Delivering Reliability of Data Sources in IoT Healthcare Ecosystems

Argyro Mavrogiorgou, Athanasios Kiourtis, Dimosthenis Kyriazis
University of Piraeus
Piraeus, Greece
{margy, kiourtis, dimos}@unipi.gr

***Abstract*—Nowadays, the use of Internet of Things (IoT) in various different fields has made significant progress, especially in the field of healthcare, where a myriad of heterogeneous medical data sources are in use. This fact has reinforced the vision of developing new communication technologies and finding new ways to synchronize and successfully manage all these data sources. However, this vision is accompanied by several related challenges. One of these challenges refers to the fact that since all the existing IoT medical data sources are usually characterized by a high degree of heterogeneity, they are expected to be recognized as reliable at different stages, thus providing data of different levels of reliability. To effectively tackle this challenge, the present paper proposes a mechanism for capturing the reliability levels of different IoT medical data sources, so as to automatically decide whether these will be considered as reliable or not, and thus their data will be kept for further analysis. In this context, in this mechanism three (3) discrete stages are implemented, facilitating both the data reliability and the availability estimation of these data sources, making finally feasible the manipulation of these sources and the estimation of their overall reliability levels. The prototype associated with this paper provides an example of this mechanism, demonstrating in detail each discrete stage.**

I. INTRODUCTION

It is an undeniable fact that healthcare is one of the major areas of application of Internet of Things (IoT). In particular, in recent years technology has focused heavily on how medical devices and health monitoring devices, clinical laptops and remote controls can contribute to better patient health and more efficient healthcare that, in turn, can lead to better medical care systems [1]. Currently, healthcare is one of the fastest areas that adopt IoT technologies, offering better personalized services, reducing operating costs and improving patient care and quality of life [2]. For that reason, nowadays there is a large expansion of the IoT medical market, resulting in a multitude of heterogeneous devices connected to the health world. However, these devices are typically characterized by a high degree of heterogeneity [3], producing large amounts of heterogeneous health and fitness data [4]. Hundreds of healthcare organizations are daily dealing with challenges in extracting data from various types of medical devices, affecting both patient care and medical research [5].

However, all of these healthcare organizations face many difficulties in successfully managing all this data, since this data may not only be heterogeneous but also have different levels of reliability [6]. Even if all of the data becomes interoperable, not all of it should be retained for reuse, as it is extremely critical since it leads medical decisions making [7]. Rather, it would be more prudent and effective to take into account the different levels of reliability that such data may have, thus analyzing only those data that have high levels of reliability. Therefore, the challenge lies in the difficulty of determining the reliability of these large amounts of data. However, appraisal of the reliability of the underlying medical data sources, as well as of their data output, are treated mainly as black boxes in the healthcare sector, and little attention is given to their reliability when integrated into larger systems. The use of these data sources without proper reliability assessments may have serious health implications, while the absence of their reliability could reduce the degree of successful interpretation and significance of the results and findings that are produced based on their output data [8].

There is no doubt that reliability has gained extremely high popularity. Perhaps this is due to medical errors that have immediate and possibly deadly effects on human affairs. We all know the stories of patients who received the wrong drug or the right medicine at the wrong dose because the wrong disease was diagnosed by medical staff with inadequate training in administering a specific test. Therefore, improving the reliability of medical systems is probably much more urgent than improving, for example, the quality of a video game. Patientcare, for example, in the field of nursing, is another extremely sensitive area where reliability has found a fertile ground [9]. Therefore, the reliability of technological systems is very important for medical engineering. Failure of medical systems and data sources may result in adverse effects that could be an injury or death of a patient, and may have serious legal consequences. This results in the challenge of using reliable medical devices and data sources or equipment that must have a high level of reliability. The intensity of failure increases with the age of medical equipment, thus requiring technological repair and monitoring of equipment. It is reported that up to 80,0% of medical equipment currently used in public health organizations is worn out or is outdated, making it difficult to guarantee not only the reliability and effectiveness but also the safety of medical equipment [10]. In the same notion, in another study [11] it is pointed out that data from various sources indicate that some hospitalized patients suffer from treatment-induced injuries, most of which are due to system/device failures.

To address all the aforementioned challenges, in this paper a solution is proposed for capturing the reliability of the heterogeneous IoT medical data sources that exist, of both known and unknown nature (i.e. data source type). To this end, a mechanism is proposed for assessing the reliability of these data sources, by facilitating the automatic estimation of both the data sources and their produced data reliability levels. More particularly, the mechanism by using as an input both the availability of the connected data sources and the reliability of their produced data, it calculates the reliability of each corresponding data source. Thus, it concludes whether the latter will be qualified as reliable or not, and thus its data will be obtained for further utilization and analysis. This mechanism has been evaluated through three (3) different experiments in order to assess its effectiveness for capturing the reliability of IoT medical data sources.

The rest of this paper is organized as follows. Section II describes the study of the related work regarding reliability and the existing reliability researches that have been implemented in the healthcare sector. Section III describes the proposed mechanism for capturing the reliability of heterogeneous IoT medical data sources, whereas in Section IV, it is described a representative use case of the proposed mechanism. Finally, Section V is analyzing our conclusions and plans.

II. RELATED WORK

With the growing population and aging society in several countries, healthcare providers aim to enhance the quality of healthcare services, while balancing risk mitigation and service costs. As a result, various new information technologies and innovative communication methodologies have evolved to improve the healthcare sector. These technologies increase the quality of services, thus helping to reduce the cost of the healthcare systems and increase the quality of healthcare services. To measure and evaluate the reliability of systems, different reliability metrics exist, which are important for monitoring reliability growth, performing risk analysis, and decreasing warranty costs. In this context, reliability metrics assess the degree to which a software product consistently performs its intended function without failure (i.e. it assesses the probability of software failure or the rate at which software errors will occur). Therefore, reliability metrics are important for estimating reliability, as they provide quantitative indicators for reliability management, evaluation and validation, trade-off among cost, schedule, monitoring testing process, and interpretation of reliability behavior. In order to estimate reliability, the corresponding metrics may derive either from the failure occurrence expressions (i.e. software reliability) or from the derived data (i.e. data reliability).

In this context, various methods of reliability have been proposed in the literature on the world of IoT, and more particularly in the healthcare sector. All these methods are attempting to measure the reliability of IoT medical data sources used for various health purposes, using either software reliability or data reliability metrics. In more detail, the authors in [12] discussed the reliability of Fitbit devices, assessing the reliability between Fitbit Flex devices and two (2) different other similar devices, recording their activity based on sitting time measurement, and the time spent at different intensities of activity, against a validated triaxial accelerometer. In the same context, the authors in [13] examined the reliability of ten (10) activity trackers for measuring steps in laboratory and free-living conditions, by estimating the Intra-Class Correlation (ICC). In [14] the authors evaluated the reliability of two (2) criteria on two (2) different foot stiffness devices in different test approaches by measuring the Coefficient of Variation (CV), the ICC, as well and the Standard Error of Measurement (SEM) of these devices. In addition, the authors in [15] evaluated the reliability of a medical mobile application using a gravity strength test, recording the corresponding ICC and SEM of its produced data. In the same context, the authors in [16] recorded the ICC measurement to evaluate the intrinsic and interactive reliability of a device used by patients with diabetes, calculating both the ICC and the SEM of the data generated by this device. What is more, the purpose of the study in [17] was to determine the reliability of automated devices that measure systolic blood pressure of the foot and arm index of a patient, while the purpose of the study in [18] was to determine the internal reliability of a mobile device goniometer in the measurement of lumbar flexion. Furthermore, the goal of the study in [19] was to assess the validity and reliability of commonly used temperature devices compared with rectal temperature in individuals exercising in a controlled, high environmental temperature indoor setting, and then resting in a cool environment, by estimating the ICC, the SEM, and the CV. In addition to the above, in [20] the authors presented a simple method of decomposition that could be easily applied on complex medical systems, through which the effect of the subsystems or components on the reliability of the overall system could be easily calculated, estimating the metrics of Mean Time To Failure (MTTF), Mean Time To Repair (MTTR), and Availability. In addition, the authors in [21] analyzed a database of failures of many types of medical equipment, so as to study the dependence of failure rate on equipment age and on time since repair, whereas in the same notion, the authors in [22] presented various criteria and methods for evaluating the reliability of medical equipment. What is more, the authors in [23] presented a reliability analysis of a standby complex system in a dairy plant. Finally, the authors in [24] analyzed the field data for medical imaging systems during the warranty period, while the authors in [25] presented the results of the early reliability prediction for Philips medical systems based on field data.

All the aforementioned approaches have implemented several features regarding the reliability estimation among heterogeneous IoT medical data sources. However, all these approaches lack of sufficient flexibility and adaptability to solve challenges arisen from dynamically gathering data from both known and unknown devices and automatically estimating their reliability levels. Apart from this, none of the existing approaches measures the reliability of the underlying data sources based upon both their software and their data reliability, an innovation that takes place in the proposed approach. For that reason, in our approach an innovative mechanism is proposed for automatically capturing the reliability levels of both known and unknown data sources, and finally decide whether they will be considered as reliable or not, and thus their data will be kept for further analysis.

III. PROPOSED APPROACH

In our approach, an innovative mechanism is proposed for capturing the reliability of heterogeneous IoT medical data sources of both known and unknown nature (i.e. data source type), in order to finally collect data only from the reliable ones. In more detail, based on the proposed mechanism, the reliability estimation of the connected data sources in combination with their data takes place. This process is of major importance, as it is not sufficient to keep all the derived data and use it for further analysis, as many of it may have derived either from unreliable data sources, or from reliable data sources that are faulty and error prone. For that reason, it is necessary to measure and evaluate the reliability of all the produced data, so as to finally keep only the reliable data that comes from only reliable data sources. In order to achieve that, it is more effective to estimate both the data sources' reliability themselves, and the reliability of their produced data. For that purpose, the mechanism implements three (3) discrete stages: (i) Data Sources Availability, (ii) Data Reliability, and (iii) Overall Reliability, as depicted in Fig. 1. In short, in the first stage the calculation of the data sources' availability occurs, followed by the second stage, where the calculation of the reliability of the produced data of the corresponding data sources takes place. Finally, in the third stage the combination of the results of the two (2) aforementioned stages occurs, so as to calculate the overall reliability levels of each connected data source based upon both its availability levels and the reliability levels of its produced data. Consequently, the final result is made available, representing whether each connected data source is considered as reliable or not and its data will be kept for further analysis.

It should be noted that the proposed mechanism requires as an input the connection of the available IoT medical data sources, as well as the available data sources' produced data, so as to furtherly use it for calculating the corresponding reliability levels. In order to achieve that, the current mechanism exploits the approach proposed in [26]. Since the current mechanism requires an input from the mechanism proposed in [26], it has to take into consideration the same requirements. For that reason, it must consider that the available IoT medical data sources must be Bluetooth-enabled, whereas they must always contain open Application Programming Interfaces (APIs), so as to be able to be connected to the mechanism and offer their data.

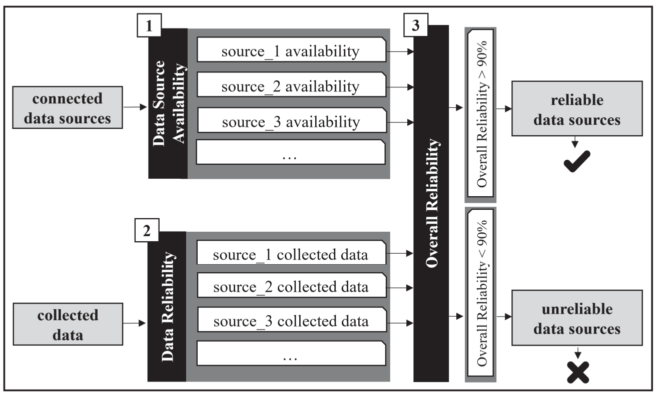

Fig. 1. Architecture of the mechanism

A. Data Sources Availability

In the first stage, the availability of the connected data sources is evaluated, where the mechanism calculates the reliability levels of the connected data sources. Although the research in [27] outlines that there exists a wide range of metrics for capturing the data sources' reliability, this mechanism, in order to calculate each different connected data source's reliability, measures only the metric of the availability (or mission capable rate) of them, as it is the most representative metric [28]. To this context, a wide range of availability classifications and definitions exist [29], however in this mechanism the most suitable one that is going to be measured is the operational availability. Operational availability (i.e. Availability) represents the ratio of the system uptime to total time, given mathematically by the equation (1), where the Operating_Cycle is the overall time period of operation being investigated, and Uptime is the total time in which the system was functioning during the specific Operating_Cycle.

$$Availability = Uptime \,/\, (Operating\_Cycle) \qquad (1)$$

Therefore, the mechanism calculates the availability of each connected data source, calculating the corresponding values, setting a timer to record how often each data source communicates with the mechanism and provides its data.

B. Data Reliability

However, as mentioned above, it is not sufficient enough to measure only the data sources' availability for deciding whether the latter is considered as reliable or not, but it is more effective to measure also the reliability of these data sources' data. For this reason, the mechanism applies the second stage, where it records the reliability measurements [30] of each different collected dataset. More specifically, among the different types that exist for measuring data reliability [31], in the current mechanism the Test-Retest Reliability (TRR) is used, since it is the most suitable one. Therefore, based on the basic features of TRR [31], the measurements are taken by a single person on the same item (i.e. type of data source), under the same conditions, and in a short period, evaluating the reliability across this period. In order to calculate the TRR of the connected data sources' data, the SPSS library [32] is used, calculating the corresponding ICC measurement, since the data may contain either interval or ratio data [33]. More specifically, the method of two-way random effects, absolute agreement, and single rater/measurement (i.e. ICC (2,1)) is implemented, obeying the corresponding conditions [31].

C. Overall Reliability

As a result of all the above, in the final stage, the overall reliability of the collected data is calculated. Thus, as soon as the ICC of each different dataset is calculated, its results are combined with the results of the availability (i.e. Availability) that derived upon the corresponding data source, so as to finally decide whether each data source, and as a result its derived data, are considered as of high quality or not. To this end, it should be noted that in order to consider the final results (i.e. Overall_Reliability) as trustful and reliable, these must exceed the set threshold of 90,0%. In more detail, Overall Reliability is calculated mathematically by equation (2), where it equals with the sum of the data source's Availability that is multiplied with a weight of 0.5, and the corresponding ICC of the data of this data source that is multiplied with a weight of 0.5. With regards to the set weights, these were chosen based upon the research results that were acquired during relevant experiments that were performed in the past. These results revealed that both the Availability and the ICC should have the same weights, since they were considered to be the same characteristic and decisive for the calculation of the Overall_Reliability results.

$$Overall\_Reliability = (Availability * 0.5) + (ICC * 0.5) \quad (2)$$

Based on the calculated results of the Overall_Reliability, all the data whose reliability exceeds the predefined threshold are retained in the mechanism for further exploitation and use. On the contrary, all the data whose reliability does not exceed the predefined threshold are rejected by the mechanism.

IV. EXPERIMENTAL RESULTS

A. Evaluation Environment

The proposed mechanism was developed in Java SE using the NetBeans IDE v8.0.2 [34], and used a processing environment with 16GB RAM, Intel i7-4790 @ 3.60 GHz x 8 CPU Cores, 2TB Storage, and Windows 10 operating system. Concerning the results of the mechanism, these are depicted below, following the three (3) stages explained in Section III. What is more, in order to store all the data that is produced by the mechanism, the latter used the Derby Database [35].

B. Evaluation Dataset

In order to perform a complete testing and evaluation of the proposed mechanism, three (3) IoT medical data sources were chosen, being able to communicate through Bluetooth with the mechanism, and offering open APIs for accessing their data. In deeper detail, two (2) of these data sources were activity trackers, whereas one (1) of these was a body weight scale, as they are depicted in Table I.

TABLE I. EVALUATION DATASET

#	Name	Vendor	Type
1	Fitbit Aria	Fitbit	Body weight scale
2	Misfit Path	Misfit	Activity tracker
3	Polar A370	Polar	Activity tracker

C. Evaluation Results

1) With regards to Fitbit Aria, using the mechanism from [26], once the device was connected, the mechanism found out that the connected device already existed in its private registry, and thus it was characterized as a device of a known type (i.e. known device). As a result, the overall process of evaluating its reliability levels was bypassed, since as a known device, it was already known that it is a quite reliable data source.

2) With regards to Misfit Path, using the mechanism from [26], once the device was connected, the mechanism found out that the connected device did not exist in the mechanism's private registry. Consequently, the coupled device was characterized as an unknown device. Using the same mechanism, the current mechanism retrieved the data from the connected device, aiming at estimating its reliability levels. As soon as all the data of the connected device was successfully collected, the overall reliability of the collected data of this device was calculated. Thus, its availability measurements were combined with the reliability measurements of its collected data, applying the first and the second stages of the mechanism that were described in Section III. More specifically, in terms of availability measurements, the mechanism recorded the device's operating time through the frequency of data transmission to the device over a total period of 30 days. Thus, during the 30 days of the experiment, the mechanism recorded the availability of the connected device, using equation (1) of Section III, assuming that the device is fully available (i.e. 100,0% availability) when it sends 288 measurements per 24 hours. This assumption was based on the fact that a device that sends a single measurement per minute, for a whole day (i.e. 60 measurements per hour, for 24 hours) is considered as of high-reliability and availability, according to the mean of the overall measurements that were provided during the aforementioned operation period of 30 days, in combination with the derived results of other relevant experiments that were made in the past.

However, as mentioned in Section III, along with the availability of the device, the reliability of the collected data of the device was also recorded. For that purpose, the ICC of this data was calculated by using the SPSS library, in order to calculate the final degree of reliability of the connected device, and as a result the degree of reliability of its generated data. When this procedure got complete, the mechanism produced the results of Table II, which depicts the results that were collected after performing the same experiment for 30 days. In more detail, Table II summarizes the mechanism's results including: (i) the data frequency of the connected device's transmitted data, in terms of how many measurements were gathered per day, (ii) the percentage of the data availability of this device (i.e. Availability), considering the data availability of the aforementioned fully available device, (iii) the percentage of the data reliability (i.e. ICC) that resulted from the collected data, and (iv) the percentage of the overall reliability calculated based on the results of the device's availability and its data reliability, applying the equation (2) of Section III.

As stated in Section III, in order to consider the final results (i.e. Overall Reliability) of the connected device's data as trustful and reliable, and thus keep it for further utilization, this must exceed the set threshold of 90,0%. Thus, based upon the results of Table II, it can be observed that all the data that was collected everyday was of high levels of reliability, as all of it exceeded the set threshold. As a result, the Misfit Path was considered as a device of high levels of reliability, and its data could be kept in the mechanism, in order to be exploited and used from the corresponding platforms that will use the mechanism.

Fig. 2, Fig. 3, and Fig. 4 visualize the results of Table II, depicting the percentages of overall availability, overall data reliability, and overall reliability, correspondingly, for each one of the 30 days of the experiment upon the Misfit Path device.

TABLE II. MISFIT PATH RELIABILITY MEASUREMENTS

Day	Data Frequency	Data Source Availability (%)	Data Reliability (%)	Overall Reliability (%)
1st	280	97,2	95,2	96,2
2nd	281	97,5	93,1	95,3
3rd	280	97,2	96,0	96,6
4th	276	95,8	94,2	95,0
5th	280	97,2	96,3	96,7
6th	276	95,8	89,2	97,0
7th	279	96,8	85,6	91,2
8th	280	97,2	96,3	96,9
9th	281	97,5	98,1	97,8
10th	280	97,2	87,2	92,2
11th	279	96,8	88,5	92,6
12th	280	97,2	95,7	96,4
13th	280	97,2	96,3	96,7
14th	278	96,5	90,5	93,5
15th	281	97,5	98,1	97,8
16th	280	97,2	87,2	92,2
17th	280	97,2	96,3	96,7
18th	280	97,2	95,7	96,4
19th	279	96,8	94,0	95,4
20th	281	97,5	98,1	97,8
21st	280	97,2	95,7	96,4
22nd	280	97,2	96,3	96,7
23rd	280	97,2	95,4	96,3
24th	275	95,4	95,1	95,2
25th	280	97,2	87,2	92,2
26th	279	96,8	89,5	93,1
27th	279	96,8	90,2	93,5
28th	280	97,2	87,2	92,2
29th	281	97,5	93,9	95,7
30th	280	97,2	95,5	96,3

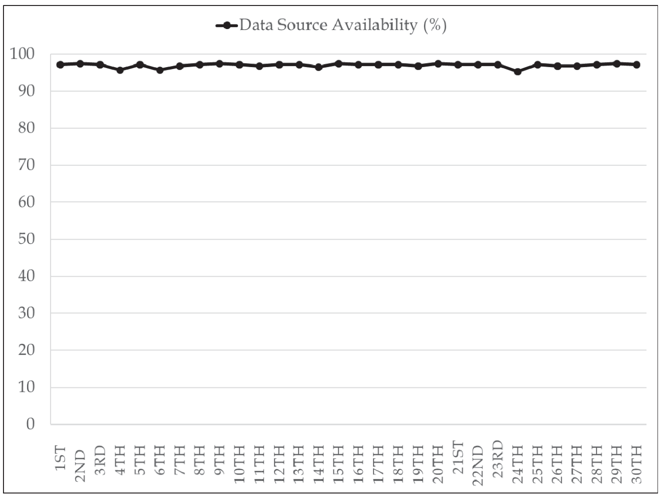

Fig. 2. Overall Misfit Path availability results

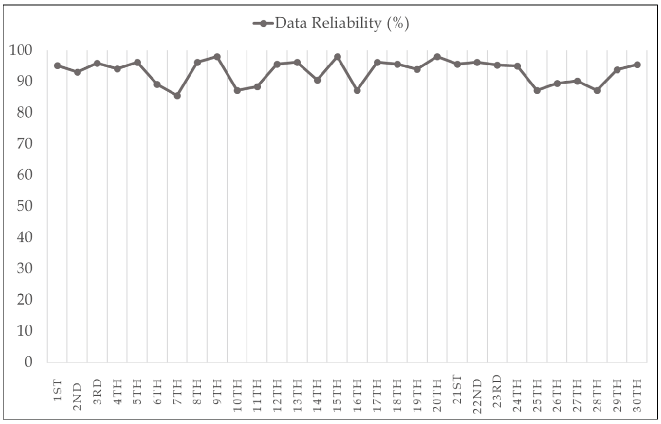

Fig. 3. Overall Misfit Path data reliability results

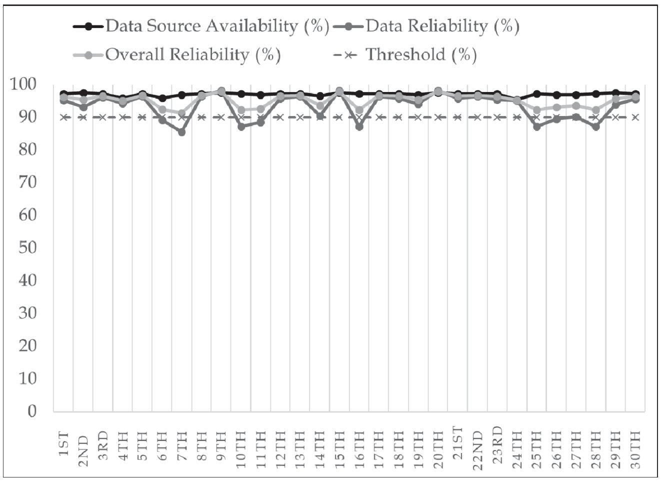

Fig. 4. Overall Misfit Path overall reliability results

3) With regards to Polar A370, using the mechanism from [26], once the device was connected, the mechanism found out that the connected device did not exist in the mechanism's private registry. Consequently, the coupled device was characterized as an unknown device. Following the same procedure as in the previous example, as soon as all the data of the connected device was successfully collected, the overall reliability of the collected data of this device was calculated, combining its availability measurements with the reliability measurements of its collected data. In more detail, in terms of availability measurements, the mechanism recorded the device's operating time through the frequency of data transmission to the device over a total period of 30 days. Therefore, during the 30 days of the experiment the mechanism recorded the availability of the connected device, using equation (1) of Section III, assuming as in the previous example, that the device is fully available (i.e. 100,0% availability) when it sends 288 measurements per 24 hours.

In sequel, along with the availability of the device, the reliability of the collected data of the device was also recorded. For that purpose, the ICC of this data was calculated by using the SPSS library, in order to calculate the final degree of reliability of the connected device, and as a result the degree of reliability of its generated data. When this procedure got complete, the mechanism produced the results of Table III that depicts the results that were collected, after performing the same experiment for 30 days. In more detail, Table III summarizes the mechanism's results including: (i) the data frequency of the connected recognized device's transmitted data, in terms of how many measurements were gathered per day, (ii) the percentage of the data availability of this device (i.e. Availability), considering the data availability of the aforementioned fully available device, (iii) the percentage of the data reliability (i.e. ICC) that resulted from the collected data, and (iv) the percentage of the overall reliability calculated based on the results of the device's availability and its data reliability, applying the equation (2) of Section III.

TABLE III. POLAR A370 RELIABILITY MEASUREMENTS

Day	Data Frequency	Data Source Availability (%)	Data Reliability (%)	Overall Reliability (%)
1st	230	79,8	93,8	86,8
2nd	240	83,3	92,6	87,9
3rd	242	84,0	91,2	87,6
4th	240	83,3	92,2	87,7
5th	241	83,6	89,2	86,4
6th	238	82,6	91,5	87,0
7th	231	80,2	90,0	85,1
8th	230	79,8	92,6	86,2
9th	240	83,3	90,2	86,7
10th	232	80,5	91,1	85,8
11th	230	79,8	90,3	85,0
12th	240	83,3	90,2	86,7
13th	242	84,0	91,2	87,6
14th	242	84,0	90,2	87,1
15th	242	84,0	92,3	88,1
16th	230	79,8	93,5	86,6
17th	240	83,3	90,7	87,0
18th	241	83,6	91,6	87,6
19th	240	83,3	92,6	87,9
20th	231	80,2	95,5	87,8
21st	242	84,0	92,6	88,3
22nd	230	79,8	93,8	86,8
23rd	230	79,8	92,3	86,0
24th	245	85,0	93,7	89,3
25th	240	83,3	91,4	87,3
26th	231	80,2	90,2	85,2
27th	231	80,2	89,5	84,8
28th	230	79,8	92,4	86,1
29th	231	80,2	93,3	86,7
30th	240	83,3	93,7	88,5

As stated in Section III, in order to consider the final results (i.e. Overall Reliability) of the connected device's data as trustful and reliable, and thus keep it for further utilization, these must exceed the set threshold of 90,0%. However, based upon the results of Table III, it can be observed that most of the data that was collected everyday did not exceed this threshold for a little percentage, whereas there were some days that the device's reliability was satisfying the set threshold. As a result, even if the threshold was not reached due to a small degree of difference, the Polar A370 would not be considered as a device of high levels of reliability, and its data would not be kept in the mechanism. Therefore, this device's data were totally erased by the mechanism.

Fig. 5, Fig. 6, and Fig. 7 visualize the results of Table III, depicting the percentages of overall availability, overall data reliability, and overall reliability, correspondingly, for each one of the 30 days of the experiment upon the Polar A370 device.

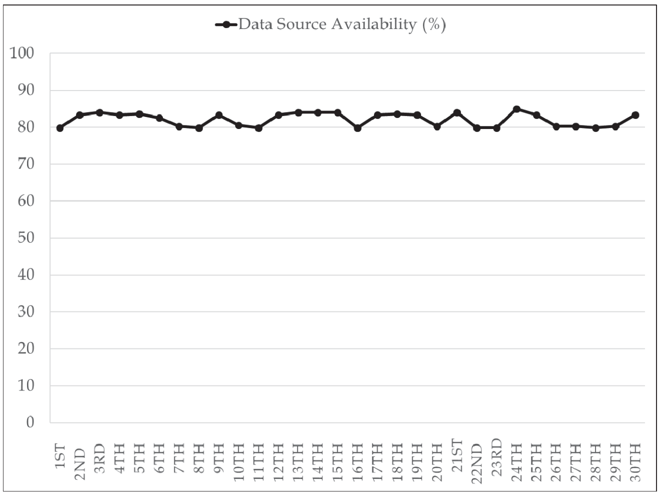

Fig. 5. Overall Polar A370 availability results

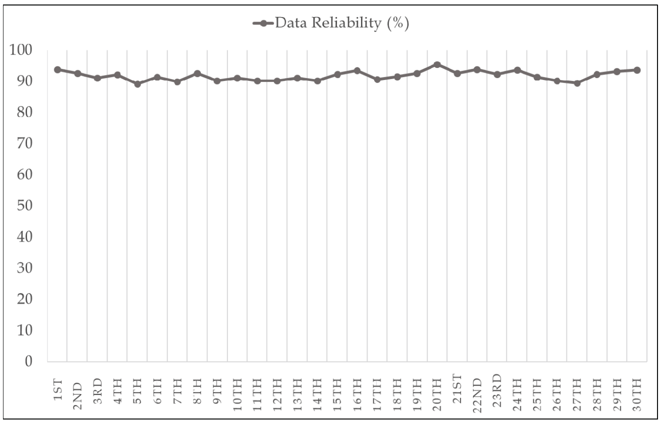

Fig. 6. Overall Polar A370 data reliability results

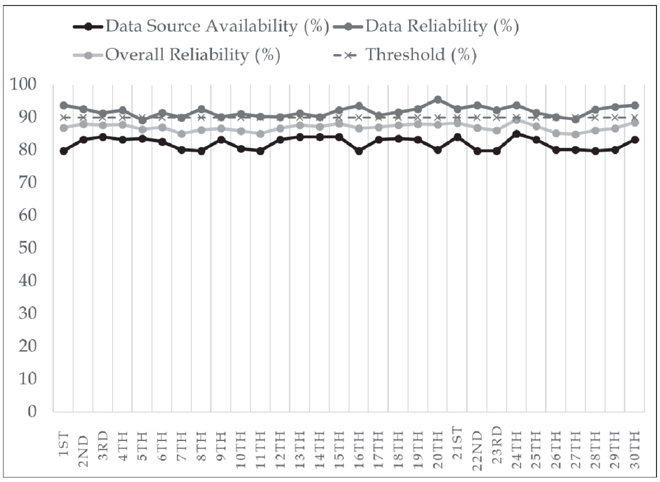

Fig. 7. Overall Polar A370 overall reliability results

D. Discussion Of Results

Concerning the final results of the 1st experiment (i.e. using the Fitbit Aria device), it can be argued that the mechanism can easily identify the reliability levels of a known device as it is already known that it is a reliable device, since the mechanism has already recognized it as a reliable device. However, the key innovation of the mechanism lies in the fact that it can easily assess the reliability levels of connected unknown devices, concluding that they can be either reliable or unreliable devices. This fact can be verified by the 2nd experiment (i.e. using the Misfit Path device) and the 3rd experiment (i.e. using the Polar A370 device). More particularly, through the 2nd experiment, the mechanism successfully evaluated the reliability levels of the connected unknown device, recognizing that it was a reliable device, thereby collecting and maintaining its data for reuse. However, through the 3rd experiment, even though the collected data of the connected device was found to have a very high degree of reliability, the availability of the device itself was not very high. As a result, the overall reliability of the device was affected, making it an inadequate device for the mechanism. Irrespectively of the final result, in all the experiments the reliability assessment was successful, as shown in the results of Tables II and III, where the mechanism in all the cases successfully calculated the overall reliability percentages of the connected devices. In deeper detail, based on the results that were recorded over the 30 days of the experiments, Table IV depicts the average (i) of the data transmission frequency values from the connected unknown devices to the mechanism, (ii) the rate of availability of the data, (iii) the percentage of data reliability derived from the collected data from those devices, and (iv) the percentage of the overall reliability calculated based on the results of the device availability and the reliability of the data.

TABLE IV. RESULTS OF MEASURED OVERALL RELIABILITIES

Mean Data Frequency	Mean Data Source Availability (%)	Mean Data Reliability (%)	Mean Overall Reliability (%)
2nd Experiment [Misfit Path]			
279	97,0	93,2	95,2
3rd Experiment [Polar A370]			
236	82,0	91,8	86,9

More specifically, for the 2nd experiment, based on the results of Table IV, it is observed that during the 30 days of the experiment, the average availability of the Misfit Path device was 97,0%, with more than half a day reaching almost the percentage of 98,0%. This proved that the Misfit Path device was almost always available, sending all of its data to the mechanism. The latter can be verified by the frequency of data that was sent by the device, where the average data rate was 279 measurements per day, reaching almost 100,0% of the perfect sending frequency of 288 measurements. In addition, based on the percentage of corresponding reliability of data retrieved daily from the device, it is observed that there were some days when the calculated percentage of reliability of the device data was not very high (~ 89,0%), indicating that the data that was collected contained some incorrect values. On the contrary, some other days the device produced quite reliable results, since the reliability of the collected data was extremely high (~ 96,0%). However, since the overall reliability was calculated based on both the availability of the Misfit Path device and the reliability of its collected data, it can be seen that the average reliability of the device is above the set threshold of 90,0%. More concretely, it has an average overall reliability of 95,2% (Table IV), which indicates that Misfit Path was a device with high levels of reliability.

Regarding the 3rd experiment, based on the results of Table IV, it is observed that during the 30 days of the experiment, the average availability of the Polar A370 was 82,0%, with no day availability exceeding the set threshold of 90,0%. This proved that this device was not always available, thus not sending to the mechanism all of its collected data, as the average frequency of its sent data was 236 measurements, a number far below the 100,0% of the perfect sending frequency of the 288 measurements. As a result, given the perfect conditions that the Polar A370 should send 288 measurements, and since in the current experiment it sent on average only 236 measurements per day for a total period of 30 days, this indicates that many measurements were not sent to the mechanism due to the fact that the device was not available. However, even if the daily measurements were less than expected, the estimated reliability of these measurements was extremely high, exceeding the percentage of 90,0%. Therefore, it can be observed that all the data that was collected by the Polar A370 device was extremely reliable, with an average reliability of 91,8% (Table IV). However, since the overall reliability was calculated based on both the availability of the Polar A370 device and the reliability of its daily collected data, it can be seen that even if the collected data was very reliable, the availability of the device was not as high as expected. This fact indicated that several measurements were lost, due to the fact that even if the measurements were recorded by the device, they were never sent to the mechanism. Thus, the overall reliability was not as high as expected, as it did not exceed the set threshold of the 90,0%, having as an average overall reliability 86,9% (Table IV). This suggests that the Polar A370 was not a device of high reliability. However, one could assume that since the data that was collected was fairly reliable, the portion of data that was not sent, due to the lack of device availability, would also be sufficiently reliable. This would lead to a higher overall reliability, indicating that Polar A370 would be a device of high reliability. However, since the mechanism is the one that ultimately decides the final results based on its outputs, and since this hypothesis could not be verified by it, the mechanism correctly predicted based on the available data that the Polar A370 was not a reliable device.

Based on all the aforementioned, it becomes clear that based on the collected data, the current mechanism provided quite reliable results, as all the outputs were also calculated manually and compared with the above results, verifying this fact. In particular, both the availability of the connected devices (i.e. Misfit Path and Polar A370), as well as the reliability of their collected data, and as a result the overall reliability of these devices, were manually calculated on a daily basis for 30 days, so as to compare these results with the results of the proposed mechanism. This was the main reason why the same number of days was chosen for the evaluation of the mechanism, in order to perform an objective comparison of the results. More particularly, with regards to the manual results, these are assumed to be of high precision, regarded as reference points, as these indicate the results that the proposed mechanism should produce (ideally) based on its overall application. Table V depicts the results of the manual calculation (i.e. manual results), compared with the results of the proposed mechanism (i.e. automatic results).

TABLE V. MANUAL AND AUTOMATIC OVERALL RELIABILITY RESULTS

Mean Data Frequency	**Mean Data Source Availability (%)**	**Mean Data Reliability (%)**	**Mean Overall Reliability (%)**
Manual results			
2nd Experiment [Misfit Path]			
280	97,2	93,2	95,2
3rd Experiment [Polar A370]			
288	100,0	97,0	97,5
Automatic results			
2nd Experiment [Misfit Path]			
279	97,0	93,2	95,2
3rd Experiment [Polar A370]			
236	82,0	91,8	86,9

In more detail, Table V shows the manually recorded reliability results of the 2nd and the 3rd experiments (i.e. the Misfit Path and the Polar A370) in combination with the corresponding automatic results of the proposed mechanism.

As for the Misfit Path, as shown in Table V, the manually calculated availability rate (97,2%) shows a minimal difference of 0,2% with the corresponding automatic rate of the mechanism (97,0%), thus not affecting the overall reliability of the Misfit Path device at all. This difference in the 0,2 degrees of percentage is due to the fact that the mechanism mistakenly rejected a measurement that should not have been rejected, thus affecting the total number of the recorded measurements. In addition, in terms of data reliability rates, as the mechanism used the SPSS tool for their calculation, and the same tool was used in the manual results, the calculated results were identical. Thus, as the calculation of the overall reliability depended on all of the aforementioned measurements, the manually calculated final values were not differentiated, as illustrated in Table V, since the Misfit Path device's overall reliability results were the same (95,2%).

As for the Polar A370, as shown in Table V, the manually calculated availability rate (100,0%) has a large difference rate (18,0%) with the corresponding availability rate calculated by the mechanism (82,0%). Therefore, due to this wide divergence, the overall reliability of the Polar A370 is affected, as in the first case (manual results) the Polar A370 is considered to have high reliability, while in the second case (automatic results) it is considered to be a device with low levels of reliability. This is because the mechanism was not able to connect to the device many times, even though it was always available, thus affecting the total number of measurements collected from it. On the contrary, in the manual results it was observed that the device had collected and sent to the mechanism all the measurements that it was expected to collect (i.e. 288 measurements), thus providing 100,0% availability. In addition, regarding the reliability of the data, it is observed that the manual results (97,0%) and the automatic results of the mechanism (91,8%) differ. However, this is reasonable, since 288 measurements were collected from the device during the manual results, while 236 measurements were collected from the same device during the automatic results. As a result, the reliability of the data for each case was calculated on the basis of the different number of measurements collected, yielding the expected results in both cases. Therefore, since the calculation of the overall reliability depends on all the above measurements, the manual results show that the Polar A370 is considered as a high reliability device, producing a very high overall reliability percentage (97,5%), well above the set threshold of 90,0%. On the contrary, in the automatic results, due to the limited availability of the device, the mechanism had concluded that the device was of low reliability. Therefore, additional experiments should be carried out in order to obtain a broader and more comprehensive view of calculating the overall reliability of the devices, by applying different degrees of weights between the availability and the data reliability. However, based on all the outputs, it can be concluded that through the provided mechanism it is effective to decide whether a connected device is considered reliable or not, by combining (i) the results of its availability and the number of the collected measurements, and (ii) the corresponding reliability of the collected data.

On the basis of all the aforementioned results, it is observed that the proposed mechanism has produced quite accurate and reliable results, validating its purpose for estimating the reliability levels of both known and unknow IoT medical data sources, and thus keeping only the reliable data that comes from only reliable data sources. Based on the three (3) different experiments that were analyzed in this Section, it has been demonstrated that the mechanism is capable of operating equally efficiently and effectively in the three (3) different possible scenarios of an incoming data source. Therefore, the mechanism is able to easily and automatically identify the reliability of: (i) a known data source, proving that if the type of the data source is known in advance, it can very easily estimate its reliability levels, (ii) an unknown data source, proving that although the type of the data source is not known in advance, it can easily estimate its reliability levels, and finally collect its data if it considers it to be a reliable data source, and (iii) an unknown data source, proving that although the type of the data source is not known in advance, it can easily assess its reliability levels, but do not collect its data if it considers it as an unreliable data source.

V. Conclusions

In this paper, a mechanism was proposed for capturing the reliability of heterogeneous IoT medical data sources of both known and unknown nature. Shortly, through this mechanism a 3-stepped approach was implemented for coping with this challenge. Initially, the calculation of the data sources' availability occurred, followed by the reliability calculation of the produced data of the corresponding data sources. Finally, the combination of the results of the two (2) aforementioned stages occurred, so as to calculate the overall reliability levels of each available data source, calculating whether each one of these would be considered as reliable or not, and its retrieved data would be kept for further analysis. This mechanism was evaluated through a specific experiment, concluding that it was sufficient enough for assessing data sources' reliability.

A key aim of our future work is to improve the current mechanism so as to be able to calculate the overall reliability of a data source by following a more global approach, taking into account not only its availability and the reliability of its input, but also the reliability of its output. In particular, data reliability will be purified so as to determine the quality of the data, in order to identify any errors related to compliance with specific constraints, ensuring that the collected data comply with certain predefined rules or constraints (e.g. compliance with specific types of data, compliance with values' representation, etc.), thus eliminating the corresponding errors by applying corrective actions to the detected incorrect data. By combining the results of these operations with the corresponding results of the reliability of this data and their data sources' availability, the overall reliability of the data sources will be predicted in a maximum degree. Moreover, an update of the mechanism relies on the way in which the reliability procedure concludes to the final results, by dynamically adjusting the value of the set threshold. In more detail, in each scenario where the mechanism is implemented, the overall slope of the data source reliability measurements (e.g. exponential, linear, etc.) will be recorded to dynamically adjust the set threshold according to each different scenario.

ACKNOWLEDGMENT

The research leading to this result has received funding from the European Union's Horizon 2020 research and innovation programmes under grant agreement No 727560 (CrowdHEALTH project) and grant agreement No 826106 (InteropEHRate project).

REFERENCES

[1] Population health outcomes, http://www.healthcatalyst.com/population-health-outcomes-3-keys-to-drive-improvement

[2] The role of IoT in the healthcare industry, https://hackernoon.com/the-role-of-internet-of-things-in-the-healthcare-industry-759b2a1abe5

[3] Pham, C., Lim, Y., & Tan, Y. (2016). Management architecture for heterogeneous IoT devices in home network. In 2016 IEEE 5th Global Conference on Consumer Electronics, pp. 1-5

[4] Mezghani, E., et al. (2015). A semantic big data platform for integrating heterogeneous wearable data in healthcare. Journal of medical systems, 39(12), pp. 185

[5] The Evolution of Smart Health Care, https://www.hticonference.com/wp-content/uploads/2018/03/gx-lshc-hc-outlook-2018.pdf

[6] Arts, D. G., et al. (2002). Defining and improving data quality in medical registries: a literature review, case study, and generic framework. Journal of the American Medical Informatics Association, 9(6), pp. 600-611

[7] Lee, C., et al. (2017). Big healthcare data analytics: Challenges and applications. In Handbook of Large-Scale Distributed Computing in Smart Healthcare, pp. 11-41

[8] Scott, M. T., et al. (2016). The validity and reliability of global positioning systems in team sport: a brief review. The Journal of Strength & Conditioning Research, 30(5), pp. 1470-1490

[9] Gwet, K. L. (2014). Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters. Advanced Analytics, LLC

[10] Toporkov, A. A. (2007). Assessment of service life of sophisticated medical equipment. Biomedical Engineering, 41(3), pp. 122-127

[11] Leape, L. L. (1997). A systems analysis approach to medical error. Journal of evaluation in clinical practice, 3(3), pp. 213-222

[12] Reid, R. E., et al. (2017). Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. Journal of science and medicine in sport, 20(6), pp. 578-582

[13] TJM. Kooiman, et al. (2015). Reliability and validity of ten consumer activity trackers. BMC, vol. 7, pp. 24

[14] Ruggiero, L., Dewhurst, S., & Bampouras, T. M. (2016). Validity and reliability of two field-based leg stiffness devices: implications for practical use. Journal of applied biomechanics, 32(4), pp. 415-419

[15] Vohralik, S. L., et al. (2015). Reliability and validity of a smartphone app to measure joint range. American journal of physical medicine & rehabilitation, 94(4), pp. 325-330

[16] Lee, J. A., et al. (2014). Reliability and validity of a point-of-care sural nerve conduction device for identification of diabetic neuropathy. PloS one, 9(1), e86515

[17] Sonter, J., et al. (2015). Inter-rater reliability of automated devices for measurement of toe systolic blood pressure and the toe brachial index. Blood pressure monitoring, 20(1), pp. 47-51

[18] Bedekar, N., et al. (2014). Inter and intra-rater reliability of mobile device goniometer in measuring lumbar flexion range of motion. Journal of back and musculoskeletal rehabilitation, 27(2), pp. 161-166

[19] Ganio, M. S., et al. (2009). Validity and reliability of devices that assess body temperature during indoor exercise in the heat. Journal of athletic training, 44(2), pp. 124-135

[20] Tsarouhas, P. H. (2010). Measuring the efficiency of medical equipment. World Academy of Science, Engineering and Technology, 4(44), pp. 340-345

[21] Baker, R. D. (2001). Data-based modeling of the failure rate of repairable equipment. Lifetime Data Analysis, 7(1), pp. 65-83

[22] Toporkov, A. A. (2008). Criteria and methods for assessing reliability of medical equipment. Biomedical Engineering, 42(1), pp. 11-16

[23] Singh, J., & Goel, P. (1995). Availability analysis of a standby complex system having imperfect switch-over device. Microelectronics Reliability, 35(2), pp. 285-288

[24] Ion, R. A., Sonnemans, P. J., & Wensing, T. (2006). Reliability prediction for complex medical systems. In RAMS'06. Annual Reliability and Maintainability Symposium, pp. 368-373

[25] Early reliability prediction based on Field Data, http://alexandria.tue.nl/extra2/afstversl/tm/roelfsema2004.pdf

[26] Mavrogiorgou, A., Kiourtis, A., & Kyriazis, D. (2019). A plug 'n' play approach for dynamic data acquisition from heterogeneous IoT medical devices of unknown nature. Evolving Systems, pp. 1-21

[27] Mavrogiorgou, A., et al. (2018). Capturing the Reliability of Unknown Devices in the IoT World. In 5th International Conference on Internet of Things: Systems, Management and Security, pp. 62-69

[28] Availability and the Different Ways to Calculate It, http://www.weibull.com/hotwire/issue79/relbasics79.htm

[29] Guide to the expression of uncertainty in measurement, https://www.bipm.org/en/publications/guides/gum.html

[30] Types of Reliability, http://changingminds.org/explanations/research/design/types_reliability.htm

[31] Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of chiropractic medicine, 15(2), pp. 155-163

[32] Java SPSS Reader, http://spss.pmstation.com/docs/spssr/api/index.html

[33] Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. Tutorials in quantitative methods for psychology, 8(1), pp. 23

[34] Netbeans, https://netbeans.org

[35] Apache Derby, https://db.apache.org/derby/