Privacy and eHealth-enabled Smart Meter Informatics

Georgios Kalogridis and Saraansh Dave
Toshiba Research Europe Ltd, Telecommunications Research Laboratory
32 Queen Square, Bristol, BS1 4ND, UK
Email: {george, saraansh.dave}@toshiba-trel.com

***Abstract*—The societal need for better public healthcare calls for granular, continuous, nationwide instrumentation and data fusion technologies. However, the current trend of centralised (database) health analytics gives rise to data privacy issues. This paper proposes sensor data mining algorithms that help infer health/well-being related lifestyle patterns and anomalous (or privacy-sensitive) events. Such algorithms enable a user-centric context awareness at the network edge, which can be used for decentralised eHealth decision making and privacy protection by design. The main hypothesis of this work involves the detection of atypical behaviours from a given stream of energy consumption data recorded at eight houses over a period of a year for cooking, microwave, and TV activities. Our initial exploratory results suggest that in the case of an unemployed single resident, the day-by-day variability of TV or microwave operation, in conjunction with the variability of the absence of other cooking activity, is more significant as compared with the variability of other combinations of activities. The proposed methodology brings together appliance monitoring, privacy, and anomaly detection within a healthcare context, which is readily scalable to include other health-related sensor streams.**

I. Introduction

Changes in demographics, such as longer life expectancy and smaller working populations (in some countries), alongside a technological shift towards ubiquitous sensor technologies can be the start of major transformation in the healthcare industry. OECD data [1] shows that the average spending on health per capita has risen over the last few decades with USA expenditure increasing from $1000 to approximately $7500 between 1980 and 2007. This represents total expenditure on healthcare in the USA of 16% of GDP [1]. This increase in costs will be further accelerated due to changing demographics as the number of people dependent on care increases as a proportion of those that can support them, termed as a 'Silver Tsunami' [2].

In order to address this challenge there has been interest in using sensors and other data sources to remotely monitor individuals for a more personalised health service. This has a number of advantages, for example, allowing medical staff to monitor specific patients without requiring an overnight stay in a hospital. Notable examples of such sensor-based technologies include wearable systems for (ageing population) health monitoring and prognosis [3], personalised diabetes treatment [4], treatment of heart failure patients [5], and pulse oximetry (blood oxygen level monitoring) [6].

Further, there are possibilities of focusing on the general public by developing systems which monitor activity and daily living habits in order to improve health and well-being [7]. For example, a large Australian study has shown that a strategy for reducing the risk of abnormal glucose metabolism could focus on the reduction of sedentary behaviours such as watching television [8]. Linking these types of observable behaviours, to specific illnesses is difficult and requires carefully designed medical trials for reliable results. However, the analysis and treatment of various data sets that can be used to make health related inferences is equally important and challenging. There are a number of challenges associated with this 'model-based' approach to care [2]; data acquisition, models of sensors and monitoring processes, estimation and classification, information fusion, sampling and interpolation, modelling behaviour, detection of anomalies, and alarm fatigue.

The flip side of (continuous and nationwide) instrumentation for health monitoring, and the focus of this paper, is privacy. Health or health-relevant data (e.g. mental health, genomics, and dietary habits) is one of the most sensitive categories of data. In this sense, privacy may be perceived as a technology barrier of health analytics and biomedical engineering. Thus the full potential of (big) sensor data-powered care may only be realised if privacy issues are addressed. This process includes designing secure data storage and communication systems, privacy-preserving distributed (e.g. cloud) computing, access control and data ownership, data fusion control and decisional interference.

This paper aims to progress work in eHealth privacy domain by means of stream informatics to help infer health/well-being related lifestyle patterns and anomalous (privacy-sensitive) events. The proposed system is codenamed as 'INformatics For sensoRS' (INFeRS), and it is designed to run at the network edge. In particular, this work analyses real granular appliance level energy consumption data that are linked to patterns of cooking and TV activities. In doing so we use real empirical data to bring together appliance monitoring, privacy, and event detection within a healthcare context. The rest of the paper is organised as follows; §II discusses related work, §III presents the INFeRS system and experiment, §IV presents the stream mining methodology and the initial statistical analysis, §V presents a method of detecting atypical events, §VI discusses the implications to privacy and eHealth, whilst §VII summarises with a conclusion.

II. Related Work

The issue of privacy is becoming more relevant today in the era of social networks and complex engineered systems such as eHealth and the Internet of Things (IoT), which rely on large amount of data collected from customers (e.g. via sensors) to provide an optimised service.

Data anonymity and de-identification (i.e., the process of limiting access to links to personally identifiable information) has been widely proposed as a privacy-preserving technique. However, de-identification assumes a trusted authority and it may not preclude an inference to personal information. Another popular approach for privacy involves rule-based access control, which may be either hard-coded or dynamic (e.g. expert-systems) [9]. A more systemic approach to the medical privacy protection is proposed in [10], which is based on the Privacy by Design (PbD) framework. PbD encourages embedding privacy ensuring features as the default functionality, designing user-centric solutions, and constructing privacy enhancing functionalities in a positive-sum manner.

From a data mining perspective, a fundamental requirement involves privacy of non-obvious personal information. In [11], the authors propose two privacy-preserving methods: value-class membership and value distortion. Both methods allow users to provide modified values for sensitive fields in the database. It can be argued that the privacy is larger if the information distance between the original and modified values is larger [12]. However, such methods assume a trade-off between privacy and precision. A common engineering challenge, regardless of the adopted approach, involves the evaluation of privacy (or sensitivity). In the data mining community, this problem essentially relates to the problem of publishing statistics at certain timestamps [13], which is similar to releasing streaming data using k-anonymity, l-diversity, or differential privacy [14]. In [15] it is further argued that typical behaviour does not violate privacy. Only detected outliers can breach individual privacy. Thus, they use the K-divergence to measure the deviation from typical values and, hence, the potential privacy violations.

Interestingly, the notion of detecting atypical behaviour is a key requirement in the healthcare context. This involves a reliable (early) detection of unexpected anomalies, such as falls, myocardial infarction, and stroke. A common approach to this area involves the identification of statistical outliers, ‘surprises’ [16], or incongruent events [17].

This paper combines some of the above techniques within a novel information system (called INFeRS) which analyses stream sensor data to detect atypical (rare) events. On this basis, INFeRS helps detect and classify both eHealth and privacy related context, based on which information personal eHealth sensor information flows may be configured as appropriate. While the presented analysis focuses on eHealth implications of home appliance operation, INFeRS can be readily applied in different streams of health sensor data.

III. INFeRS Approach

A. The System

The INFeRS system architecture, given in Fig. 1, is designed to enable context-based eHealth and privacy data flow control. The main assumptions are as follows.

- An eHealth stakeholder (e.g. NHS) is interested in a (statistical) summarisation of sensor information. That is, raw data remain confidential (e.g. they may be securely stored in private network locations).

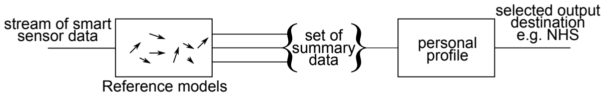
stream of smart sensor data
Reference models
set of summary data
personal profile
selected output destination e.g. NHS

Fig. 1. High level INFeRS system architecture.

- The control of data summary flows from a personal profile to multiple eHealth stakeholders is ruled by the INFeRS anomaly detection algorithm.
- Key parameters of the data summary flow control involve the level of privacy required, the level of stakeholder trust, and the level of medical importance.

B. The Case of Energy Data

To support the premise of using home appliance energy consumption sensor data in a general public eHealth system, we discuss a number of key features from [2].

- Economical feasibility and scalability: Mandated deployments of smart meter infrastructures (e.g. in Europe) provide an opportunity to acquire aggregate house consumption data on a nationwide level. Further, the use of Non-intrusive Appliance Load Monitoring (NiALM) algorithms [18], originally developed for energy disaggregation applications, can be used to analyse Activities of Daily Living (ADLs) [19].
- Unobtrusiveness and usability: Sensors should be transparent and avoid to interfere with user lifestyle (e.g. requiring too frequent battery charging). Energy sensors are ideal in this sense as they enable activity monitoring without constraining ADLs. However, this comes at the cost of increased uncertainty.
- Continuity: Smart meters collect energy data frequently (e.g. once a minute), which is important for patient-specific trends and interventions.

IV. Data Mining

A. Hypothesis and Experimental Setup

The basic hypothesis of this study is that the analysis of granular (timestamped) cooker, microwave and TV sensor data may help infer ADLs and health information. To make inferences, we analyse stream data by studying a) the dynamics, e.g. the variability of event occurrence probability distributions, and b) the variability from different event spaces and the discovery of the event spaces that exhibit the most significant variability. From the point of view of privacy protection, it is interesting to observe how atypical behavioural patterns are detected, and how such a detection might control the communication of such patterns to different stakeholders.

As a proof-of-concept we analyse timestamped data for cooker, microwave and TV data that have been collected twice a minute from eight houses in Bristol City, UK, for approximately the duration of one year. All eight houses were chosen so that they have one occupant, which helps assume a good correlation between appliance usage and the behaviour of a single person.

B. Behavioural Patterns

In our setting the number of appliances is 3. We can thus define a space of 4 possible events, including the zero event where no appliance operates. Further, we consider than any subset of the set of events may be used to define an event space. In general, for a set of N streams, a number of $2^N - 1$ event spaces may be defined (excluding the empty event space).

To mine the events from given energy data, we implement a sequential batch window algorithm that counts the usage of each event, for each separate hour interval of each day. Given the twice a minute interval between subsequent measurements, each event may be used a maximum of 120 times per hour. Finally, we mine two types of behaviour: 1) capturing the duration of the usage of each appliance, and 2) capturing the number of times of switching on an appliance (including switching on from standby).

C. Boxplot Evaluations

Initial exploration of the data sets is around daily patterns of appliance usage. Fig. 2 shows a boxplot of the duration of TV usage for House 1 over the whole time period (approximately 1 year) broken down into hourly figures. The horizontal line on each bar shows the median value of all data points in that time slot, whilst the filled circular marker shows the average usage. The upper and lower boundaries of the box show the first and third quartiles whilst the whiskers show data within 1.5 times of the interquartile range. Any points outside this range are shown as circular points. The triangular markers indicate the group average (of all the houses) over the same one year period split into 24 hours. This graph provides a number of insights into occupant behaviour. From the hours of 10 p.m. to 9 a.m. the average duration of TV usage for house 4 was higher than the group average. Comparing these trends with community averages as they build up over time can help identify health risks associated with lifestyles. For example, the consistently high level of TV usage along with the time at which this occurs points to inconsistent sleeping patterns and sedentary lifestyle patterns. Fig. 3 shows the same type of data for the microwave appliance. Again, what we see is a greater use of the microwave compared to the group average. This can help build on the observations in Fig. 2 where a high use of the microwave may hint towards poor diet which combined with high level of TV use, may point to health risks associated with diet and sedentary lifestyles.

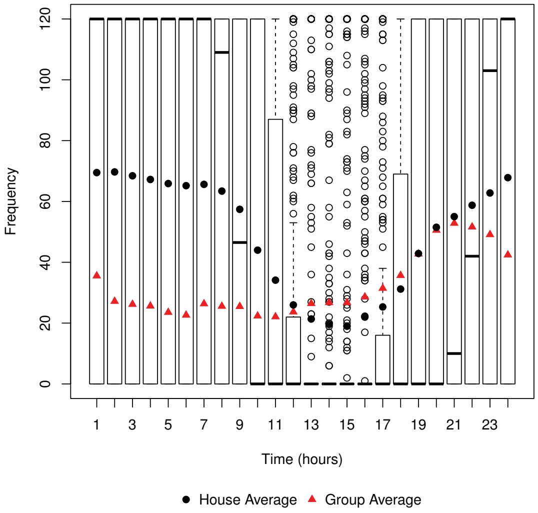

Fig. 2. Boxplot of the duration of TV usage for house 1 along with group average values.

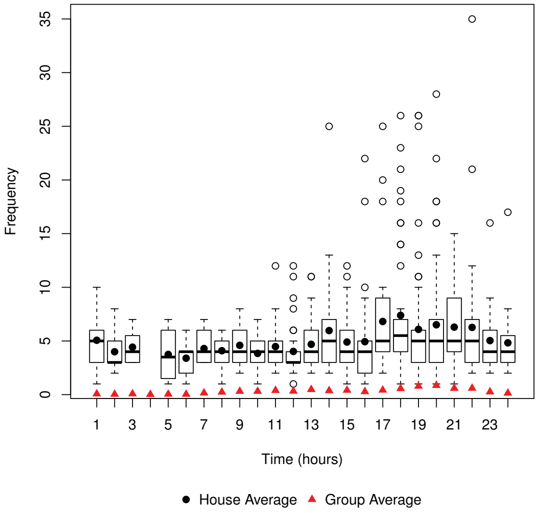

Fig. 3. Boxplot of the duration of Microwave usage for house 1 along with group average values.

V. Atypicality

A. Empirical Probability distribution

The method of atypicality, which we use to detect rare events, is based on the concept of Empirical Probability distributions (EPDs) [15], which is adapted here as follows.

Definition 1: (Interval EPD.) Let X be the alphabet of an event space defined by a multi-sensor stream p. The EPD, P_p, of p is defined by $P_p(b) = N(b|p)/N$ for any $b \in X$, where $N(b|p)$ is the number of occurrences of the b event within p.

Consequently, for any particular time interval the sensor data can be sampled and the number of occurrences for each event b can be counted in order to obtain the EPD. The method of EPD is further used to define an averaged distribution.

Definition 2: (Average EPD.) Assume that L distinct measurements have been obtained providing with L different sample paths of the multi-sensor stream. The average EPD, Q_p, is obtained by counting number of events across all L sample paths (ideally, $L \to \infty$).

In our setting $L = 8$ is the number of different houses participating into the experiment.

B. Divergence

Given an average EPD Q_p, the deviation of an interval EPD P_p from Q_p may be measured by functions which are in general called divergence measures. The simplest one is called the relative entropy $D(P_p||Q_p)$ [20] and it is given by $D(P_p||Q_p) := \sum_b P_p(b) \log \frac{P_p(b)}{Q_p(b)}$. We note that the relative entropy is infinity if there exists a symbol b such that $P_p(b) \neq 0$ and $Q_p(b) = 0$. In this paper we use a divergence measure called K-divergence [20] given by

$$K(P_p||Q_p) := \sum_b P_p(b) \log \frac{2P_p(b)}{P_p(b) + Q_p(b)}. \qquad (1)$$

The advantages of the K-divergence as compared to the relative entropy is that it is always defined, for all values of P_p and Q_p, and its value is between zero and one. It is noted that if the K-divergence increases, the relative entropy increases as well. Next we define atypicality.

Definition 3: If Q_p is the average EPD of a number of samples and P_p is the interval EPD of one sample p, the atypicality of p is the deviation of P_p from Q_p.

Given that in this paper we use K-divergence, atypicality is given by (1). We note that, using a data mining language, a sample p provides a *bag of (basic) events*. Thus, a large atypicality indicates an *atypical bag of events*, and a small atypicality indicates a *typical bag of events*. We also suppose that a typical bag contains mostly *cyclical events*, and an atypical bag contains some rare events.

One way to understand atypicality is to look at the definition of the K-divergence. Each term $\log \frac{2P_p(b)}{P_p(b)+Q_p(b)}$ in (1) will be greater than zero if $P_p(b) > Q_p(b)$, and it will be zero if $P_p(b) = Q_p(b)$; otherwise it will be negative. Thus, $P_p(b) >> Q_p(b)$ is a sufficient condition for a spike in atypicality.

C. Evaluations

Observing the complete set of atypicality evaluations, both for mining appliance duration data and switch-on appliance data, across different hours and days, both with or without the zero event, for all 8 houses, allow us to make a number of remarks: 1) The appliance duration data (On-On events) provide statistical properties for hourly and daily sampling, as the Off-On events are more sparse. 2) While daily atypicalities have value in summarising ADLs, hourly atypicalities provide a better insight of the diurnal cycle. 3) The exclusion of the zero event intensifies potential spikes of atypicality, as the zero event is typically (and by far) the most frequent event. These observations are highlighted in Fig. 4, where NZ signifies the exclusion of the zero event from the EPD. We further make a distinction between spikes (e.g. for On-On NZ events between 50–100 days for House 1, and across all days for House 8) and consistently high atypicality (e.g. during the last 100 days on House 1). The former case is more likely to be linked to atypical behaviour, whereas in the latter case is clearly a case of false alarm due to unknown non-behavioural (technical) problems. Finally it is worth observing that the hourly atypicality patterns of House 1 appear to have smaller values between 9-15 hours, which is consistent with the observations make in Fig. 2 and Fig. 3.

Further we apply the notion of atypicality in different event spaces, in order to explore combinations of interval EPDs that exhibit a greater variability, as compared with their average EPD. That is, we expand on the cases where the EPDs of duration or of switch-on events is taken across either the 3 appliance events (cooker, microwave, TV) or across the 4 events, if the zero event is included. All possible combinations of event spaces and their corresponding atypicalities is shown in Fig. 5. In this figure we observe that the atypicality of the event space comprising the duration of the microwave and the absence of any other activity (TV and cooker) is more significant as compared with the atypicality of other combinations of event spaces. This infers that atypical behaviour is more likely to be observed in the case where both the microwave usage and the absence of other appliance usage is taken into account. While this may lack statistical significance (e.g. due to alarm fatigue), there is still value in regarding this empirical result as a proof-of-concept or initial evidence for a hypothesis that would need to be medically tested (in large scale trials).

Table I summarises the results of the atypicality alarms of different bags of events, across all houses and for both the appliance duration and switch-on events. The given summary is readily obtained by counting the number of days where atypicality exceeds a threshold value, thus triggering an alarm. Here, we take the threshold value of 0.4. Similar to Fig. 5, we observe that high atypicality becomes more frequent in the cases where TV or microwave, and no other appliance is considered. This means that a) the usage of other appliances reduces the number of high atypicalities, and b) the zero event increases the number of high atypicalities (even through it may reduce their intensities as previously discussed).

VI. Privacy and eHealth Implications

The atypicality analysis above illustrates how often rare (atypical) events occur within a multi-sensor stream. Atypical events are important because they are likely to compromise health sensor data privacy. To understand the importance of atypical events better, one may refer to the Shannon's definition of information saying that the amount of information conveyed by an event A is $\log(1/P(A))$. Thus, the smaller the probability of the event the larger amount of its information. That is, an atypical event carries a large amount of information. This implies that it is more important to protect an atypical event as compared to a typical one.

From an eHealth perspective, the techniques for detection and inference presented scale as the number of sensors increase. However this increase in sensor information requires an underlying system to manage the data and extract meaningful insights. The atypicality detection method has shown how multiple data streams can be analysed together to identify patterns or events that are unexpected. In the previous section we introduced a method of counting occurrences above a given threshold, as shown in Table I. By developing this approach, a trigger mechanism can be used to generate an action or alert. Consider the following example: an atypical event (e.g. cooker and microwave event for House 1) occurs and is detected by an in-home application. This triggers an action whereby a summary of the event detected is presented to the householder via an App. The user can then make a decision based on this and decide if it needs to be sent onward to other stakeholders such as the family doctor. This might be more critical if the

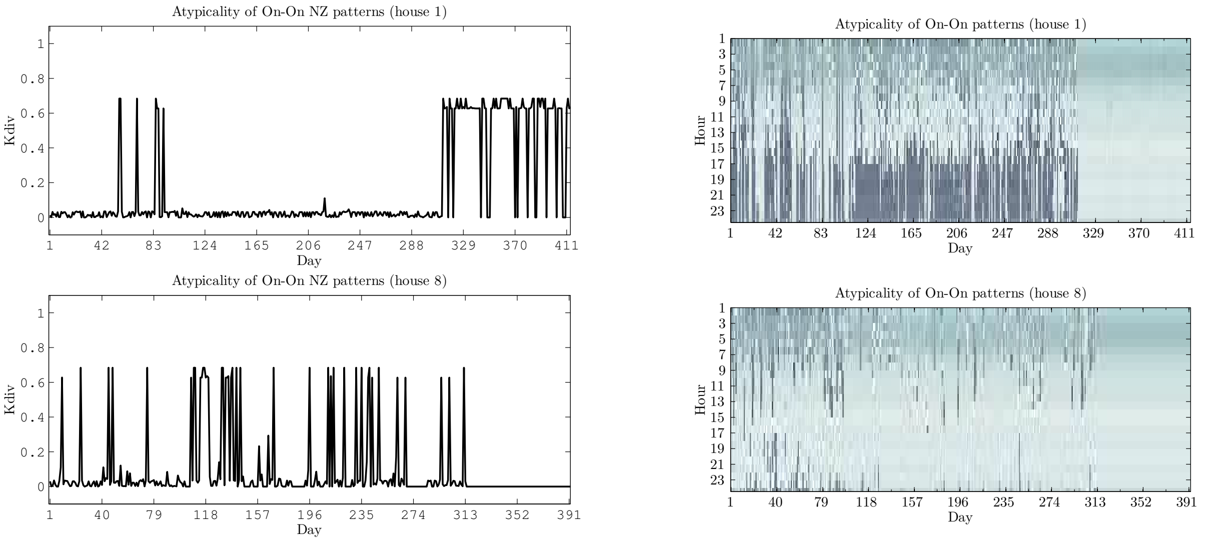

Fig. 4. Atypicalities for daily NZ appliance duration (left) and hourly switch-on events (right) for House 1 (top) and House 8 (bottom).

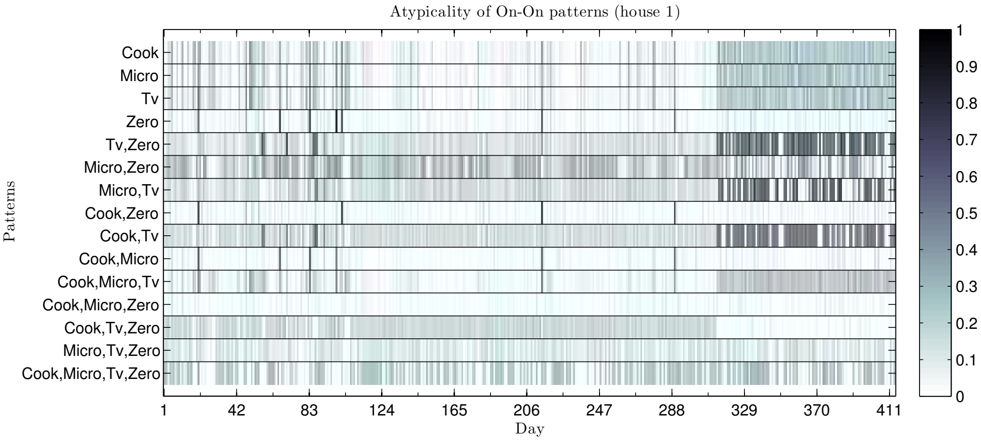

Fig. 5. Atypicality of all distinct bags of duration of operation (On-On) appliance events for House 1.

TABLE I. YEARLY FREQUENCIES OF DAYS EXHIBITING ATYPICALITIES LARGER THAN 0.4, FOR DIFFERENT BAGS OF ON-ON & OFF-ON EVENTS.

House No.	1	2	3	4	5	6	7	8
Cook	0 & 0	3 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0
Micro	0 & 0	3 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0
TV	0 & 0	3 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0
Zero	8 & 0	15 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0
TV,Zero	9 & 25	0 & 1	8 & 18	68 & 27	0 & 0	4 & 29	10 & 7	50 & 76
Micro,Zero	0 & 25	15 & 15	0 & 0	27 & 27	0 & 0	11 & 22	15 & 15	125 & 24
Micro,TV	4 & 0	0 & 0	33 & 33	54 & 14	0 & 0	4 & 28	8 & 0	20 & 105
Cook,Zero	4 & 0	14 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0
Cook,TV	9 & 0	0 & 8	18 & 19	41 & 41	0 & 0	4 & 13	10 & 13	46 & 31
Cook,Micro	7 & 0	11 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0
Cook,Micro,TV	0 & 0	3 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0
Cook,Micro,Zero	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0
Cook,TV,Zero	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0
Micro,TV,Zero	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0	42 & 42	0 & 0	0 & 0
Cook,Micro,TV,Zero	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0	0 & 0

householder is an at-risk patient who needs support and/or monitoring from healthcare personnel.

Whilst the sensor data used in this paper is predominantly around appliance energy consumption it can easily be extended to include other data streams such as pulse rate, electrocardiogram data, and body temperature. This would be able to provide links between personal parameters alongside domestic activity resulting in richer context and more directed atypical event detection.

The key limitation of this work is the assumed link between detecting activities in the house and health related issues. There is evidence from other fields of research where links between activities, diets, and health related issues has been found. However the inference of activities and thus health risks through the detection of atypical appliance usage is an area which has not been explored in detail. The methods developed here can be applied to such a situation and we urge greater collaboration between medical trials, sensor technology, and informatics in order to develop our understanding in this area.

VII. Conclusion

The paper has proposed INformatics For sensoRS (INFeRS): a system to enable context-based health and privacy data flow control. The methods and algorithms discussed have shown that sensor information can highlight atypical behaviour which can be used to identify health related issues. For example, sedentary behaviour can be identified by analysing TV patterns both in a single house and in comparison to a group. All the information being collected through different data sources gives rise to privacy concerns. The method proposed identifies atypical events allowing the numerous data streams to be reduced to usable information. The proposed system will then allow for summary data to be sent to appropriate stakeholders as opted by the householder.

The methodology presented here can be used to provide insights for designing larger scale medical trials and explore the link between activity inference from sensors to specific health risks and medical conditions.

In the future we will investigate the eHealth contextual analysis of environmental sensor data in the case where more that one users are present. Future research will also focus on the challenges of false alarms and the classification of specific health risks with the most appropriate data sources. We envisage that this research direction will help close the gap among nationwide eHealth instrumentation, health indications, atypical events, and their connection to privacy analytics.

Acknowledgment

The authors would like to thank Dr. Eirini Spyropoulou for contributing to this work through technical discussions, and the Directors of Toshiba Telecommunications Research Laboratory for their support.

References

[1] G. F. Anderson and P. Markovich, "Multinational Comparisons of Health Systems Data." The Commonwealth Fund, 2009., New York, Tech. Rep., 2009.

[2] M. Pavel, H. B. Jimison, H. D. Wactlar, T. L. Hayes, W. Barkis, J. Skapik, and J. Kaye, "The role of technology and engineering models in transforming healthcare," *IEEE Reviews in Biomedical Engineering*, vol. 6, pp. 156–77, Jan. 2013.

[3] A. Pantelopoulos and N. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 40, no. 1, pp. 1–12, Jan 2010.

[4] E. Geisler and N. Wickramasinghe, "The Role and Use of Wireless Technology in the Management and Monitoring of Chronic Diseases," IBM Center for The Business of Government, DC, Tech. Rep., 2007.

[5] D. Scherr, P. Kastner, A. Kollmann, A. Hallas, J. Auer, H. Krappinger, H. Schuchlenz, G. Stark, W. Grander, G. Jakl, G. Schreier, and F. Fruhwald, "Effect of home-based telemonitoring using mobile phone technology on the outcome of heart failure patients after an episode of acute decompensation: Randomized Controlled Trial," *Medical Internet Research*, vol. 11, no. 3, p. e34, 2009.

[6] S. Lam, K. L. Wong, K. O. Wong, W. Wong, and W.-H. Mow, "A smartphone-centric platform for personal health monitoring using wireless wearable biosensors," in *7th Int. Conf. on Information, Communications and Signal Processing, ICICS 2009*, Dec 2009, pp. 1–7.

[7] S. A. Lowe and G. Ólaighin, "Monitoring human health behaviour in one's living environment: a technological review." *Medical Engineering & Physics*, vol. 36, no. 2, pp. 147–68, Feb. 2014.

[8] D. W. Dunstan, J. Salmon, N. Owen, T. Armstrong, P. Z. Zimmet, T. A. Welborn, A. J. Cameron, T. Dwyer, D. Jolley, and J. E. Shaw, "Physical Activity and Television Viewing in Relation to Risk of Undiagnosed Abnormal Glucose Metabolism in Adults," *Diabetes Care*, vol. 27, no. 11, pp. 2603–2609, 2004.

[9] N. Zakaria, L. K. Yew, N. Alias, and W. Husain, "Protecting privacy of children in social networking sites with rule-based privacy tool," in *High Capacity Optical Networks and Enabling Technologies (HONET)*, Dec 2011, pp. 253–257.

[10] A. Cavoukian and K. Emam, "A Positive-Sum Paradigm in Action in the Health Sector," Office of the Information and Privacy Commissioner of Ontario, Tech. Rep., March 2010.

[11] R. Agrawal and R. Srikant, "Privacy preserving data mining," in *ACM SIGMOD International Conference on Management of Data*, vol. 29. ACM, 2000, pp. 439–450.

[12] G. Kalogridis, R. Cepeda, S. Z. Denic, T. Lewis, and C. Efthymiou, "Elecprivacy: Evaluating the privacy protection of electricity management algorithms," *IEEE Trans. on Smart Grid*, vol. 2, Dec 2011.

[13] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacy-preserving Data Publishing: A Survey of Recent Developments," *ACM Comput. Surv.*, vol. 42, no. 4, pp. 14:1–14:53, Jun 2010.

[14] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "L-diversity: Privacy Beyond K-anonymity," *ACM Trans. Knowl. Discov. Data*, vol. 1, no. 1, Mar 2007.

[15] G. Kalogridis and S. Z. Denic, "Data Mining and Privacy of Personal Behaviour Types in Smart Grid," in *IEEE 11th Int. Conference on Data Mining Workshops*, ser. ICDMW'11, December 2011, pp. 636–642.

[16] L. Itti and P. F. Baldi, "Bayesian Surprise Attracts Human Attention," in *Advances in Neural Information Processing Systems, Vol. 19*. Cambridge, MA: MIT Press, 2006, pp. 547–554.

[17] D. Weinshall, A. Zweig, H. Hermansky, S. Kombrink, F. W. Ohl, J. Anemuller, J.-H. Bach, L. V. Gool, F. Nater, T. Pajdla, M. Havlena, and M. Pavel, "Beyond Novelty Detection: Incongruent Events, When General and Specific Classifiers Disagree," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1886–1901, 2012.

[18] E. Vogiatzis, G. Kalogridis, and S. Z. Denic, "Real-time and low cost energy disaggregation of coarse meter data," in *4th IEEE/PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, Oct 2013, pp. 1–5.

[19] H. Song, G. Kalogridis, and Z. Fan, "Short paper: Time-dependent power load disaggregation with applications to daily activity monitoring," in *1st IEEE World Forum on Internet of Things (WF-IoT)*, March 2014, pp. 183–184.

[20] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, Inc. New York, NY, USA, 2006.