Hindawi
Security and Communication Networks
Volume 2018, Article ID 2476968, 12 pages
https://doi.org/10.1155/2018/2476968

Research Article

The Motif-Based Approach to the Analysis of the Employee Trajectories within Organization

Evgenia Novikova, Yana Bekeneva, and Andrey Shorov

Saint Petersburg State Electrotechnical University "LETI", Professora Popova Str. 5, Saint Petersburg, Russia

Correspondence should be addressed to Yana Bekeneva; yana.barc@mail.ru

Received 18 December 2017; Revised 10 April 2018; Accepted 15 April 2018; Published 20 May 2018

Academic Editor: Félix Gómez Mármol

The analysis of the employees' movement within organization building is an important task of the investigation of the business processes existing in the organization, including provision its cyberphysical security. In the paper, the motif-based approach to behavior pattern description and anomalies in organization staff movement is proposed. The *motif* of the employees' movement represents a combination of the spatial and temporal attributes of the movement enforced by attributes of the visited controlled zone. The usage of motifs enables transformation of the raw logs from the proximity sensors of the access control system containing only identifiers of the controlled zones into semantically meaningful list of the activities. This approach is demonstrated with an application to the 2016 VAST Mini-Challenge 2 data set, which describes movement of the employees within organization building.

1. Introduction

The Internet of Things (IoT) technologies provide enormous opportunities in the remote management of different objects starting with homes finishing with complex industrial objects. Industrial IoT unites computer networks and industrial objects equipped with built-in sensors and special data collecting and exchanging software and thus allows controlling critical objects such as nuclear plants, dams, and sewage treatment facilities without human in the loop. To avoid downtime and to provide critical object security, it is necessary to introduce technologies detecting and predicting cyberphysical risks. It is necessary to monitor functioning of the technological equipment and people implementing its maintenance. The analysis of the movement of the employees of critical infrastructure and hazardous industries assists in monitoring observance of the safety and access control policies and is useful in detection of insider threat [1, 2]. However, existing access control systems and employee activity monitoring systems are mostly targeted for employee productivity assessment by calculating working hours, estimating dynamics of employee lateness, and evaluating user activity on work place [3–5]. They do not allow revealing patterns of the employees' trajectories and establishing the possible activities without any prior information about existing business processes in the organization.

The paper presents an approach to the analysis of the employees' trajectories inside the organization extracted from the logs of the access control sensors. The underlying idea of the proposed approach is that employees make moves within organization because of some professional responsibilities such as going to their working place, visiting weekly meetings, or some individual needs such as going for lunch or café break during the work day. Thus, every move of the employee can be explained by some *motif* that is described by a combination of the temporal and spatial attributes of every move including specific features of the visited room, such as presence of the conference halls, offices, and refreshing facilities. Usage of the motifs helps to transform spatiotemporal data about employees' trajectories into a sequence of the higher level abstraction activities, enabling thus efficient and semantically meaningful analysis of their routes. The moves that could not be recognized as a motif-based movement are assumed to be anomalous. In the proposed approach, motifs of the employee movement are extracted from logs of the access

control proximity sensors and attributes of the controlled zone.

Specifically, the main contribution of the authors is an approach to the analysis of the movement of critical infrastructure staff that fuses trajectory data, attributes of the controlled zones, and employees' position in the organization.

The rest of the paper is organized as follows. Section 2 discusses the related work on approaches to patterns and anomaly detection in trajectories of moving objects. In Section 3, the authors describe the proposed approach to anomaly detection in employee movements. In Section 4, case study used to evaluate approach is presented, results are discussed, and directions of the future research are defined. Conclusion sums up the contributions.

2. Related Works

In the modern access control systems, the analysis of the employees' movement is performed by constructing their routes and heat maps of movement, calculating distances walked, and monitoring compliance of the routes to the travel regulations on the basis of the specified schedules defined for controlled areas or employee's role profile [2–4, 6–9]. In some cases, they allow identification of the frauds associated with a simultaneous registration of several employees or registration of missing colleagues [6]. The tasks of forming patterns of the employees' routes as well as construction of the existing business processes from the logs of the access control sensors are not solved. The detection of anomalies is done mainly on the basis of rules defined in accordance with the description of employee's role profile meaning that the analyst needs to have data on the existing business processes in order to form such rules or labeled data set describing all possible cases of the normal or abnormal personnel behavior to train automated analysis models. The latter is not a trivial task as in the major cases labeled data describing trajectories of the moving objects are not available.

Different data mining techniques are adopted for the analysis of the trajectories of the moving objects [10]. When an analyst does not possess labeled data describing normal and abnormal behavior, the most widely used approach is to apply clustering-based techniques to investigate trajectories [11]. The obtained clusters are then used to describe normal behavior of the moving objects, while outliers can be used to detect anomalies. The clusters may be found by centroid based approaches, hierarchical models, or density-based approaches [10–14]. In [15], authors construct a probability tree to mine movement patterns of the person movement inside the room. In [16], the approach to discover user daily activity patterns from GPS trajectories using association rules is proposed. The authors describe an algorithm recognizing visited place from stops and moves of GPS trajectories. The Apriori algorithm is used to extract user activity patterns.

In [17], the visual clustering algorithms are applied to identify visually and analyze areas/time periods with anomalous distributions of pedestrian flows. The contour maps are adopted to describe the distribution of pedestrian movement in terms of entry/exit areas. The visualization-driven approach to the movement analysis of the employees of critical infrastructure is presented in [18, 19]. It consists of two stages: finding groups of employees with similar behavior and detection of the anomalies. The groups of similarities are detected using Kohonen self-organizing map; spatiotemporal patterns of the behavior are presented using Gannt-based visualization technique; anomalies are assessed as deviations from detected cluster centroids. The similar approach to the visualization of the employees' trajectories is described in [20]. The Gannt-based visualization of the employees trajectories is supported by a 3D Building View which represents itself a 3D plan of building and shows employee trajectories throughout a day.

In [21], the authors present a method for searching for anomalies in the routes of vessel movements based on the analysis of motifs. The process includes the following stages: representation of the trajectories in the form of motifs, transformation of the source low-level data space into a high level feature space of motifs, and classification of the routes on the base of the motifs. Similar approach is implemented in [22]; the motif-based anomaly detection is supported by usage of predefined rules.

In this paper, the authors propose an approach based on principles similar to the principles proposed in [21, 22]. However, [21, 22] are focused on analysis of the vessels trajectories, and thus subject domain differs significantly from one investigated in this paper, requiring elaboration of another concept of the *motif*. In [18–20] authors study similar problem; however, the proposed approaches construct visual behavior patterns in terms of visited zones, while approach proposed in the paper allows abstracting from these low-level data and present behavior model in terms of high level motif-based activities, such as meeting with colleagues, working in office, and going for lunch. Moreover the proposed in the paper approach allows forming formal description of the employee's activity suitable for training automated analysis models.

3. The Approach Description

In general case the logs from proximity card readers describing employees' movement have the following format:

$$\langle \text{timestamp, employee ID, controlled zone ID, additional data} \rangle \quad (1)$$

Additional fields may describe status (entrance permitted or denied), employee access level, and so forth. The logs could be complemented by the control zone plan, including description of the rooms located within a controlled zone, employees' position within organization hierarchy, and the location of the employees' work place within the controlled zones. In the proposed approach, the authors fuse data from controlled zone plan, location of the employee's work office, his/her position in the organization, and logs from the proximity card readers to extract *motifs* of the employees' movement. Motifs are used to explain the goal of every move of the employee: going to the working place in the morning or

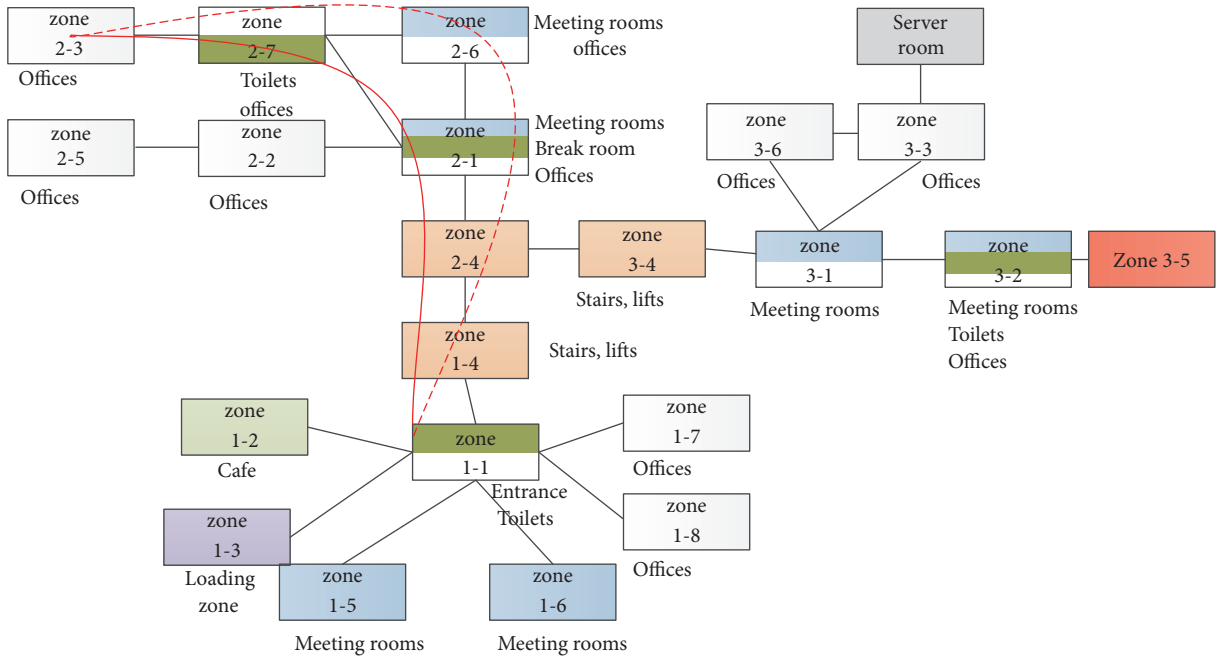

Figure 1: Graph of the controlled zones.

leaving building during lunch time, visiting weekly meetings with colleagues, and so on.

3.1. The Motif Features Description. In the proposed approach, *a motif* represents a combination of the spatial and temporal attributes of the staff movement enforced by attributes of the visited controlled zone.

The spatial attributes of the motif are constructed on the basis of the controlled zones visited by an employee and their attributes. The plan of the controlled zones allows constructing graph of the adjacent zones. It is considered that the route of the employee between two nonadjacent controlled zones would consist of the zones that constitute the shortest path between selected zones, as people tend to use the shortest way to reach their destination. Thus, irregular visit of the zone not included in the shortest path could be considered abnormal. Figure 1 demonstrates two possible routes of the employee from building entrance to zone 2-3, where his/her office is located, the one marked by the solid line is the shortest one and represents a sequence of the most frequently met zones in his/her route, while the second one marked by dotted line is not the shortest path and zone 2-6 is never met in his trajectory.

The attributes of the controlled zones may vary from the organization to organization. Some organization premises may contain only work offices, rooms with food, or refreshing facilities. Some buildings may have zones with specific equipment such as servers and machines. All controlled zones could be separated in the sets of zones having similar attributes. The authors also suggest using location of the employee's working place as extra attribute of the zone. In general case it is possible to form following sets of controlled zones:

Z_j^w is a set of the controlled zones where working places of the employees of j department are located.

$z_{ij}^w \in Z_j^w$ is a zone where working place of the i employee is located.

$z_{ij}^r \in Z^r$ is the closest controlled zone with refreshing facilities to the working place of the employee, where Z^r is a set of all controlled zones equipped with refreshing facilities.

$z_{ij}^f \in Z^f$ is the closest controlled zone with food services to the working place of the employee, where Z^f is a set of all controlled zones equipped with food facilities.

$z_{ij}^s \in Z^s$ is the closest controlled zone with stairs and lifts to the working place of the employee, where Z^s is a set of all controlled zones equipped with stairs and lifts.

Figure 1 shows zones with their attributes given in the labels. A certain problem arises when the zone has several attributes meaning that it is necessary to select possible goal of the visit assessing mean duration of the employee's staying within it.

To determine the temporal characteristics of the motif, the timestamp in the source log is transformed in the following way:

$$\text{timestamp} \longrightarrow (\text{wd}, \text{time\_offset}), \quad (2)$$

where wd stands for the day of the week, that is, Monday, Tuesday, and Wednesday, represented by a number; time_offset is a time offset from the beginning of the day

Table 1: Example of the raw proximity sensors logs.

Timestamp	Employee id	Zone ID
2016-05-31 00:05:00	vawelon001	1-1
2016-05-31 00:20:00	earpa001	1-1
2016-05-31 02:26:40	earpa001	1-6
2016-05-31 02:31:41	vawelon001	1-6
2016-05-31 03:12:00	earpa001	1-1

Table 2: Example of a journal of the proximity sensor logs after the transformation.

wd	time_offset (sec)	t_dur (sec)	employee_id	Zone ID
3	300	8801	vawelon001	1-1
3	1200	7600	earpa001	1-1
3	8800	2720	earpa001	1-6
3	9101	3379	vawelon001	1-6
3	11520	3680	earpa001	1-1

in seconds. The duration of staying within controlled zone wraps up the temporal description of the motif.

Summing up all mentioned above, it is possible to describe motif in the following way:

$$\text{Motif} = \left\{ f_i, \text{dep\_id}, \left\{ \text{wd}_j, \text{time\_offset}_j, t\text{\_dur}_j \right\}_{j=1}^{n} \right\}, \quad (3)$$

where f_i is the zone attribute; dep_id is employees' department; wd_j is day of week; time_offset_j is time offset from the beginning of the day in seconds; $t\text{\_dur}_j$ is duration of staying the zone with given attribute f_i. As the authors assume that the motif may have temporal periodicity, a set of temporal constraints describing when the given motif takes place is defined. For example, an employee can visit meetings held within his/her department every Monday and Wednesday in the morning at 10.30.

3.2. The Motif Features Extraction. Obviously, the employees make scheduled activities at different moment of time; for example, they may come to work at 7.55 am, at 7.58 am, or even at 7.43 am. They can have lunch for 30 minutes or 47 minutes. When extracting patterns, it is necessary to consider possible variations in the temporal attributes of the moves in order to produce more generalized description of their motifs. The possible solution of the problem is to cluster raw features of the employee trajectories and to use cluster centroids to as motif description. The obtained motif descriptions can be applied further to classify logs of the proximity sensors as motifs.

In order to perform this procedure, raw proximity logs are transformed into the following format:

$$\begin{aligned} &\langle \text{timestamp, employee ID, controlled zone ID} \rangle \\ &\quad \longrightarrow \langle (\text{wd, time\_offset}), t\text{\_dur, employee ID,} \\ &\quad \text{controlled zone ID} \rangle . \end{aligned} \quad (4)$$

Tables 1 and 2 show examples of the raw proximity logs and their view after transformation, correspondingly.

After the raw logs transformation, they are grouped by the employee department; this is done because the motifs may vary from department to department. Then the authors assign each zone a vector of possible attributes. To do this, a set of attributes is presented as Boolean vector. If the zone has an attribute, the corresponding flag is set to true; otherwise it is set to false. The next preprocessing step consists in grouping transformed logs from the proximity sensor according to the value of the Boolean vector of the zone attributes. This allows extracting clusters from the set of logs generated by the proximity sensors of the zones with similar characteristics for the employees belonging to one department.

The data preprocessing steps including grouping logs by department and zone attributes allow working only with numeric vectors. Thus different distance metrics could be applied to extract clusters centroids that can be used as motifs description. In the approach the authors use simple Euclidian distance. Firstly, the hierarchical clustering technique is applied to detect number of possible clusters and then the flattening operator is applied to extract centroids of the clusters. This approach produces mean values for temporal parameters such as time of the visit of the given controlled zone and duration of staying in it, that is, time_offset parameter and t_dur parameter, correspondingly. However, to obtain generalized description of the motif, more relaxed constraints for these parameters are needed. To meet this requirement, the authors use samples of the proximity logs belonging to one cluster to calculate confidence interval for each parameter.

The semantic meaning of the motif is derived from the attributes of the controlled zone. If the controlled zone has one attribute only, it can be used as a basic name of the motif; in the case the controlled zone has several attributes, an analyst can choose the one most suitable for extracted motif. The motifs semantically close to each other can be united by joining temporal attributes of the clusters' centroids. Examples of the cluster centroids and corresponding attribute zones and possible motif names are shown in Table 3.

3.3. The Motif-Based Analysis of the Employee Trajectories. The usage of the motifs enables transforming raw logs from the proximity sensors containing only identifiers of the controlled zones into semantically meaningful list of the activities. This is done by labeling raw data with extracted motifs. The logs with similar motifs are united into one log, preserving the timestamp of the first log with given motif. The scheme of the process is shown as follows:

$$\begin{aligned} &\langle \text{timestamp, employee ID, controlled zone ID} \rangle \\ &\quad \longrightarrow \langle \text{timestamp, employee ID, motif ID} \rangle . \end{aligned} \quad (5)$$

The transformed logs from proximity sensors equipped with motif labels can be used to model behavior of the employees belonging to one department. Currently the Alpha algorithm is applied to form the Petri net N describing employee behavior model. It is defined as follows: $N = \{P, T, F\}$, where P is the set of positions, T is the set of

Table 3: Examples of the motifs extracted from proximity sensors logs and additional data about controlled zones.

Motif name	Zone attributes	Department ID	Weekday	Time offset	Duration of staying (secs)
Working on place 1	Zones with work places	Security department	0-4	30207 ± 427	14680 ± 1380
Working on place 2	Zones with work places	Security department	0-4	33257 ± 527	9727 ± 587
Working on place 3	Zones with work places	Security department	0-4	30207 ± 427	4680 ± 360
Meeting with colleagues	Zones with colleague offices	Security department	0, 2	30600 ± 347	1560 ± 307
Going out to refresh	Zones with coffee machines and toilets	Security department	0-4	37525 ± 781 55754 ± 639	810 ± 127
Passing through zone	Zones with lifts, stairs, and zones included in the shortest path to the working place	Security department	0-4	28300 ± 53 43622 ± 101 46455 ± 74 61940 ± 132	148 ± 63
Entering/exiting the building	Zone with entrance/exit	Security department	0-4	28200 ± 61 43322 ± 59 46505 ± 54 61440 ± 123	148 ± 52

transitions (motifs), and F is the set of arcs joining positions and transitions. Figure 2 presents a Petri net constructed for the raw logs from proximity sensors (Figure 2(a)) and Petri net constructed for transformed logs marked with motif labels (Figure 2(b)). It is clearly seen that the second Petri net is more obvious, and it can be rather easily understood by an analyst especially in the case when role profile of the employees is not clear.

Thus, the proposed approach allows identification of the common behavior patterns existing within department, for example, walking to work place, rest areas, and attending meetings, on one hand; on the other hand, it can be used for detecting possible deviations in employees movement that can be considered as anomalous ones. These deviations need further analysis as they can be signs of the rare motif or signs of some suspicious activity.

4. Experiments and Effectiveness Evaluation

To evaluate the approach, the authors use a dataset provided within the VAST Challenge 2016: Mini-Challenge 2 [23]. It contains logs of the proximity card readers that cover separate zones within building. When an employee with proximity card enters a controlled zone, his/her card is detected and recorded. The dataset contains a two-week set of logs. The logs from the proximity sensors are supported by building layout of the offices, cafes, lifts, stairs, and controlled zones. The list of employees gives information about their department and office assignments.

According to the proposed approach, the raw logs were transformed to the following format $\langle$(wd, time_offset), t_dur, employee ID, controlled zone ID$\rangle$ firstly and then were divided into sets grouped by department. The set of attributes of the controlled zones is presented as Boolean vector. The following attributes of the controlled zones for the given department were defined: (1) presence of the offices where employees of the same department reside; (2) presence of the offices where employees of other department reside; (3) presence of the employee working place; (4) presence of the cafes; (5) presence of the toilets; (6) presence of lifts or stairs; (7) presence of the meeting rooms. Then the logs were grouped by vectors of attribute and clustered. In the next step, the confidence interval for each temporal attribute of the extracted cluster centroid was calculated and possible motif name was assigned according to the values of the controlled zone attributes.

Let us consider the motif-based analysis of the employees' movement for one department in detail.

The typical day of the Engineers Department starts around 8 am. The working day for the majority of employees ends approximately at 4-5 pm; a small part of the engineers starts their working day approximately at 4-5 pm and finishes it approximately at 11 pm-12 am. Figure 3 gives an overview on the typical day of the engineers; their routes are shown using visual model proposed in [20]. The employee's route is presented as a sequence of colored segments, whose length corresponds to staying in the particular zone and the color denotes the identifier of the controlled zone.

Table 4 shows clusters extracted for zones equipped with stairs and lifts (zone 4 is located on the 1st floor and 2nd floor). Going upstairs to the second floor where their workplaces are located they visit firstly zone 1-4 on the first floor and zone 2-4 on the second floor; going downstairs they visit firstly zone 2-4 and then 1-4, correspondingly. Interestingly, the duration of staying in the zone depends on what zone is visited firstly. Thus going upstairs they spend approximately 40 seconds in

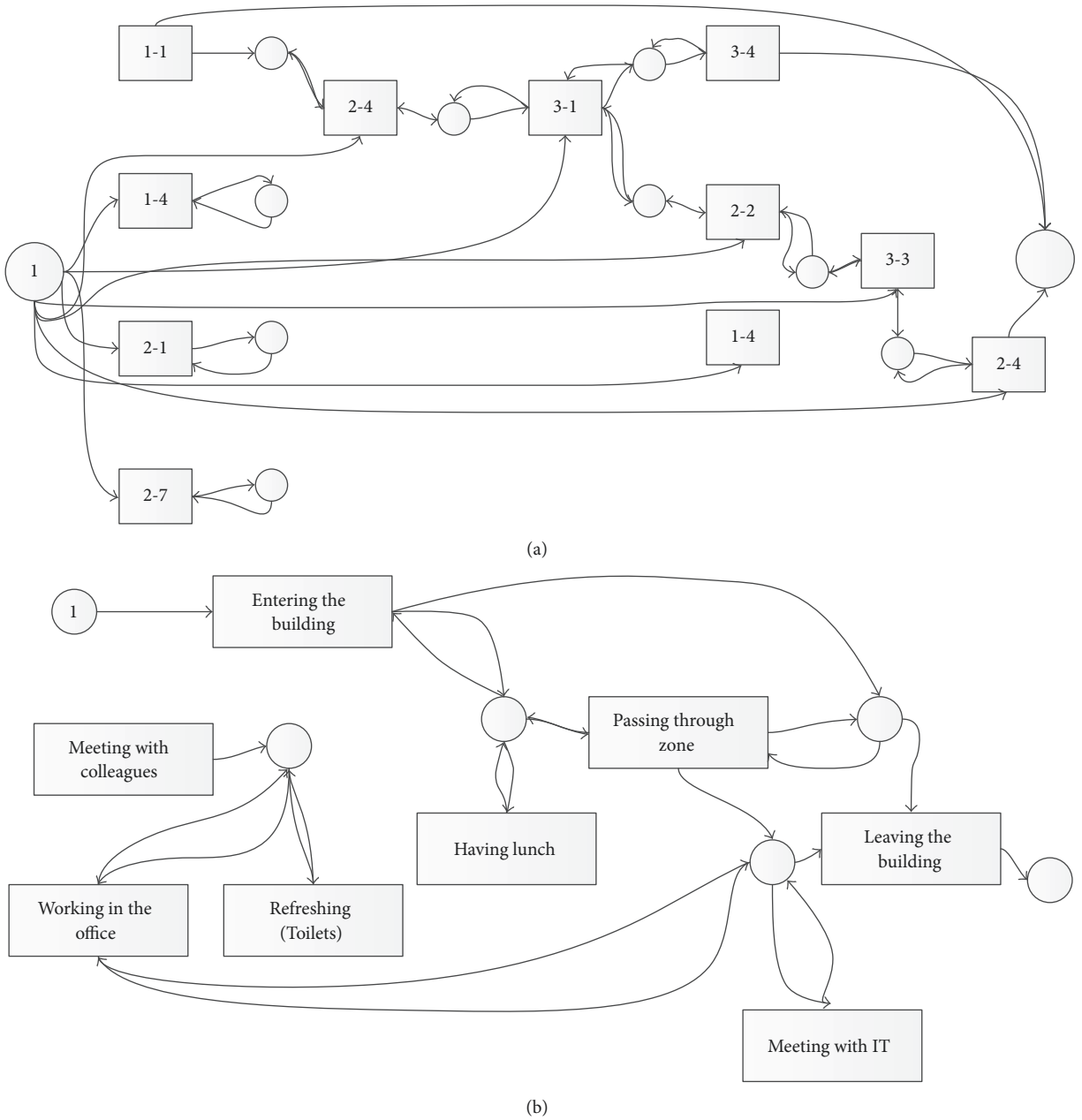

Figure 2: The Petri Net illustrating behavior model constructed from the raw logs (a) and from motif-labeled logs (b) of the proximity sensors.

zone 1-4 and only 2 seconds in zone 2-4, and, vice versa, going downstairs they spend 40 seconds in zone 2-4 and 2 seconds in zone 1-4.

The majority of the engineers start their working day around 7-8 am. Each employee first passes zone 1-4, located on the 1st floor, then zone 2-4, located on the 2nd floor, then he/she gets to their workplace. Obviously, the successive passage of these zones at the specified time can be combined into one motive, "coming to work." Similarly, all movements of employees in the period from 12 to 13 pm are associated with a lunch break and visit of a café (Figure 4). Thus, the successive passage of zones 1-4 and 2-4 or vice versa as well as through passage of certain zones can be combined into one motive. The authors obtained also four clusters of the motives that correspond to the late coming and leaving working place in the second part of the day. These results are fully consistent with the fact that there is a small group of engineers whose working day begins at 4 pm and ends around midnight. They have day routine very similar to the routine of the first group of the employees; the major distinction is that their activities are shifted relative activities of the first group of engineers. Thus, they can be united with the previously defined motives by adding extra temporal condition defining the possibility of the activity in the second part of the day. In

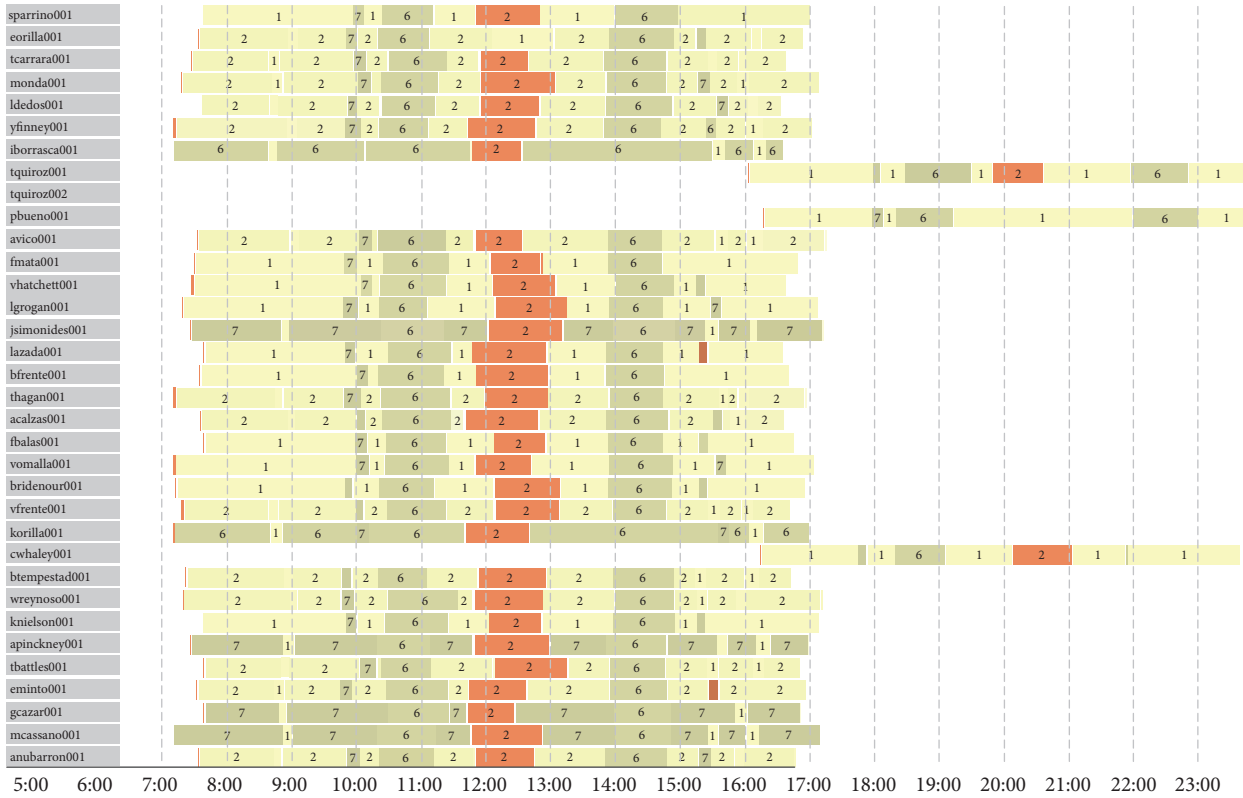

Figure 3: Typical day of Engineering Department.

zones 1-4 and 2-4, one anomaly in the engineers' movement was identified (Table 4). Figure 4 shows distribution of the detected clusters of the moves in the zones containing lifts and stairs and corresponding motives assigned to them; the detected anomaly was filtered out to make visualization more clear. The zones visited firstly on the way to the destination zone form the column on the right side of the figure; the zones visited after form the column on the left side of the figure.

The similar analysis was implemented for all zones having other attributes (workplaces, cafes, toilets, etc.). The workplaces of the employees of the Engineering Department are located in different controlled zones. Each employee spends most of the time at his/her workplace, but during the day they make moves when attending meetings, visiting colleagues, cafes, or toilets. The workplaces of some employees are located in the same zones with meeting rooms and rooms with coffee machines. Therefore, the staying duration of an employee in the zone may vary depending on whether his/her working place is located there. Let us consider the example of zone 2-1 which is usually passed through by engineers; however, it hosts workplaces of some employees, meeting halls, and break room as well. The detected motives of visits of this zone are shown in Figure 5. The moves associated with the motif "passing through" are highlighted by black circles; they all have similar duration, while moves associated with other motives such as "meetings" and "visiting coffee breaks" have different durations. A group of anomalous moves is also clearly seen in Figure 5.

Figure 6 shows behavior models of the employees of the Engineers Department presented using Petri net. The first one is constructed on the basis of the raw logs of the proximity sensors (Figure 6(a)) and the second one is constructed on the basis of transformed logs with assigned motives (Figure 6(b)). The moves with motives "going through" and "going to ..." with name of the destination zone are omitted in order to obtain more clear understanding of the employee's activities. Working with raw logs, the analyst only can see a set of zones visited by employees and transitions between them. It is hard to understand the goal of these moves and, therefore, to classify them as benign or suspicious one. Moreover, showing all zones visited by the employee during the work day can make Petri net unreadable for analyst, presenting information on zones passed by the employee on the way to destination zone. The Petri net constructed on the basis of the transformed logs is more clear as it contains activities of the higher level. Such graph is compact and easy to understand and analyze.

After mapping motives to the logs, six records with unrecognized motives were identified (Table 5).

All anomalies identified for this department are related to the atypical duration of staying of one employee in the areas typical for staff members of this department.

To evaluate the approach accuracy, the authors used the results of the VAST Mini-Challenge 2 provided by the organizers [23]. According to the data provided, the data set

Table 4: Examples of the motifs extracted for Engineering Department and zone with stairs and lifts.

List of zones with chosen attribute	Number of clusters	Average duration	Average time shift	Motif
1-4, 2-4	1	40496 ± 1968 (40 s)	27064 ± 364 (approx. 7,5 am)	Coming to work (1-4)
1-4, 2-4	2	2000 ± 129 (2 s)	27104 ± 213 (7,52 h)	Coming to work (2-4)
1-4, 2-4	3	40851 ± 2019 (40 s)	43217 ± 912 (12 h)	Going to lunch (2-4)
1-4, 2-4	4	2000 ± 154 (2 s)	43249 ± 847 (12,01 h)	Going to lunch (1-4)
1-4, 2-4	5	41208 ± 1928 (41 s)	46012 ± 483 (13 h)	Returning from lunch (1-4)
1-4, 2-4	6	2000 ± 124 (2 s)	46672 ± 763 (13 h)	Returning from lunch (2-4)
1-4, 2-4	7	40985 ± 1843 (40 s)	57602 ± 244 (16 h)	Coming to work (late) (1-4)
1-4, 2-4	8	2000 ± 103 (2 s)	57645 ± 251 (16 h)	Coming to work (late) (2-4)
1-4, 2-4	9	33388 ± 1211 (33 s)	61084 ± 253 (17 h)	Going home (2-4)
1-4, 2-4	10	2000 ± 112 (2 s)	61332 ± 294 (17 h)	Going home (1-4)
1-4, 2-4	11	41250 ± 1043 (41 s)	72291 ± 358 (20 h)	Going to lunch (2-4)
1-4, 2-4	12	2000 ± 107 (2 s)	72333 ± 411 (20 h)	Going to lunch (1-4)
1-4, 2-4	13	41000 ± 985 (41 s)	75700 ± 683 (21 h)	Returning from lunch (1-4)
1-4, 2-4	14	2000 ± 103 (2 s)	76363 ± 541 (21 h)	Returning from lunch (2-4)
1-4, 2-4	15	41130 ± 1211 (41 s)	85818 ± 784 (24 h)	Going home (2-4)
1-4, 2-4	16	2000 ± 95 (2 s)	85860 ± 241 (24 h)	Going home (1-4)
1-4, 2-4	17	58889000	86371 (23,85 h)	Anomaly

Table 5: Anomalies for Engineering Department.

ID of employee	Zone	Day of week	Time shift	Duration
cwhaley001	2-4	3	86371	58889000
cwhaley001	2-1	4	82685	62455000
cwhaley001	2-1	3	82218	63042000
cwhaley001	2-1	5	82085	63175000
cwhaley001	2-1	4	81334	64046000
cwhaley001	2-1	6	82158	235422000

contained the following types of the anomalies concerning usage of the proximity card:

(i) Loss of a proximity card by an employee, which is expressed in the long-term presence of a proximity card in the same zone

(ii) Usage loss of a proximity card by an employee, which is characterized by simultaneous usage of two proximity cards after issuing a new one

(iii) Employees visiting atypical areas.

Each anomaly is described by a set of records generated by proximity sensors; the set capacity depends on type of the anomaly and time of the work day when it occurred. For example, if employee forgets to use the proximity card at the end of the day, the anomaly is described by one unusual record; if the employee visits atypical zone, the number of the corresponding anomalous logs depends on the number of visited zones on the way to the given atypical zone.

After mapping motifs to logs, 211 unrecognized logs for all employees from all sample containing 29763 records were detected. These logs describe all abnormal situations announced by the organizers of the contest. However, the authors detected cases when the employee stays atypical time in the zone typical for him/her. The atypical duration may be longer or shorter than typical one. These anomalies also include cases when the employee does not visit zone he/she usually visits. It should be noted that though these deviations in the staff routes were not considered abnormal within given contest, they still can be signs of the suspicious behavior. The majority of the false positives were registered for employees belonging to the *Administrative* and *Executive* departments

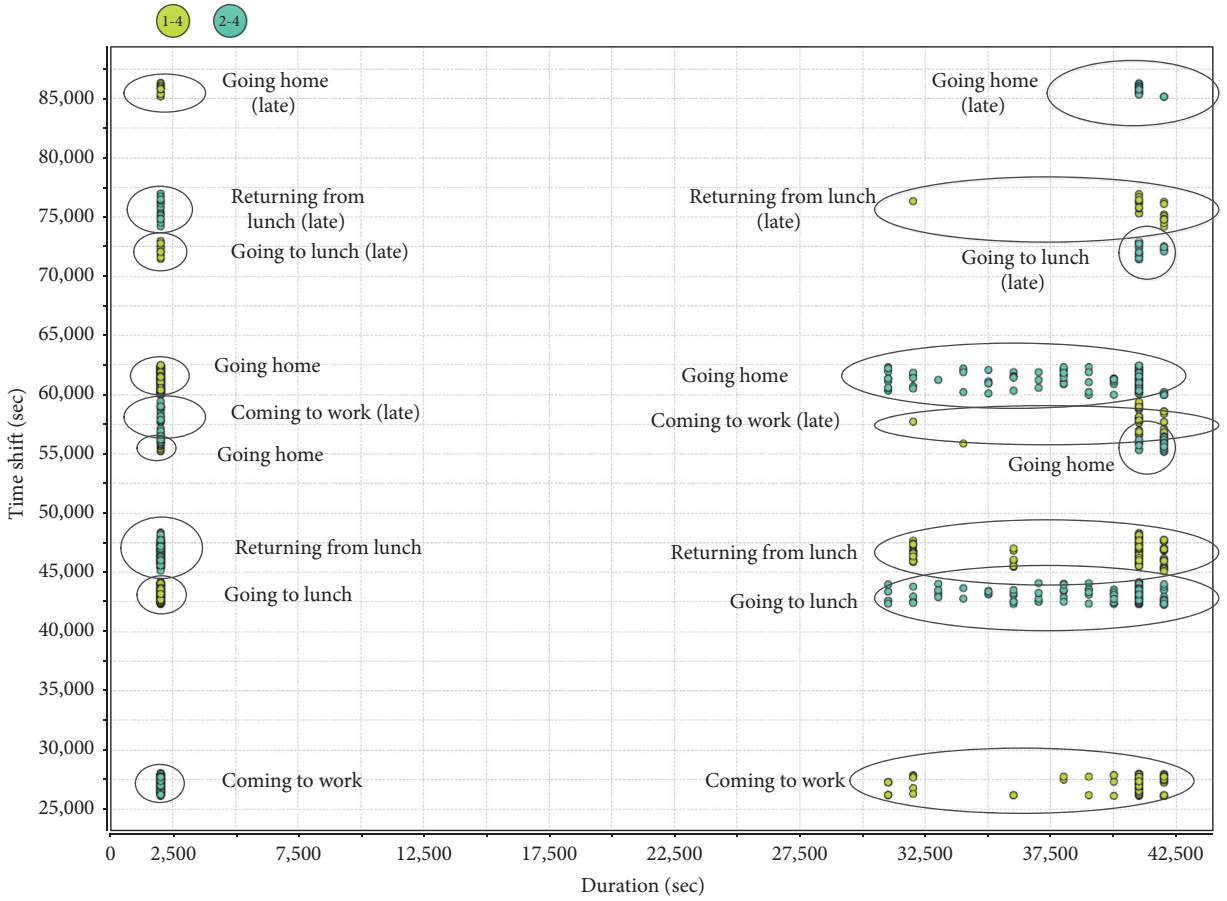

Figure 4: Clusters and corresponding motives of engineers movements in the zone containing stairs and lifts.

as these employees move rather diversely. It is also should be noted that it was rather difficult to determine possible motives of their movement. There were almost no false positives for the employees belonging to the *Security*, *Engineering*, and *IT* departments. This fact is explained by that their moves have a certain structure perhaps due to existing job routine.

To summarize the results of the efficiency evaluation, the following accuracy metrics were calculated:

$$\begin{aligned} \text{Precision} &= \frac{\text{TP}}{(\text{TP} + \text{FP})} = 0.87 \\ \text{Recall} &= \frac{\text{TP}}{(\text{TP} + \text{FN})} = 0.84 \\ F\text{-measure} &= 2 * \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} = 0.85. \end{aligned} \tag{6}$$

Thus, it is possible to conclude that the approach allows detection of the anomalous deviations in the employees' movement if they have a specific job routine. In this case, even data set containing two-week logs from proximity sensors is enough to reveal existing routine. Otherwise, it is necessary to have more data describing longer period of time with employee activity to reveal more sophisticated patterns in staff movement. Due to approach's core idea, if the actions are performed rarely, they are assigned to the separate cluster and the analyst defines whether they have a certain motif or they are suspicious ones. To be more precise, the analyst should possess other data specific to a particular organization. Such data may include a schedule of work shifts, holidays, birthdays, and so forth. A joint analysis of this type of data will allow the analyst or system to identify the motifs more accurately and, thus, reduce the number of suspicious movements.

It should be noted that the detailed description of the motifs given in the article is made for a understandable illustration of the principles of the developed approach. With a real application, a less detailed classification of movements is possible in order to follow the ethical norms and rules adopted in the organization.

To the best of the authors' knowledge, only few papers present results of the researches devoted to the analysis of the employees' routes inside the organization building [18–20]. The approaches presented in these papers are visualization-driven approaches; different visualization techniques are proposed to use to analyze proximity switch logs. The use cases presented in the papers show what kind of anomalies could be detected using them. However, these papers are lacking information on efficiency of the proposed techniques and tools implementing proposed approaches are not publicly

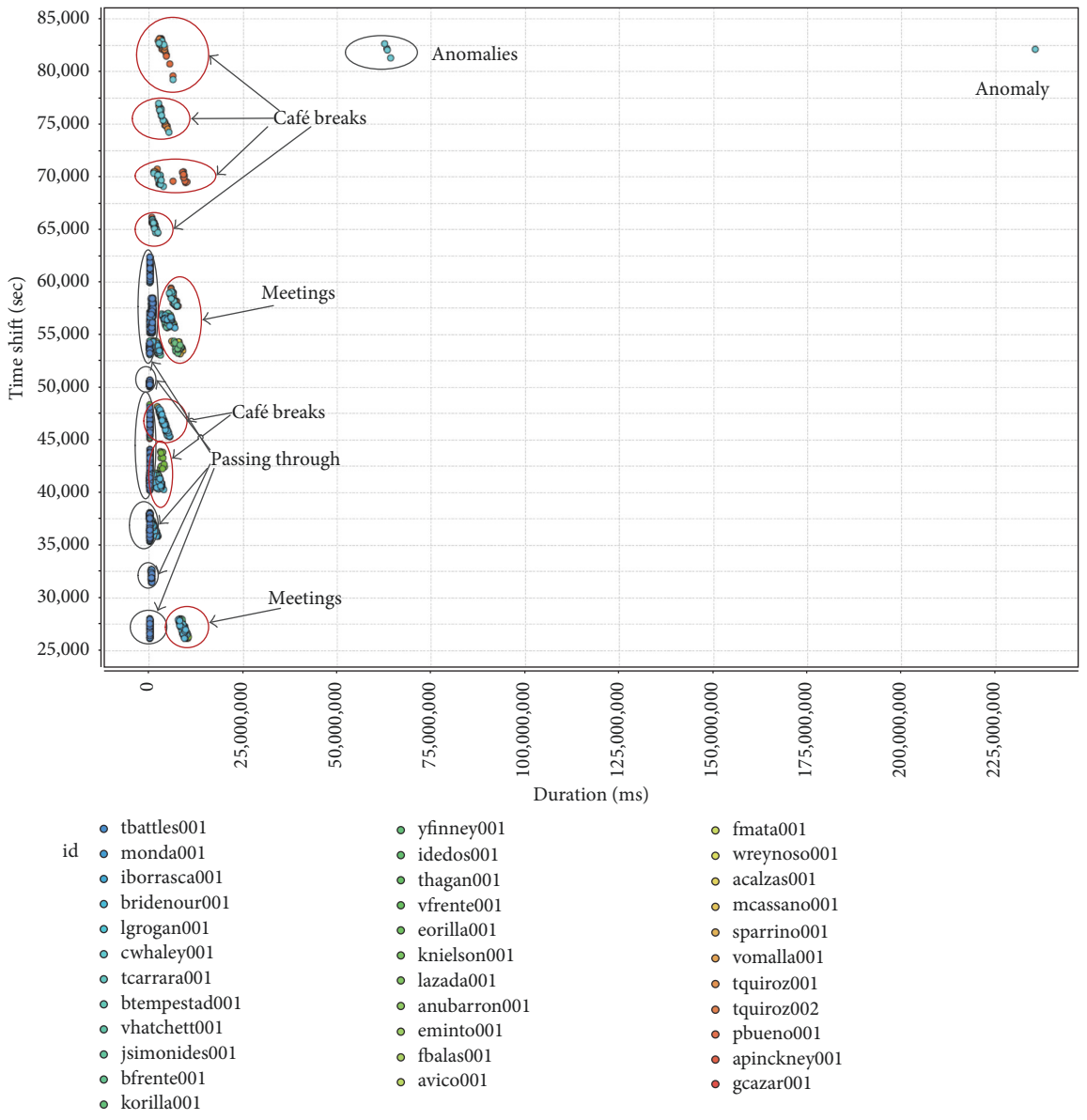

Figure 5: Movements of engineers in the zone containing working places, meeting halls, and break rooms.

available, thus making comparison of the approaches impossible. Techniques suggested in [21, 22] could be considered rather similar as they also operate with the *motif* notion, but as this notion is defined in the different subject domain, movement of the vessels, it is hard to include them in the evaluation process as they would require significant adoption to the data used in the approach suggested in the article.

According to the results of the evaluation process the authors defined the following directions of the future research work. One of the primary tasks of the future work is to test approach against real life data sets as real life data is usually noisy and messy; testing approach on different data sets allows assessing its robustness. Another important direction of the future research work is the construction of automated analysis model for detection of the anomalies in the real-time mode. The IoT technology allows distribution of the classification task between IoT devices making it possible to react on abnormal action almost immediately.

5. Conclusions

In the paper, the motif-based approach to the analysis of the movements of critical infrastructure staff that allows construction of the employees' behavior model and detection of the anomalies in their moves is proposed. The main feature of the approach is that it fuses trajectory data, attributes of the controlled zones, and employees' position in the organization to detect possible motives of the staff movement. Application of the motives allows an analyst to transform proximity sensor logs into a sequence of the higher abstraction level activities. These activities are used to construct behavior model of the employees. The moves that could not be

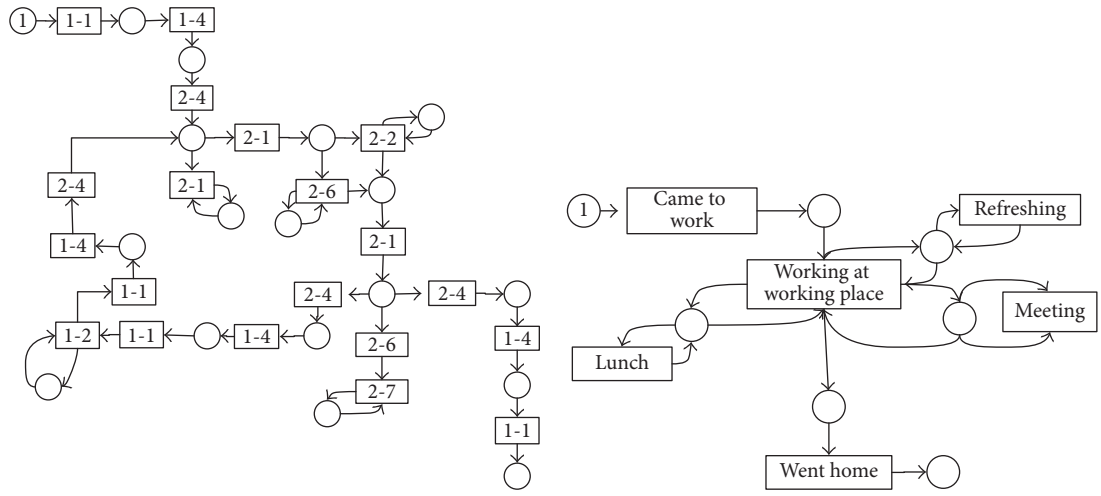

Figure 6: The Petri Net illustrating behavior model constructed from the raw logs (a) and from motif-labeled logs (b) of the proximity sensors for Engineering Department.

recognized as a motif-based movement are assumed to be anomalous. To illustrate the proposed approach and assess its efficiency, the data set provided by the VAST Challenge 2016 is used. The results obtained showed that it could be effectively used in the analysis process of the data describing movement of the employees whose role profiles are unknown as it enables constructing behavior model that is easy to interpret. Moreover it allows detecting possible deviations in the trajectories of the employees that could be sign of the potential fraud.

Further research will be devoted to the implementation of the automated classification model distributed between IoT devices, thus allowing to reduce computational load on the central control unit and produce reaction on abnormal behavior timely.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Ministry of Education and Science of the Russian Federation in the framework of the state order "Organization of Scientific Research," Task no. 2.6113.2017/6.7 and grant of the RFBR no. 16-07-00625.

References

[1] R. Rawassizadeh, E. Momeni, C. Dobbins, J. Gharibshah, and M. Pazzani, "Scalable Daily Human Behavioral Pattern Mining from Multivariate Temporal Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 11, pp. 3098–3112, 2016.

[2] X. Pan, C. Han, K. Dauber, and K. Law, "A multi-agent based framework for the simulation of human and social behaviors during emergency evacuations," *AI Society*, vol. 22, pp. 113–132, 2007.

[3] "Mircom access control," 2017, https://www.mircom.com/security.

[4] "Lenel access control," 2017, http://www.lenel.com/solutions/access-control.

[5] "Avigilon access control," 2017, http://avigilon.com/ru-ru/products/access-control/.

[6] P. S. S. Srivignessh and M. Bhaskar, "RFID and pose invariant face verification based automated classroom attendance system," in *Proceedings of the 2016 International Conference on Microelectronics, Computing and Communication, MicroCom 2016*, Durgapur, India, January 2016.

[7] "Hubstuff Employee Monitoring Software," 2017, https://hubstaff.com/employee_monitoring_software.

[8] "WaveTrend Access Control," 2017, http://www.wavetrend.net/access-control.php.

[9] "ObserveIT Insider Threat Solution," 2017, https://www.observeit.com/insider-threat-solution.

[10] J. Zendulka and M. Pešek, "Mining moving object data," *Open Computer Science*, vol. 2, no. 3, 2012.

[11] S. Kisilevich, F. Mansmann, M. Nanni, and S. Rinzivillo, "Spatio-temporal clustering," in *Data Mining and Knowledge Discovery Handbook*, pp. 855–874, Springer, Boston, MA, USA, 2010.

[12] E. M. Knorr, R. T. Ng, and V. Tucakov, "Distance-based outliers: algorithms and applications," *The VLDB Journal*, vol. 8, no. 3-4, pp. 237–253, 2000.

[13] Y. Ge, H. Xiong, Z. Zhou, H. Ozdemir, J. Yu, and K. C. Lee, "TOP-EYE: Top-k evolving trajectory outlier detection," in *Proceedings of the the 19th ACM international conference on Information and Knowledge Management (CIKM)*, J. Huang et al., Ed., pp. 1733–1736, Toronto, Canada, October 2010.

[14] R. K. Wong and K. Raymond, "Trajectory analysis based on clustering and casual structures," in *Proceedings of the Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[15] L. D. M. Lam, A. Tang, and J. Grundy, "Predicting indoor spatial movement using data mining and movement patterns," in *Proceedings of the 2017 IEEE International Conference on Big*

Data and Smart Computing, BigComp 2017, pp. 223–230, kor, February 2017.

[16] S. Khoshahval, M. Farnaghi, and M. Taleai, "Spatio-temporal pattern mining on trajectory data using ARM," *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 42, 2017.

[17] L. Li and C. Leckie, "Trajectory pattern identification and anomaly detection of pedestrian flows based on visual clustering," in *Proceedings of the 9th IFIP TC 12 International Conference on Intelligent Information Processing VIII, IIP 2016*, Springer International Publishing, Melbourne, Australia, 2016.

[18] E. S. Novikova, I. N. Murenin, and A. V. Shorov, "Visualizing anomalous activity in the movement of critical infrastructure employees," in *Proceedings of the 2017 IEEE Russia Section Young Researchers in Electrical and Electronic Engineering Conference, ElConRus 2017*, pp. 504–509, Petersburg, Russia, February 2017.

[19] E. Novikova and I. Murenin, "Visualization-Driven Approach to Anomaly Detection in the Movement of Critical Infrastructure," in *Computer Network Security. MMM-ACNS 2017. Lecture Notes in Computer Science*, J. Rak, J. Bay, I. Kotenko, L. Popyack, V. Skormin, and K. Szczypiorski, Eds., vol. 10446, pp. 50–61, Springer, Cham, Switzerland, 2017.

[20] D. McNamara, J. Tapia, C. Ma, and T. Luciani, "Spatial Analysis of Employee Safety Using Organizable Event Quiltmaps," in *Proceedings of the IEEE VIS Workshop on Temporal Sequential Event Analysis*, Baltimore, MD, USA, 2016.

[21] X. Li, J. Han, and S. Kim, "Motion-alert: Automatic anomaly detection in massive moving objects," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 3975, pp. 166–177, 2006.

[22] X. Li, J. Han, S. Kim, and H. Gonzalez, "ROAM: Rule-and motif-based anomaly detection in massive moving object data sets," in *Proceedings of the 7th SIAM International Conference on Data Mining*, pp. 273–284, April 2007.

[23] "VAST Challenge: MC2," 2016, http://vacommunity.org/2016+VAST+Challenge%3A+MC2.