar experiments varies. Considering a single node, rpi3-2 for example, $\%\Delta AvrgPow(exp)$ is only 0.4213 %, for a $\Delta AvrgPow(exp)$ of 0.006 W. Across multiple homogeneous nodes, $\%\Delta AvrgPow(exp)$ is 5.9582 %, for a $\Delta AvrgPow(exp)$ of 0.0798 W. Furthermore, one node, rpi3-2, has higher average power measurements than the other 4 nodes.

Energy observations: Figure 2(b) presents 10 boxplots of energy consumption measurements, for 5 homogeneous nodes. Each boxplot represents 100 $E(stress)$, for the same experiments as in figure 2(a). Such box-plots permit a detailed graphical representation of variability, it summarizes 5 statistical values: $min(E(exp))$, Q1, median, Q3, and $max(E(exp))$. A box spans from Q1 to Q3 and is split by the median (Q2).

Several boxplots are not symmetric, which reflects a non normal distribution in $E(stress)$ measurements. Median is used in energy analysis, as it is more representative than the average, for central tendency, in skewed distributions [25].

On a single node, repeated thorough experiments can have no overlapping $E(stress)$ measurements (e.g rpi-3 and rpi-5). On rpi3-2, $\%\Delta MdnE(exp)$ is 0.3966%, for a $\Delta MdnE(exp)$ of 0.3355 J, with an $equiv\Delta T_{month}$ estimated between 2.0521 and 2.0685 h, only. Furthermore, boxplots show that the delta between minimum and maximum $E(stress)$, on a node, can be 3 times the measured $\Delta MdnE(exp)$ (e.g rpi3-2 or rpi3-4).

Across multiple homogeneous nodes, several nodes do not have any overlapping $E(stress)$ measurements. For example, it is the case for measurements from rpi3-2 not overlapping with any other node, and for rpi3-1 and rpi3-3 not overlapping with rpi3-5. $\%\Delta MdnE(exp)$ is 6.13 %, for a $\Delta MdnE(exp)$ of 4.9043 J. The corresponding $equiv\Delta T_{month}$ is estimated between 30.0008 and 31.1439 h, meaning that remaining lifetime predictions of one month can be off by 31.1439 h for identical idle edge nodes.

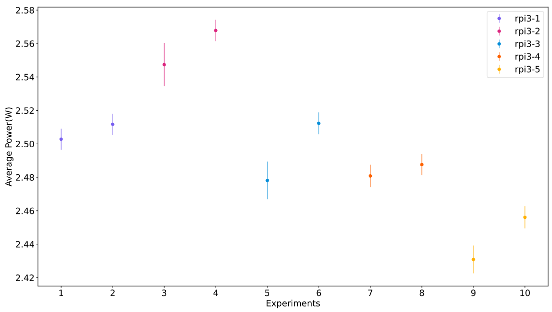
Discussion: On a single node, for idle scenario, energy measurements from repeated experiments, of 100 iterations, can be not overlapping. Conducting one experiment on a node, even if it is thorough, is not enough to represent possible variability.

Power and energy consumption variability metrics are relatively low, on a single node. In multiple papers [18], [20], idle power or energy is a baseline, deducted from total measurement to do a fine-grain analysis for workloads, without taking variability into account. Furthermore, power is multiplied by time in several works to get total energy consumption [15], multiplying the impact of non considered variability. Quantifying idle power and energy consumption variability on a single node can be important, even when it is considered low.

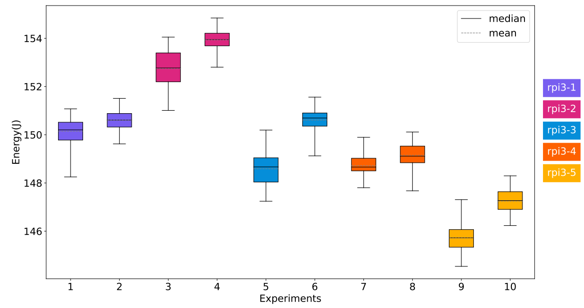
Across multiple homogeneous idle nodes, $\%\Delta MdnE(exp)$ (6.13 %) is higher than its value on a single node (0.3966 %). Similar pattern is observed with the power variability metric, $\%\Delta AvrgPow(exp)$. Conducting experiments on one node is not enough to characterize power and energy consumption for multiple edge nodes, of identical hardware. Adding nodes to the analysis could show higher variability, as energy consumption measurements for one node can be notably higher than others (e.g rpi3-2).

This high variability, across multiple homogeneous nodes, shows that power models calibrated and validated using only one or few homogeneous nodes can be surprisingly less accurate, on other identical edge nodes. The accuracy of remaining lifetime predictions can then be affected by this delta, as a node continuously uses idle power. Unexpected shorter lifetime of battery-powered edge nodes can be crucial to avoid in real deployments of edge systems, especially when nodes are not always reachable by humans. Critical examples are sensor nodes used for monitoring difficult environments for crises predictions and warnings [1], [2], [14].

We show that variability of power and energy on a single idle node exists and is relatively low. Homogeneous idle edge nodes can be operating at different power and energy values, showing a possible overall heterogeneity in energy consumption and efficiency. The insights derived from our experiments for Idle scenario are critical for analyzing and



(a) Average power and standard deviation. Experiments with min and max average power, out of 10 experiments per node.



(b) Energy measurements. Experiments with min and max average power, out of 10 experiments per node.

Fig. 3. CPU intensive "matrix product" experiments on 5 raspberry pi 3b nodes. An experiment has 100 iterations. Identical experimental protocol.

understanding energy characteristics of single and multiple homogeneous nodes, in a distributed edge system.

B. CPU intensive "matrix product" observations

Power observations: In figure 3(a), on a single node, rpi3-2 as an example, $\%\Delta AvrgPow(exp)$ is 0.8018 %, for a $\Delta AvrgPow(exp)$ of 0.0204 W. Across multiple homogeneous nodes, $min(AvrgPow(exp))$ and $max(AvrgPow(exp))$ of several nodes are not overlapping (e.g rpi3-1, rpi3-4, and rpi3-5). $\%\Delta AvrgPow(exp)$ is 5.6331 %, for a $\Delta AvrgPow(exp)$ of 0.1369 W.

Energy observations: In figure 3(b), examining variability on a single node, rpi3-2, $\%\Delta MdnE(exp)$ is 0.7684 %, for a $\Delta MdnE(exp)$ of 1.1739 J, with an $equiv\Delta T_{month}$ estimated between 7.1808 and 7.2381 h. Comparing minimum and maximum $E(stress)$ measurements, from a single node boxplots, reveals more variability than when comparing medians.

Boxplots show that several nodes do not overlap in $E(stress)$ measurements. For example, it is the case for rpi3-2 and rpi3-5 (these nodes also do not overlap for idle scenario). $E(stress)$ measurements from rpi3-2 overlap with rpi3-1 and rpi3-3 for CPU intensive scenario, when it does not for idle. Thus, variability for CPU intensive scenario can have different characteristics than idle scenario.

Across multiple homogeneous nodes, $\%\Delta MdnE(exp)$ is 5.6387%, for a $\Delta MdnE(exp)$ of 8.217 J, with an $equiv\Delta T_{month}$ estimated between 50.2648 and 52.1801 h.

Discussion: Results show that on a single node, also for this CPU intensive scenario, one experiment does not describe power and energy variability completely, even when considering 100 iterations.

On a single node, there is a low variability. On rpi3-2, $equiv\Delta T_{month}$ for $\Delta MdnE(exp)$ is 3.499 times its value for idle scenario. In literature, energy measurements for CPU workloads, from a node, can be subtracted from total energy measurements of the node, without taking variability into account [20]. For studies aiming at high accuracy and precision, especially fine-grain analysis, understanding CPU component variability, on a single node, is needed.

Across multiple homogeneous nodes, $equiv\Delta T_{month}$ for $\Delta MdnE(exp)$ in CPU intensive scenario (52.1801 h) is (i) 7.209 times its value on a single node (7.2381 h) (ii) and is 1.6754 times its value for multiple homogeneous idle nodes (31.1439 h). This demonstrates that power and energy consumption variability for CPU intensive scenarios can be important among multiple homogeneous nodes.

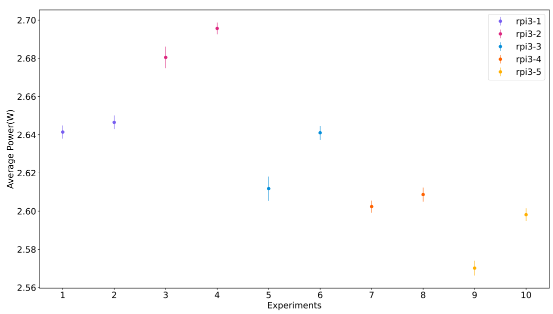
It is interesting that across multiple homogeneous nodes, $\%\Delta MdnE(exp)$ for CPU intensive scenario (5.6387 %) is lower than its value for idle scenario (6.13 %). In fact, (i) $\Delta MdnE(exp)$ for CPU intensive scenario is higher. However, (ii) $\%\Delta MdnE(exp)$ is relative to its scenario measurements ($E(stress)$ for CPU intensive scenario can be 1.9 times its value for idle scenario, using figures 3(b) and 2(b)). This demonstrates that the proposed $equiv\Delta T_{month}$ can make it easier to perceive variability, especially when it is compared among multiple scenarios.

Variability exists for the CPU intensive scenario. It can be significant across multiple homogeneous nodes. Research in edge computing with power and energy concerns is ongoing [6]. Understanding variability, on a single node and among homogeneous nodes, especially for a power hungry scenario, is necessary. It is a first step to accurately characterize energy consumption of edge nodes, and to conduct fine-grain studies.

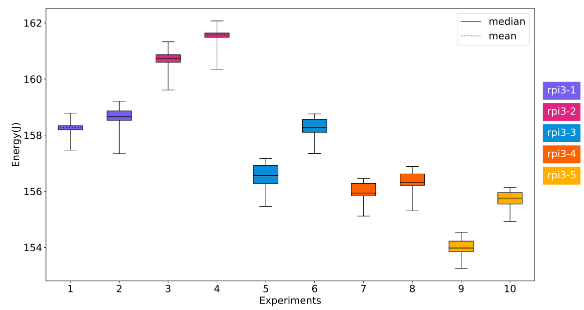
C. RAM intensive "flip" observations

Power observations: In figure 4(a), on a single node, rpi3-2, $\%\Delta AvrgPow(exp)$ is equal to 0.5659 %, for a $\Delta AvrgPow(exp)$ of 0.0152 W. Across multiple nodes, $AvrgPow(exp)$ varies, again. Where $\%\Delta AvrgPow(exp)$ is 4.8822 %, for a $\Delta AvrgPow(exp)$ of 0.1255 W.

Energy observations: In figure 4(b), on a single node, repeated identical experiments, of 100 iterations, have no overlapping $E(stress)$ measurements (e.g rpi3-3 and rpi3-5). On rpi3-2, $\%\Delta MdnE(exp)$ is 0.5591 %, for a $\Delta MdnE(exp)$ of 0.8984 J, with an $equiv\Delta T_{month}$ estimated between 5.4955 and 5.5394 h. Again, on a node, comparing minimum and maximum values of $E(stress)$ reveals higher variability, compared to using $MdnE(exp)$.



(a) Average power and standard deviation. Experiments with min and max average power, out of 10 experiments per node.



(b) Energy measurements. Experiments with min and max average power, out of 10 experiments per node.

Fig. 4. RAM intensive "Flip" experiments on 5 raspberry pi 3b nodes. An experiment has 100 iterations. Identical experimental protocol.

Across several nodes, experiments can have no overlapping $E(stress)$ measurements (e.g experiments across rpi3-1, rpi3-2 and rpi3-5). $\%\Delta MdnE(exp)$ is 4.9329 %, for a $\Delta MdnE(exp)$ of 7.5954 J, with an $equiv\Delta T_{month}$ estimated between 46.4624 h and 48.2328 h. Nodes with lowest and highest energy measurements in idle scenario (rpi3-2 and rpi3-5) keep the same order in RAM intensive scenario. Nodes 3 and 5 have overlapping $E(stress)$ measurements, but it is not the case for idle scenario. Variability for RAM intensive scenario can have different characteristics from idle scenario.

Discussion: On a single node for this RAM intensive scenario, a low power and energy variability exists, $equiv\Delta T_{month}$ for $\Delta MdnE(exp)$ (5.5394 h) is 2.6779 times its value for idle scenario. In literature, energy consumption from idle state and CPU workloads are subtracted from total measurements to get RAM energy consumption [20]. Even when the variability on a single node for RAM intensive scenario is low, it is needed for fine-grain studies, especially focusing on memory consumption.

Across multiple homogeneous nodes, the variability is important. $equiv\Delta T_{month}$ for $\Delta MdnE(exp)$ (48.2328 h) is (i) 8.7072 times its value on a single node (5.5394 h), and (ii) 1.5487 times its value across multiple idle nodes (31.1439 h). Thus, statistical measurements from one node, average power and standard deviation, commonly used in related work [12], [13], are not enough to represent existing variability. Memory intensive applications are being deployed at the edge [26]. Acknowledging and understanding variability of power and energy for RAM intensive scenarios is necessary.

V. Discussions, Implications and Usage

Edge nodes can have a limited energy budget. Quantifying the variability of power and energy consumption among homogeneous edge nodes can be a leverage to increase energy efficiency and availability in edge systems. For example, when provisioning system deployments and nodes selection, a critical node, that needs to be highly available, can be chosen, among identical hardware, taking into account its energy consumption characteristics. Another example is to consider variability in energy-aware scheduling algorithms for the edge (for clusters, up to 17 % power savings are achieved using this approach [27]). Quantifying power and energy variability can also be used to (i) increase accuracy and precision of power models and simulators, (ii) have a fine-grain analysis of energy consumption, and (iii) predict or analyze variability of complex applications. This detailed and thorough study of power and energy variability for idle, CPU and RAM intensive scenarios is needed.

The main strength of this study is that it reveals fine-grain power and energy variability, by thorough empirical experiments. Another strength is the use of ultra precise external monitoring, in addition to detailed analysis. The main weakness is the number of nodes, due to available resources. However, this means that using more nodes could reveal higher variability. Experiments for studying other node components and nodes are being designed and will be conducted in the near future.

VI. Conclusion

IoT and edge systems are adopted widely in various sciences and domains, where energy can be a key constraint, affecting availability and scalability. Energy can play a major role in the choice of edge technologies, architectures, and resources.

Many related works focus on measuring, predicting and optimizing energy consumption at the edge. However, variability of power and energy consumption for basic and fine-grain scenarios is not thoroughly investigated. Worst, several works do not account for potential power and energy variability.

We investigate, on a single and across multiple homogeneous edge nodes, the variability of power and energy consumption, under various scenarios (i.e Idle, CPU, and RAM intensive scenarios). We control experiments to limit potential variability that can exist (e.g by fixing CPU frequency, disabling WiFi, Bluetooth, LEDs) to extract variability only from our scenarios.

We quantify variability by repeating thorough experiments for the same scenario, while retrieving power and energy measurements, using ultra precise external monitoring. The study is done for Raspberry Pi 3B edge nodes, highly adopted in literature, for prototyping and real deployments.

Results show that power and energy measurements of a scenario can be different, from one node to another. On a single node, for idle, CPU and RAM intensive scenarios, the variability in energy consumption can be up to an equivalence of 2.0685 h, 7.2381 h and 5.5394 h of idle up-time, in a month, respectively. On multiple homogeneous nodes, the variability for these scenarios can be up to an equivalence of 31.1439 h, 52.1801 h and 48.2328 h of idle up-time, in a month, respectively.

Power and energy variability, when not considered, can impact lifetime predictions or lead to misleading energy consumption conclusions. This is especially true for IoT and edge systems, where energy can be limited. Understanding power and energy consumption variability is crucial.

For future works, we plan to assess the variability of other components on edge nodes (e.g I/O and network interfaces). We also plan on evaluating the variability of other types of nodes. Finally, we plan to assess the variability for multiple components stressed together, representing an application.

References

[1] I. Raïs, J. M. Bjørndalen, P. Hoai Ha, K.-A. Jensen, L. S. Michalik, H. Mjøen, M. Tveito, and O. Anshus, "Uavs as a leverage to provide energy and network for cyber-physical observation units on the arctic tundra," in *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 2019, pp. 625–632.

[2] P. Giri, K. Ng, and W. Phillips, "Wireless sensor network system for landslide monitoring and warning," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 4, pp. 1210–1220, 2018.

[3] C. Chen, A. Knoll, H.-E. Wichmann, A. Horsch *et al.*, "A review of three-layer wireless body sensor network systems in healthcare for continuous monitoring," *Journal of Modern Internet of Things*, vol. 2, no. 3, pp. 24–34, 2013.

[4] B. Lashkari, J. Rezazadeh, R. Farahbakhsh, and K. Sandrasegaran, "Crowdsourcing and sensing for indoor localization in iot: A review," *IEEE Sensors Journal*, vol. 19, no. 7, pp. 2408–2434, 2019.

[5] C. MacGillivray and D. Reinsel, "Worldwide global datasphere iot device and data forecast, 2019–2023," *IDC Market Forecast-Doc US45066919*, 2019.

[6] C. Jiang, T. Fan, H. Gao, W. Shi, L. Liu, C. Cérin, and J. Wan, "Energy aware edge computing: A survey," *Computer Communications*, vol. 151, pp. 556–580, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S014036641930831X

[7] C. Adjih, E. Baccelli, E. Fleury, G. Harter, N. Mitton, T. Noel, R. Pissard-Gibollet, F. Saint-Marcel, G. Schreiner, J. Vandaele *et al.*, "Fit iot-lab: A large scale open experimental iot testbed," in *2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*. IEEE, 2015, pp. 459–464.

[8] M. Saari, A. M. bin Baharudin, and S. Hyrynsalmi, "Survey of prototyping solutions utilizing raspberry pi," in *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2017, pp. 991–994.

[9] C. N. Cabaccan, F. Reidj, and G. Cruz, "Power characterization of raspberry pi agricultural sensor nodes using arduino based voltmeter," in *2018 3rd International Conference on Computer and Communication Systems (ICCCS)*. IEEE, 2018, pp. 349–352.

[10] I. Raïs, L. Guegan, and O. Anshus, "Impact of loosely coupled data dissemination policies for resource challenged environments," in *2022 22nd IEEE International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*, 2022, pp. 524–533.

[11] A. Dâmaso, D. Freitas, N. Rosa, B. Silva, and P. Maciel, "Evaluating the power consumption of wireless sensor network applications using models," *Sensors*, vol. 13, no. 3, pp. 3473–3500, 2013. [Online]. Available: https://www.mdpi.com/1424-8220/13/3/3473

[12] I. Raïs, O. Anshus, J. M. Bjørndalen, D. Balouek-Thomert, and M. Parashar, "Trading data size and cnn confidence score for energy efficient cps node communications," in *2020 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID)*, 2020, pp. 469–478.

[13] K. Shinghal, A. Noor, N. Srivastava, and R. Singh, "Power measurements of wireless sensor networks node," *Int. J. Comput. Eng. Sci.(IJCES)*, vol. 1, no. 1, pp. 8–13, 2011.

[14] M. J. Murphy, M. Tveito, E. F. Kleiven, I. Raïs, E. M. Soininen, J. M. Bjørndalen, and O. Anshus, "Experiences building and deploying wireless sensor nodes for the arctic tundra," in *2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*, 2021, pp. 376–385.

[15] E. Ahvar, A.-C. Orgerie, and A. Lebre, "Estimating energy consumption of cloud, fog, and edge computing infrastructures," *IEEE Transactions on Sustainable Computing*, vol. 7, no. 2, pp. 277–288, 2022.

[16] M. El Mehdi Diouri, O. Glück, L. Lefèvre, and J.-C. Mignot, "Your cluster is not power homogeneous: Take care when designing green schedulers!" in *2013 International Green Computing Conference Proceedings*, 2013, pp. 1–10.

[17] A. Benoit, L. Lefevre, A.-C. Orgerie, and I. Raïs, "Reducing the energy consumption of large-scale computing systems through combined shutdown policies with multiple constraints," *The International Journal of High Performance Computing Applications*, vol. 32, no. 1, pp. 176–188, 2018.

[18] M. E. M. Diouri, M. F. Dolz, O. Glück, L. Lefèvre, P. Alonso, S. Catalán, R. Mayo, and E. S. Quintana-Ortí, "Assessing power monitoring approaches for energy and power analysis of computers," *Sustainable Computing: Informatics and Systems*, vol. 4, no. 2, pp. 68–82, 2014, special Issue on Selected papers from EE-LSDS2013 Conference. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2210537914000171

[19] K. Kadota, I. Taniguchi, and H. Tomiyama, "Measurement of performance and energy consumption of opencv programs on raspberry pi," *Bulletin of Networking, Computing, Systems, and Software*, vol. 9, no. 1, pp. 35–39, 2020.

[20] M. S. Aslanpour, A. N. Toosi, R. Gaire, and M. A. Cheema, "Wattedge: A holistic approach for empirical energy measurements in edge computing," in *International Conference on Service-Oriented Computing*. Springer, 2021, pp. 531–547.

[21] H. A. Abdelhafez, H. Halawa, A. Almoallim, A. Ahmadi, K. Pattabiraman, and M. Ripeanu, "Characterizing variability in heterogeneous edge systems: A methodology & case study," in *2022 IEEE/ACM 7th Symposium on Edge Computing (SEC)*, 2022, pp. 107–121.

[22] J. Muñoz, F. Rincon, T. Chang, X. Vilajosana, B. Vermeulen, T. Walcarius, W. van de Meerssche, and T. Watteyne, "Opentestbed: Poor man's iot testbed," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2019, pp. 467–471.

[23] H. Knoche and H. Eichelberger, "Using the raspberry pi and docker for replicable performance experiments," in *Proceedings of the 2018 ACM/SPEC International Conference on Performance Engineering. ACM. https://doi. org/10*, vol. 1145, no. 3184407.3184431, 2018.

[24] J. Leverich, M. Monchiero, V. Talwar, P. Ranganathan, and C. Kozyrakis, "Power management of datacenter workloads using per-core power gating," *IEEE Computer Architecture Letters*, vol. 8, no. 2, pp. 48–51, 2009.

[25] G. A. Rousselet and R. R. Wilcox, "Reaction times and other skewed distributions: problems with the mean and the median," *BioRxiv*, p. 383935, 2018.

[26] Z. Zou, Y. Jin, P. Nevalainen, Y. Huan, J. Heikkonen, and T. Westerlund, "Edge and fog computing enabled ai for iot-an overview," in *2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, 2019, pp. 51–56.

[27] M. E. M. Diouri, O. Glück, L. Lefevre, and J.-C. Mignot, "Your cluster is not power homogeneous: Take care when designing green schedulers!" in *2013 International Green Computing Conference Proceedings*. IEEE, 2013, pp. 1–10.