Kensor: Coordinated Intelligence from Co-Located Sensors

Olivera Kotevska
Computer Science and Mathematics
Oak Ridge National Laboratory
Oak Ridge, TN, U.S.A.
kotevskao@ornl.gov

Kalyan Perumalla
Computer Science and Mathematics
Oak Ridge National Laboratory
Oak Ridge, TN, U.S.A.
perumallaks@ornl.gov

Juan Lopez Jr.
Cyber & Applied Data Analytics
Oak Ridge National Laboratory
Oak Ridge, TN, U.S.A.
lopezj@ornl.gov

***Abstract*—Internet of Things (IoT) is becoming more pervasive in many installations, including homes, manufacturing plants, and industrial facilities of all kinds. The data that IoT produces is a reflection of usual behavior such as daily routines and scheduled tasks, but also from unexpected behavior due to unintentional or undesirable abnormalities. Here, we focus on achieving coordinated intelligence about normal and abnormal phenomena from multiple sensors that are geographically co-located in close proximity, monitoring and controlling a set of co-located devices. Given a set of co-located sensors, we seek an intelligent approach that would automatically determine the "normal" patterns of behaviors among the correlated sensors. After normal behavior is extracted, later monitoring should detect any deviant variations over time. An example application is an entry monitoring and alert system for facilities such as nuclear reactors, where badge readers, door locks, lights, weight trackers and other co-located sensors at the entry point are collectively tracked. To address this problem, we identify the possible solution approach that can be used to solve its different variants. The implemented model is developed as a combination of rules and Markov Chain methods.**

***Index Terms*—IoT, Markov Chain, sensors, abnormality detection, security, pattern detection**

I. Introduction

We live in an era of automation present in buildings, homes, and facilities, where even the tiniest things are inter-connected and communicate each other through network protocols. Recent poll results show that the number of internet connected devices will grow exponentially in the next year and it is estimated by 2020 to rich 50.1 billions [9]. These devices have various functionalists and capabilities such as face detection cameras for door lock/unlock and mechanism to open when the user is recognized. Or, if someone is coming to walk the dog at the same time each day we can program the home automation system to unlock the front door for them, and lock it up again when they are done. We can also check our security systems status, whether the lights are on, whether the doors are locked, what the current temperature of the home is and so on. And we even ask the intelligent virtual assistant turn off the lights, brew a coffee or turn on the car. These are closely located sensors that monitor the same space from and for different or identical aspects. They communicate with each other and have capability of making decisions based on the collective co-location in some capacity. We call them **co-located sensors**. Examples include (a) human body attached sensors that measure glucose level and accordingly adjust insulin, or (b) in home environments, motion sensors in the house that detect unexpected movements, lock the door and send emergency messages. Our current focus in this paper is on (physical) access-controlled spaces in the context where there is a mechanism that determines who enters that space and what privileges that person is allowed to perform in that space or interact with the devices.

A. Coordinated intelligence

In conventional individual sensors, data is received by each sensor from its own sensor stream and decisions are based on that data. In co-located sensors, a real-time processing unit receives data from multiple sensors and decisions are made based on the reasoning from all sensors. This shared intelligence that emerges from the collaboration, collective efforts, and competition of many individual sensors and appears in consensual decision making is called collective intelligence [3]. One of the types of collective intelligence is coordinated intelligence, because collective actions or tasks require different amounts of coordination depending on the complexity of the task [12]. The coordination enables the system to function in a much more intelligent way and make more timely and more precise decisions.

B. Contributions

Our research is focused on addressing the aforementioned problem of achieving collective co-located intelligence, with little or no *a priori* user-specified definition of normal and abnormal behaviors.

The contributions of this work are summarized as follows:

- We provide a definition of the coordinated intelligence approach.

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

- Security monitoring for co-located sensors problem is formulated as a problem of Markov Chain definition and analysis;
- We adapt the Markov Chain model to realize monitoring within and across sensors with the aim of automatically capturing normal behaviors as well as detecting abnormal behaviors;
- We present results from implementation of the analysis, and the simulation results verify effectiveness of the proposed approach.

C. Organization

The rest of the paper is organized as follows. The formulation of the colocated intelligence problem is provided in Section II. In Section III, our proposed solution including algorithm design, results using and discussion are presented. Related work is covered in Section IV, while Section V concludes the paper and identifies future work directions.

II. Problem Statement

A. Formulation

Co-located in a shared physical space, a set of sensors $S = \{S_1, S_2, .., S_n\}$ measure different parameters over time $T = \{t_1, .., t_k\}$, as illustrated in Figure 1. Each sensor is associated with an event stream that represents list of events generated by the sensor and each sensor has a finite number of possible discrete event outcomes $E_1, E_2, ..., E_l$. List of events for sensor S_1 would be such as $S_1 = \{E_{11}, .., E_{1l}\}$, where each event $E_{ij} = (v_{ij}, t_{ij})$ is associated with a value v_j at time t_j. Event in this context represent state of the sensor and each sensor has discrete states.

The objective is to (a) identify event patterns of behavior for each sensor and across sensors via automated process to generate a characterization of temporal patterns and relative orderings, and (b) detect and identify deviations from the characterized event behavioral patterns to signal abnormal behavior dynamically.

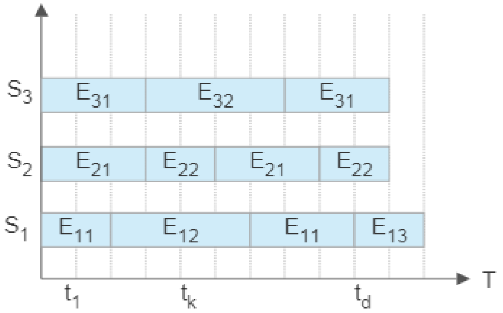

Fig. 1. Illustration of multi-sensor sensor streams

Sensor behavior patterns represent both within and across sensors: state changes can span over time within one sensor, and state changes across sensors over some time period. These are illustrated in Figure 2.

Consider the following example. In an office facility that enforces an authentication policy to enter the facility, there are three sensors that are measuring different parameters over time: (1) a sensor for checking authentication privileges if the individual is allowed to enter or not, (2) door lock sensor that unlocks if the user have the credentials to enter the office, and (3) a light sensor that activates when the user is passing by. This scenario is illustrated on Figure 3 (left side diagram), where sensors are continuously measuring and streaming the data readings.

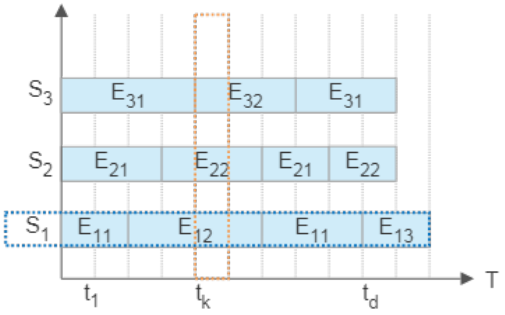

Fig. 2. Illustration of event behavior within sensor and across sensors

A list of constraints can be used when designing the solutions with respect to number of sensors and type of sensors and states.

B. Definition on normal and abnormal behavior

The challenge is to determine what is the normal behavior and, based on that, detect abnormal behavior. Both behaviors depend on situations where sensors are applied and used. To be applicable to a range of installations, we define the normal behavior in an application-agnostic way. A database of sensor data traces is taken as input that captures behavior and relations among sensors which are tagged as “normal behavior” by domain expert. For instance, the expected normal behavior of the sensors is that the door sensor is activated after the authentication is approved and the light sensor is then activated after the approved authentication and the door is unlocked, illustrated on Figure 3.

Abnormal behavior is defined as any violation or deviation from the normal. For example, with the badge sensor system of Figure 3, abnormal behavior includes:

1) sensor do not activate in the expected dependency order within each sensor stream; for example, “door open” should be followed by “door close”;
2) sensor stream behavior due to “unusual” user actions; for example a user scanning the access badge to enter, which turns the badge sensor turns “green”, and the door opens but the user decides not to enter the space;
3) multiple users enter the space without scanning the badge.
4) the light sensor is activated even before door was unlocked

In some situations, abnormal behavior may also arise from reasons that are not alarming per se. Some of the known

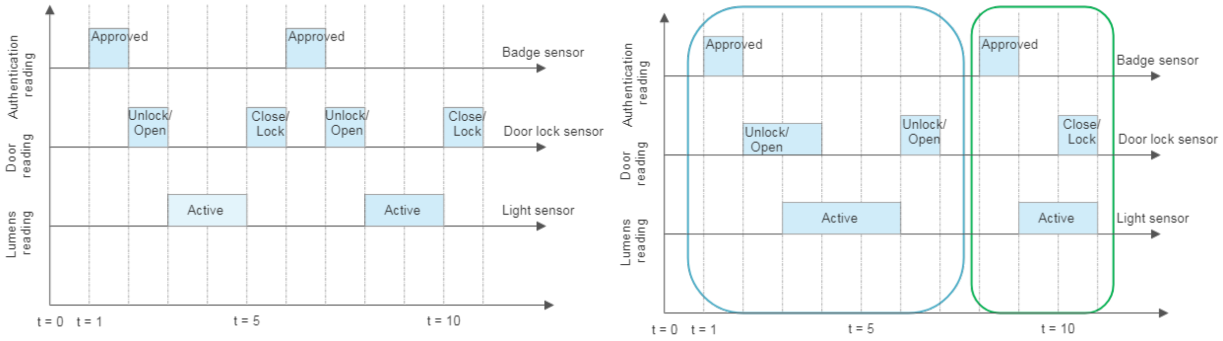

Fig. 3. Example of sensor data stream readings - normal patterns (left) and anomaly behavior (right)

reasons are battery leakage, device (e.g., light bulb) malfunctions, or losses in network connectivity, which are not necessarily indicators of abnormal behavior. This situation and other situations caused by transmission errors or node failure (or sophisticated attacks that use cross-layer fuzzy logic rules methods) are not considered in this work.

III. Our solution approach

A. *Algorithm development: Kensor*

Our solution approach is designed considering the following assumptions: 1) we know the number of sensors; 2) we know the finite set of discrete time interval action events; 3) user behavior is deterministic; With this in view, we develop an algorithm, named Kensor (for ”knowledge”-oriented smart sensor), based on a Markov Chain modeling approach that is not only domain-agnostic but also captures correlations of events among all the sensors. The set of all possible outcomes is considered the sample space. In our case, it is finite, from which each subset of a sample space is defined to be an event. We describe a sequence of possible events in which the probability of each event depends only on the state attained in the previous event. In case of the determining the event patterns across sensors Kensor approach is the create aggregated event state that is a representation of a snapshots of the states across sensors. For instance if we considerate time t_k aggregated event would be $[E_{12}, E_{22}, E_{32}]$ where $E_{12} = (v_{12}, t_{1k})$.

We represent this as a graph where nodes are events (sensor individual or aggregated) and edges are relations between events. The edge weights represent the total number of transition from one event to another in different time interval and these number of transitions as a fraction represents the transition probabilities.

We used visual evaluation metrics for event nodes, relations between events and transitional probability between them.

B. *Dataset*

We consider a scenario of an access-controlled environment with intelligent entry monitoring mechanism. This environment has three sensors to monitor an access area that includes a badge authentication device, a door locker and a light sensor. Badge sensor has three states they are *green*, *blue*, and *yellow*, the door sensor has two states *open* and *closed*, and light sensor has two states *on* and *off*.

The datasets are of two types:

1) Normal behavior is simulated using a Multinomial discrete probability distribution and domain expert knowledge to establish a normal controlled events and relations (see Figure 4), and
2) Abnormal activities are simulated dynamically using a set of rules that violates the normal behavior (see Figure 5).

Normality: The scenario for normal behavior of entering the access-controlled area is represented as follows. Suppose there are two events (t_1, E_1) and (t_2, E_2), where time $t_1 < t_2$ and it is normal for event E_1 to appear before event w_2. One example is shown in Figure 4, which represents the flow of events: when the badge reader is green, the door opens, and when the door opens, the person walks towards the room, and the light turns on. When the badge reader is yellow or red, the door is closed and light turns off after a timeout period.

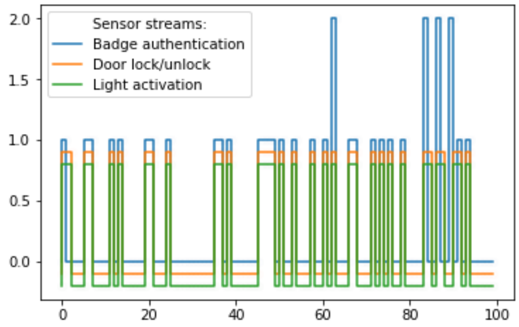

Fig. 4. Normal behavior of sensor streams

Abnormality: Abnormal behavior is created to introduce a number of unusual scenarios into the dataset. A set of rules that violate normal expected behavior is used. For example, one rule is the door is closed when the badge light is green and when the light is on, this is illustrated in Figure 5.

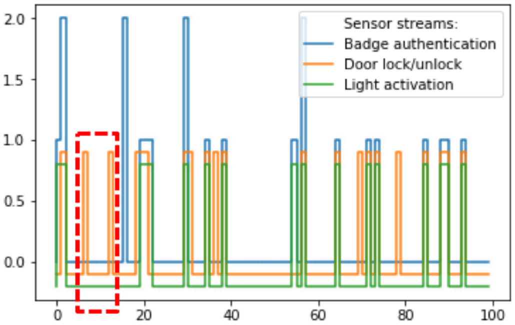

Fig. 5. Abnormal behavior of sensor streams. Marked with red is a scenario the door was open with out successful badge authentication and light is off.

C. Results

The access-controlled environment presented earlier has installation with three types of sensors: a badge authentication device, door lock/unlock mechanism, and light sensor activation. These sensors are operational all the time and generate event streams whenever any user action activates or interacts with them. We applied Kensor to the normal data set to represent the normal baseline behavior. Although, transitional probabilities can change over time it shows the number and type of expected events and relations between them, shown on Figure 6. For the same scenario, the normal behavior aggregated across sensors at a given time is shown on Figure 7. From both figures we can observe that the flow of event transition follows the normal previously defined scenario.

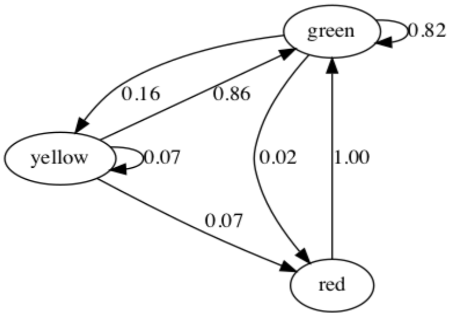

Fig. 6. Normal pattern behavior of single sensor badge authentication

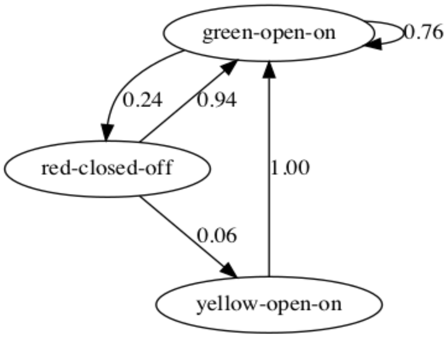

Fig. 7. Normal pattern behavior of aggregated multi-sensor badge reader, door and light activation aggregated states

To evaluate the proposed Kensor algorithm for capturing the abnormal behavior we performed the following analysis:

1) Comparison between normal graph and graph at time t_d;
2) Comparison over time between graph at time t_d and t_k;

The following abnormal behaviors are introduced to test detection:

- **Abnormality 1** The badge authentication recognizes the ID but the door does not open.
- **Abnormality 2** The door opens but lights do not turn on.
- **Abnormality 3** After positive authentication, the sensors does not turn to idle (yellow) state.
- **Abnormality 4** Authentication is positive, and the lights are on, but the door does not open.

We experiment with the aforementioned types of abnormal state changes. In a scenario with ten events of Abnormality 1, the graph of results is shown in Figure 8.

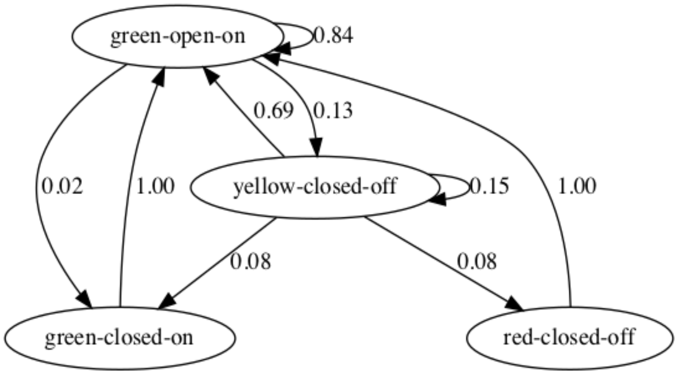

Fig. 8. Aggregated multi-sensor behavior with one abnormal state

Another scenario we have is by introducing Abnormalities 2, 3, and 4, the graph that presents this is shown on Figure 9.

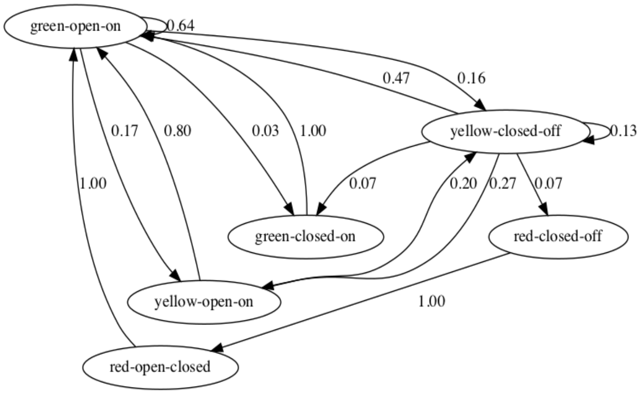

Fig. 9. Aggregated multi-sensor behavior with three abnormal states

If we take a screen shot at time t_d of the sensor states we should see a graph as presented on Figure 7 the normal behavior, while if it looks different than that something like the graph on Figure 8 we can notice that, there are new nodes and based on that to determine what happened and which previous state was the initiator and how often that new state transition occurred. We can go back or forward at time and observe the changes in state and relations between sensors. This is graph change over time is represented on Figure 10 and Figure 11.

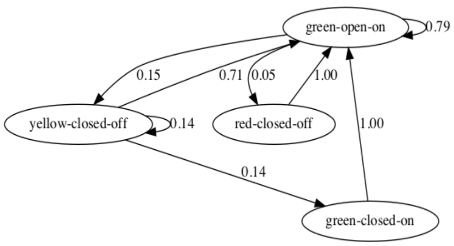

Fig. 10. Aggregated multi-sensor behavior for period of $[t_0 : t_{50}]$ time stamps

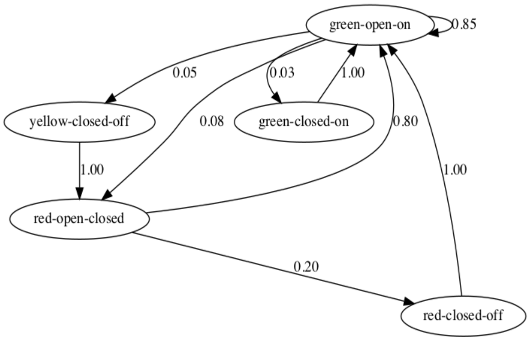

Fig. 11. Aggregated multi-sensor behavior for period of $[t_{51} : t_{100}]$ time stamps

To evaluate the solution we used transitional matrix comparison and visual graph inspection.

D. Discussion

With this approach we are tying to approach this problem in an application agnostic manner. The algorithm is applicable to similar problems of identifying abnormal behavior when given baseline of normal behavior. So for any given event data set we can represent graph of aggregated states and relations at certain time and over time period. But to determine the abnormalities is necessary to have baseline that represent the normal behavior. For our example we used controlled simulated behavior with rules to ensure the normality, but for real world cases and deployment we need data set such as building telemetry data already collect for LEED certification such as entry/exit, occupancy, activity type, etc. Abnormality detection and characterization has been tested at multi-sensor streams level. Our approach determines the difference between two Kensor graphs using the following modes: 1) Static mode of comparison - baseline with certain time; and 2) Dynamic streaming mode of detecting and tracking changes in the graph over time. Both of these approaches consider graph change detection such as node additions, node deletions, edge additions, edge deletions, edge weight changes as metrics for evaluating abnormal behavior compared with the known normal behavior. If a new node and edge is added or removed compared with the baseline, it is considered unexpected behavior. Similarly, changes to edge weights compared to the normal require specification of user-desired thresholds on tolerable versus abnormal deviation. We are planing to have statistical analysis to justify the evaluation metrics and show only the difference between graphs in the next development phase.

IV. Related work

Related work in the literature is reviewed here under two criteria: 1) co-located sensors and 2) multi-sensor pattern detection methods.

Identification of relations among sensor streams in home settings [4] has been studied using Allen's rules and association mining algorithms to identify temporal relations among sensors. Detection of frequent patterns by considering connectivity among sensors has been studied in IoT environments [1] in which they represented the problem as a graph; their method incrementally detects frequent sub-graph patterns by using frequent sub-graph pattern information generated in prior windows of a sliding window. An intelligent air quality system was developed [13] in indoor settings using the k-nearest neighbour classification algorithm. Anomaly detection in smart home settings using temporal relationship analyses and such as likelihood of an event occurrence given another event occurrence has been previously explored using probabilistic models for implementation was proposed by [5]. This problem has been addressed in personal health monitoring systems such as human activity detection for health monitoring that includes wearable sensors is presented by [6]. Also, identifying the abnormal behavior such as malware from normal was presented by [11]. Our approach is different from the preceding approaches in the sense that we consider event pattern changes not only within a single sensor stream but also across all sensors at a certain time stamp.

We also identify the possible solution approaches to solve this problem and organize them in categories as presented in Table I. Logic and probabilistic methods are based on expert knowledge for development. Data-based methods are driven by the data streams. While hybrid methods are a combination of both, they use data-driven and logic-based methods to identify patterns in the data. Our algorithm belongs to this category as we use a Markov Chain model which is a data-driven method, and we also use temporal logic-based rules for sensor event stream aggregation.

V. Conclusion and future work

In emerging cyber-physical systems, the numbers and types of sensors that are co-located in close geographical proximity offer opportunities to create increased amount of coordinated intelligence. Here, we focused on defining the problem of generating coordinated intelligence from co-located sensors, and applied it to an entry monitoring system in an access-controlled area. The problem is formulated as (1) converting "normal data" into assimilated patterns that are derived automatically, and (2) detecting and identifying deviations from the derived normal patterns, to be characterized as "abnormal

TABLE I
OVERVIEW OF SOLUTION APPROACHES

Types	**Methods**	**Techniques**	**Language and algorithms**
Logic-based	Probabilistic, Logical rules	Finite automata, Markov models	Python, Java
	Inductive-logic programming	First-order predicate calculus	PROGOL
	Complex Event Processing	Event calc. [7]	OnProM, ISEQ, Esper
	Temporal logic	Hierarchical lang. [8]	Allen', TSKR rules
Data-based	Association	Frequent/sequential pattern mining, Fuzzy rules [5]	Apriori, SPADE, PrefixSpan
	Relationship	Correlation, Causation [1]	Correlation, Granger
	Similarity	Clustering, Classification [2]	Naive Bayes, DBScan
Hybrid	Temporal logic and Association	Temporal relations between events and association [4]	Expert rules and Apriori algorithm
	Association and similarity	Temporal association rules and clustering [10]	Hierarchical clustering and association

events”. As one of the effective solutions, we used a Markov Chain-based modeling approach names Kensor to create a graph of aggregate state transitions that characterize normal operation, and identify deviations of the graphs of abnormal operation from the graphs of normal operation. Initial experiments have shown the ability to digest normal behaviors from sensor streams in an application-agnostic fashion, and then also detect aberrations in entry access behaviors in subsequent feeds of sensor streams.

Although multi-sensor fusion in the co-located context appears simple at first, it is clear that there are many complexities that arise when we consider more variants of abnormalities, such as temporal ordering, overlapping spans across events, and so on. In the future, the problem offers a range of additional directions to extend the work at the intersection of time series analysis, temporal logic, graph analytics and statistics.

ACKNOWLEDGMENT

Research sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U. S. Department of Energy.

REFERENCES

[1] Kyoungsoo Bok, Jaeyun Jeong, Dojin Choi, and Jaesoo Yoo. Detecting incremental frequent subgraph patterns in iot environments. *Sensors*, 18(11):4020, 2018.

[2] Stefan Dernbach, Barnan Das, Narayanan C Krishnan, Brian L Thomas, and Diane J Cook. Simple and complex activity recognition through smart phones. In *2012 Eighth International Conference on Intelligent Environments*, pages 214–221. IEEE, 2012.

[3] Wikipedia Collective intelligence.

[4] Vikramaditya Jakkula and Diane J Cook. Mining sensor data in smart environment for temporal activity prediction. *Poster session at the ACM SIGKDD, San Jose, CA*, 2007.

[5] Vikramaditya Jakkula and Diane J Cook. Anomaly detection using temporal data mining in a smart home environment. *Methods of information in medicine*, 47(01):70–75, 2008.

[6] Muhammad Usman Shahid Khan, Assad Abbas, Mazhar Ali, Muhammad Jawad, Samee U Khan, Keqin Li, and Albert Y Zomaya. On the correlation of sensor location and human activity recognition in body area networks (bans). *IEEE Systems Journal*, 12(1):82–91, 2018.

[7] Michael Körber, Nikolaus Glombiewski, Andreas Morgen, and Bernhard Seeger. Tpstream: low-latency and high-throughput temporal pattern matching on event streams. *Distributed and Parallel Databases*, pages 1–52, 2019.

[8] Fabian Mörchen. A better tool than allens relations for expressing temporal knowledge in interval data. In *Workshop on Temporal Data Mining at the Twelveth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 25–34, 2006.

[9] NCTA-Growth of IoT. Ncta growth in the iot, 2016.

[10] Sarah Osama, Marco Alfonse, and Abdel-Badeeh M Salem. Mining temporal patterns to discover inter-appliance associations using smart meter data. *Big Data and Cognitive Computing*, 3(2):20, 2019.

[11] Meltem Ozsoy, Caleb Donovick, Iakov Gorelik, Nael Abu-Ghazaleh, and Dmitry Ponomarev. Malware-aware processors: A framework for efficient online malware detection. In *2015 IEEE 21st International Symposium on High Performance Computer Architecture (HPCA)*, pages 651–661. IEEE, 2015.

[12] Anita Williams Woolley, Christopher F Chabris, Alex Pentland, Nada Hashmi, and Thomas W Malone. Evidence for a collective intelligence factor in the performance of human groups. *science*, 330(6004):686–688, 2010.

[13] Yujiao Wu, Taoping Liu, Sai Ho Ling, Jan Szymanski, Wentian Zhang, and Steven Weidong Su. Air quality monitoring for vulnerable groups in residential environments using a multiple hazard gas detector. *Sensors*, 19(2):362, 2019.