# Group Activity Analysis for Persistent Surveillance

*Jeffrey S. Lin, Ariel M. Greenberg, Clayton R. Fink, Wayne L. Bethea, and Andreas G. Savvides*

*To help combat terrorist and insurgent threats, the DoD is deploying persistent surveillance systems to record the activities of people and vehicles in high-risk locations. Simple observation is insufficient for real-time monitoring of the vast amounts of data collected. Automated systems are needed to rapidly screen the collected data for timely interdiction of terrorist or insurgent activity. Effective analysis is hampered by the similarity of actions of individuals posing a threat to actions of individuals pursuing a benign activity. The analysis of the activity of groups of individuals, with requirements for team coordination, can potentially increase the ability to detect larger threats against the background of normal everyday activities. APL, in collaboration with Yale University, is developing sensor-independent approaches and tools to robustly and efficiently analyze complex group activities.*

## INTRODUCTION

Persistent video surveillance systems are used routinely for retrospective analysis of an attack. By using sophisticated facial recognition capabilities, surveillance systems might also be used to identify persons of interest at portals. The challenge is to use these systems to detect threatening activities by unknown actors in sufficient time to proactively respond to the threat and prevent an attack. The ability to meet this challenge requires posing different sets of questions and developing approaches to answer those questions. We pose two complementary questions: "If I know what activities I am looking for, how do I search for them?" and "If I do not know what I am looking for, to what should I pay attention?"

The need for tools for reasoning about databases of temporally labeled actions and transactions is a general need for many persistent surveillance applications including video, distributed sensor network, and electronic communication data streams. Event graph[1] and probabilistic Petri net[2] approaches for multiagent activity recognition have been described for video analysis. We propose a two-pronged iterative analysis approach including an extension of the event graph representation for detecting targeted group behavior and analysis of routine behaviors. Although we discuss in this article the application of these approaches to video analysis, we explicitly decouple the analysis tools from the feature extraction of the raw data and emphasize the formulation of models that are easily created, modified, and understood by the analyst. This "sensor-independent" implementation allows the algorithms and tools that are developed to be incorporated into non-video and multimodal persistent surveillance systems.

A notional analysis hierarchy is shown in Fig. 1 for a persistent video-surveillance application. Other analogous layers can be defined for other surveillance applications such as cell phone or e-mail communications. The vast quantities of raw video data acquired are systematically reduced by each layer of processing. Each successive layer extracts increasing abstractions of the data but necessarily loses information and introduces errors and uncertainty.

The image segmentation layer separates raw pixels into regions that share sufficient similarity (e.g., color, texture, temporal continuity) to be considered distinct from each other. Strong shadows and occlusions are two factors that may cause segmentation errors because the boundaries of the object are ambiguous. The entity classification and identification layer classifies the image/video regions as a physical entity such as a building, forest, vehicle, or person. Some systems may go as far as to identify the particular object, such as a specific individual, through feature matching to a database. Once an entity is classified, its location can then be tracked over time. The previous uncertainties of segmentation, classification, and identification are propagated into the tracking algorithms, leading to continuity and ambiguity errors of the tracks. The spatiotemporal activity and

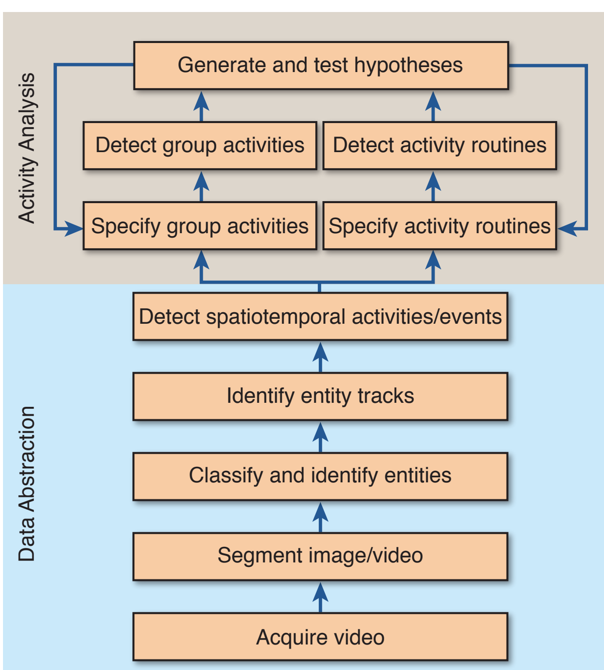


**Figure 1.** A notional processing hierarchy for the analysis of persistent video-surveillance data.

event detection layer remains a particularly active area of research and is focused on identifying the activity of individual entities in the video, with uncertainties generated in the accuracy of the activity interpretation.

Although not every persistent surveillance system includes the layers discussed, these layers do illustrate the hierarchy of data abstractions required to ultimately yield a database of actions and transactions, potentially from a heterogeneous suite of sensors, each tagged with data fields such as entity classification and identification; activity classification, start time, and end time; and a collection of relevant uncertainty measures. Although the types of activities and their detectable attributes and confidences will depend on the particular sensor system, the approaches for reasoning about the detected activities can be general.

## APPROACH

In many cases, the analysis of the actions of individuals is insufficient to discriminate threatening activity from benign activity. The analysis of the activity of groups of individuals, with requirements for team coordination, can potentially increase the ability to detect larger threats against the background of normal everyday activities. The top three layers in Fig. 1 represent our two complementary approaches to provide tools for analysts to interactively and iteratively build and refine queries against a database or streaming data to identify complex activities that may pose a threat. In the first, for targeted adversary goals, we develop a model of the expected group activity and then search the data for matches. In the second, we develop approaches to detect and describe routine behavior to understand the activity patterns of both our adversaries and the general population among which they operate.

## GROUP ACTIVITY QUERY

The top layers in Fig. 1 are expanded upon in Fig. 2. Once the analyst selects a targeted adversary goal and estimates the constraints, a model of hypothesized group activity can be developed through a planning analysis from the perspective of the adversary. The goal can be decomposed into subgoals, which can be further decomposed into tasks and subtasks. Each task or subtask is then assigned to a role to be assumed by an entity (e.g., person, vehicle, or location). We describe a task involving only one entity as an action by that entity, and a task involving more than one entity as a transaction between those entities. The detect group activities layer matches the roles and tasks in the specified group activity against entities and actions/transactions extracted by the abstract data layers for the given sensor system. By broadly defining an entity to be a person, vehicle, or

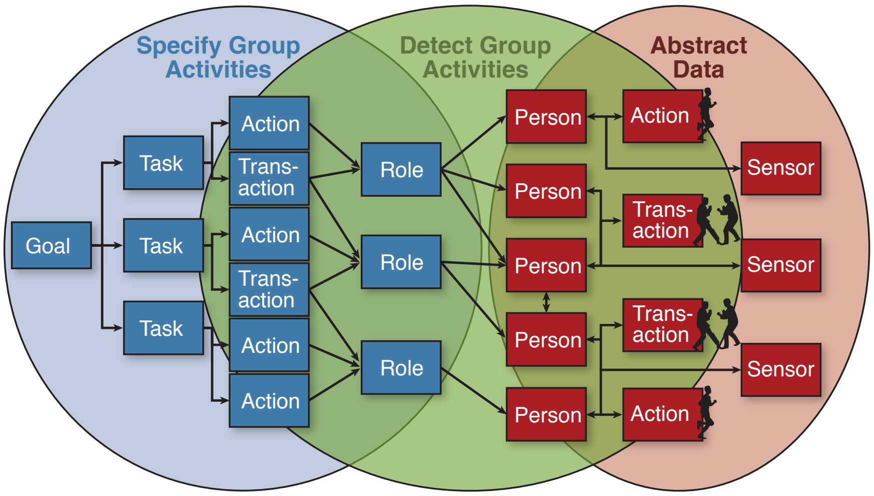


**Figure 2.** The detection of a specified group activity matches roles of the specification with observed individuals, and actions and transactions required to achieve tasks with observed actions and transactions.

location, the specified group activity is general to many applications and includes spatial relationships of people and vehicles with specific geographic locations, regions, or boundaries.

The specified group activity includes more details than are represented in Fig. 2. Most plans for coordinating multiple people toward a common goal have timing constraints. Some tasks must precede other tasks, and some tasks must be performed simultaneously. In addition, there are contingencies, with optional tasks substituting for other tasks. While multiple roles in the plan may be taken by one entity, other roles may require distinct entities.

The matching of the specified group activity to the action/transaction database presents several challenges. The computational complexity of the search for matches must be carefully managed, as the databases and streaming rates for persistent surveillance systems can grow large. This complexity is compounded by the need for inexact matching of the specification to the database, due to errors both in the specification and the database. The errors in the specification result from incomplete knowledge of the adversary's true constraints and options. The errors in the database include the abstraction errors mentioned but also include errors of omission because some activities may not be observed.

We focus the development of approaches for detecting expected group transactions on an open-air drug-deal scenario, inspired by an episode of HBO's dramatic series *The Wire*. The adversary's goal in this scenario is to complete an exchange of drugs for money. There are several constraints on execution of this conspiracy. First, to make detecting a transaction more difficult for the police, both the money and drugs must not be exchanged between the same two people. Second, to prevent theft, the customer should not be able to observe where, or with whom, the drugs are stored. By distributing the transactions over both time and space and by involving multiple individuals, the conspirators make it difficult for an observer to understand what is happening. The detection task is made more difficult against the background of everyday transactions of residents in the neighborhood, which, on a single-transaction scale, are indistinguishable from those transactions of the drug deal.

## TEST DATA

We have developed a simulation (see Box 1 and Fig. 3) running in the Virtual Battlespace 2 (VBS2) multiuser gaming environment to generate data for testing and evaluating our approaches and algorithms. The use of a gaming simulation as a data source offers many benefits:

- The gaming environment can accommodate both non-player characters (NPCs), with their behaviors controlled by finite-state-machine (FSM) models, and human players, with unpredictably creative behavior.
- A simulation gives control over the number of executed group activities and the complexity and scale of background individual activity.
- A simulation provides a complete symbolic record of all activity, eliminating the need for developing or selecting data abstraction software.
- Uncertainties inherent in sensing and data abstraction (e.g., noise, errors, and omissions) can be modeled as degradations of the accuracy and confidence of the ground-truth activity.

All of our experiments to date have used only the simulated activity of NPCs. Human players will be introduced later to evaluate the robustness of the inexact matching approaches we are developing. We define the behavior of each NPC using an FSM, with the transactions between NPCs emerging based on the individual responses. We have defined FSMs to produce the drug-deal scenario, as well as several background behaviors that draw from the same set of transactions within the drug deal: a flower purchase and giving scenario, a hot dog vending and purchase scenario, and a friendly wave.

## GROUP ACTIVITY SPECIFICATION

While developing our approach for specifying targeted group activities, we seek an intuitive and expressive

## BOX 1. SIMULATING GROUP ACTIVITY IN VIRTUAL ENVIRONMENTS

The multiuser virtual environments used to create online simulated-world games are also used for training, mission rehearsal, telepresence, visualization, and data generation. Game designers generate relatively complex behavior for NPCs—the computer-controlled agents in the game—with modeling constructs such as FSMs and behavior-based control. To generate coordinated group activities for our test database, we selected the VBS2 environment (Fig. 3a), used widely by the U.S. military for training, with NPC behavior controlled by FSMs.

An FMS captures a behavior model with a preselected set of internal states, such as waiting, eating, and sleeping. The FSM switches between these states according to rules governed by the current state, possible next states, external conditions, and chance. By carefully defining FSMs controlling the behavior of two NPCs, we can orchestrate desired transactions between the NPCs. Although we specify that an NPC is able to engage in a transaction, we do not know exactly when, for how long, or with whom the transaction will take place. We can approximate personality types by modifying the probability of transitioning between states for individual NPCs so that different NPCs prefer different activities as well as prefer to assume different roles in an activity.

Whereas generating desired individual actions of NPCs is relatively simple with FSMs, generating coordinated group transactions is not as simple. Behavior prescription in modern game design is egocentric, i.e., the atomic unit of activity is that of a person or a team with the environment (objects, terrain). No modeling construct explicitly represents a transaction; instead, each NPC's participation in a transaction is coded separately, with stimulus defined in one agent and the response in another. Group activity emerges as a result of synchronicity between individual NPC state transitions. As an example of different approaches to transaction prescription, contrast panels b and c of Fig. 3. Figure 3b shows the desired transactions between roles in our simulation. Figure 3c shows the expanded FSMs of the stash, runner, and cashier roles and the implied transactions between self-centered states and state transition criteria across the roles. As the number and complexity of the FSMs increase, the role transaction diagram ideally would be generated by yet-to-be-developed consistency-checking algorithms to validate the design of the FSMs and their transactions. A consequence of implicitly specifying transactions is lack of direct control of the transaction frequency. Transaction frequency is moderated by three interdependent factors: FSM state-transition probability, resource/counterpart availability, and duration of transaction. We achieve the desired overall transaction rate by iteratively customizing these factors.

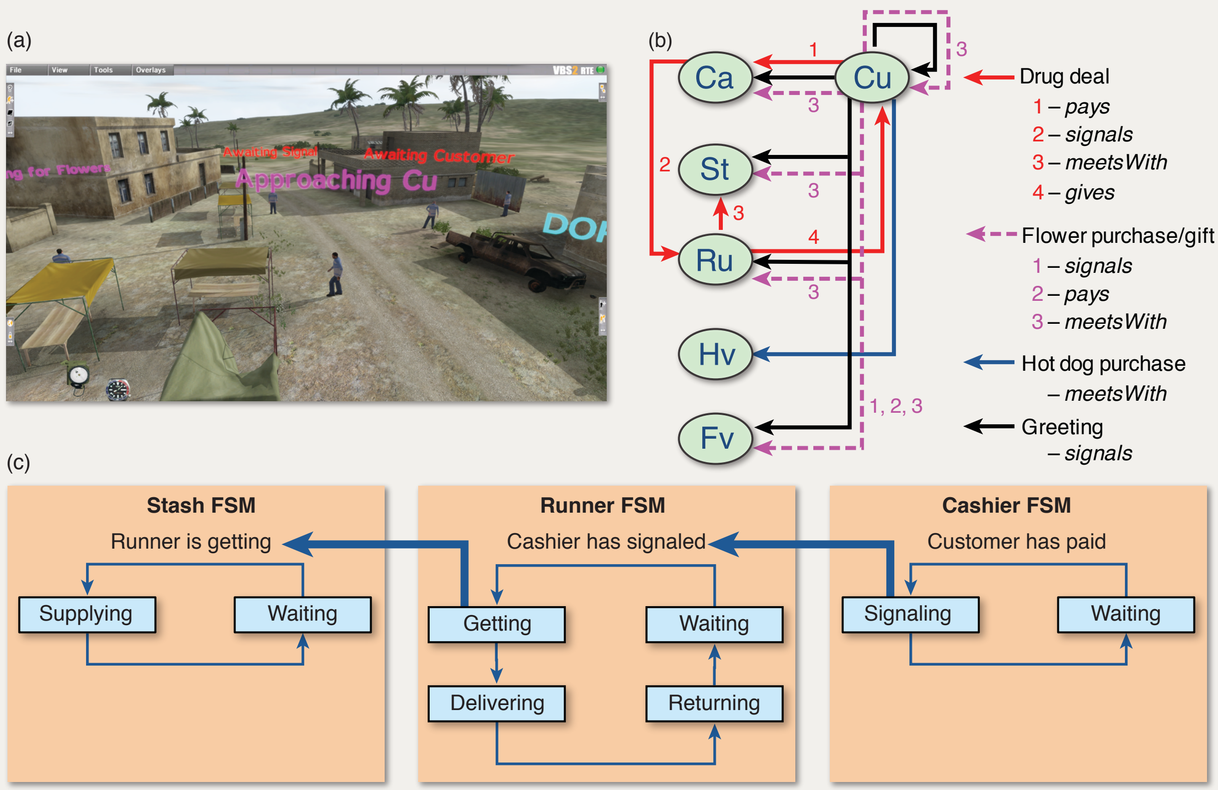


**Figure 3.** (a) An instant during the VBS2 simulation, with individuals tagged with states and transitions. (b) Transaction-oriented representation of group transactions between the roles of customer (Cu), cashier (Ca), runner (Ru), stash (St), hotdog vendor (Hv), and flower vender (Fv). See Box 2 for detailed descriptions of the transactions. (c) Implied transactions (shown as large arrows) between states in one FSM and satisfying state transition criteria in other FSMs.

notation of the goal–task–role decomposition shown in Fig. 2, along with the temporal constraints and relationships of the tasks. By leveraging analyst familiarity with graphical representations of social networks, we express the task–role relationships as a graph, with nodes representing individuals and edges representing actions (an edge from a node to itself) or transactions (an edge between nodes). The specified transaction network for our drug-deal scenario is shown in Fig. 4a. We specify that the customer pays the cashier, the cashier signals the runner, the runner goes to the stash, and the runner gives the drugs to the customer. The transaction network shows what (and potentially where) transactions must take place but does not show when.

The temporal constraints are specified by using another graph, a simple temporal network[3] (STN) (Fig. 4b). Each node-pair in the STN represents an action/transaction edge in the transaction network. The left node in the node-pair represents the beginning of the activity and the right node, the end of the activity. The directed edges in this graph indicate precedence, with the arrow pointing from the preceding activity to the following activity. The minimum and maximum allowable time intervals (in seconds) are shown as labels on the edges and node-pairs. Figure 4b specifies that the customer payment to the cashier is the first transaction, and the runner delivering to the customer is the last transaction. The lack of an edge between transactions B and C indicates that their relative ordering is not specified: if the runner anticipates the drug order, he may visit the stash before getting a confirmatory signal from the cashier.

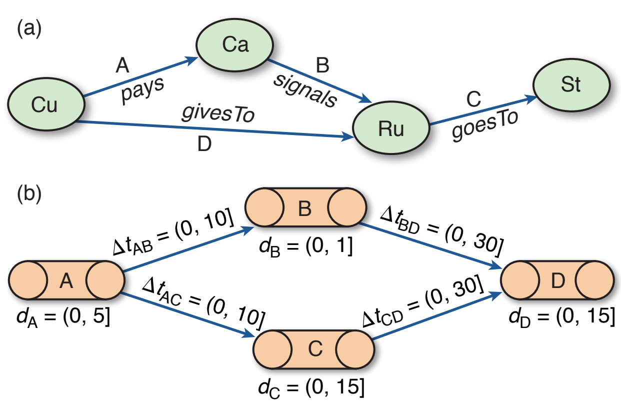


**Figure 4.** A specified group activity comprises (a) a group activity network of individuals (nodes) and transactions (edges) and (b) the constraint STN. The temporal constraints are expressed as allowable time intervals (in seconds) for $d_i$, the duration of, and $\Delta t_{ij}$, the delays between, the transactions. For example, the customer (Cu) paying the cashier (Ca) is transaction A. This transaction takes up to 5 s and is followed up to 10 s later by the cashier signaling (B) the runner (Ru) to deliver drugs to the customer.

The temporal relations described by Allen and Ferguson[4] and used by Hongeng and Nevatia[1] do not include numerical temporal constraints. Hongeng and Nevatia mention the potential expressive power of numerical temporal constraints while deferring implementation due to representation and algorithmic complications.[1] In an application with a large number of transactions, these numerical temporal constraints are critical in pruning the search space of the query. If one activity is specified as preceding another activity without any constraint on the time lag, every pair of activities must be evaluated, resulting in an explosion of both returned matches and search time as the database size increases.

We have implemented the capability to specify a group activity in our prototype Group Activity Network Analysis (GANA) software, leveraging APL software previously developed for rapid, iterative query refinement against a social-network database. This software has extensive user-centered capabilities. The first is ontology-assisted queries (see Box 2 and Fig. 5), enabling the user to construct a group activity specification in terms of problem-specific concepts that expands into queries against the full set of relevant database fields. Another capability is the direct interaction with the analyst by using graphs, with interactive visual construction of graph queries, and return of database matches as graphs (Fig. 6). Unseen by the user, GANA generates textual database queries (e.g., structured query language or "SQL") directly from the graphical representations created by the user, executes the query against the database, and processes the returned records of matching activity for displays as graphs.

The group transaction network and STN form the basis for our group activity detection approach. We are currently addressing the challenges of robust group activity detection, including:

- Data abstraction errors that corrupt otherwise matching database information
- Individuals or activities not being observable by the data collection system
- Alternative paths to accomplishing the same targeted adversary goal
- Incorrect assumptions resulting in partial mismatches of the transaction network and/or STN

The potential sources of detection error are in both the data abstraction and the activity analysis, suggesting a consistent uncertainty management approach across the layers. The analyst is permitted to assign measures of uncertainty to each part of the activity specification. When the uncertainty management is fully implemented, the analyst will be able to rapidly screen surveillance data by iteratively posing queries of varying specificity and then sorting returned matches by an integrated measure of overall (data abstraction and activity specification) uncertainty.

## BOX 2. ONTOLOGY-ASSISTED QUERY

The explicit and expressive semantics of an application area's concepts, together with their relationships represented through logical formalisms and inference, constitute a knowledge representation known as an ontology. Ontologies allow automated processing of data and information in a logical, well understood, and predictable way. In the drug-deal scenario there are roles of customers, cashiers, runners, and stashes, and the relationships among those roles are the transactions *pays*, *signals*, *givesTo*, and *meetsWith*. In GANA we use ontology-assisted queries to visually explain the defined concepts and relationships to the user to facilitate graph query construction and to enable automated expansion of queries based on the ontology.

One semantic construct GANA takes advantage of is the subsumption semantic relation, i.e., the *is-a* relation in knowledge representation, to assist in query construction and query execution. Subsumption in classes means that an instance of the subsumed class can be used in any place that an instance of the subsuming class can be used. For example, an instance of a woman can be used anywhere an instance of a person can be used within a system, because a woman is a person. In the GANA drug-deal scenario there can be a *meetsWith* transaction, a *givesTo* transaction, and a *pays* transaction, each of which describe parts of a drug-deal scenario and are represented by a number of edges in the ontology graph. In an ontology we represent these transactions as successively more specific or specialized versions of kinds of transactions through the subsumption relationship. Therefore, a *givesTo* transaction is more specific or specialized than a *meetsWith* transaction, and a *pays* transaction is more specific or specialized than a *givesTo* transaction. Stated another way, a *pays* transaction *is-a givesTo* transaction, and a *givesTo* transaction *is-a meetsWith* transaction (Fig. 5a). By using subsumption, GANA can assist the user in exploring (Fig. 5b) and visually constructing (Fig. 5c) a desired query, or it can automatically execute an appropriately expanded set of queries that leverage the semantic information encoded in the ontology.

Another semantic construct GANA will take advantage of is the symmetry semantic relation. Symmetry means that for all classes $x$ and all classes $y$, $x$ *relatesTo* $y$ implies $y$ *relatesTo* $x$, where *relatesTo* is a semantic relation. In the GANA drug-deal scenario a *meetsWith* relation may be described as symmetric in the ontology, which means if customer *meetsWith* cashier it is implied (and can be inferred) that cashier *meetsWith* customer. This would allow a user to explore a graph schema in much more flexible and dynamic ways. Subsumption and symmetry are just two of the semantic constructs that GANA takes advantage of in providing ontology-assisted graph query. Some other constructs GANA could take advantage of through its use of ontology technology include reflection, inverse-relation, transitivity, equality, and disjointness.

(a)
```
<owl:ObjectProperty rdf:ID=”meetsWith”>
<rdfs:label>meets with</rdfs:label>
<rdfs:domain rdf:resource=“#Person”/>
<rdfs:range rdf:resource=“#Person”/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:ID=“givesTo”>”’
<rdfs:label>gives to</rdfs:label>
<rdfs:subPropertyOf rdf:resource=“#meetsWith”/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:ID=“pays”>
<rdfs:label>pays</rdfs:label>
<rdfs:subPropertyOf rdf:resource=“#givesTo”/>
<rdfs:domain rdf:resource=“#Customer”/>
<rdfs:range rdf:resource=“#Cashier”/>
</owl:ObjectProperty>
```

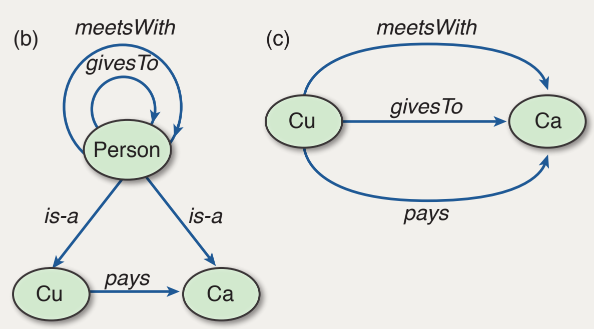


**Figure 5.** (a) The Web Ontology Language (OWL) definition of the *meetsWith, givesTo,* and *pays* transactions in a drug-deal context. (b) These transactions are shown in relation to the cashier (Ca) and customer (Cu) roles shown in the ontology graph for the user. (c) The options for transactions to specify in the user graph query, as generated by the GANA use of the ontology.

With the relative ease of acquiring enormous quantities of data, the next challenge becomes performing the database searches in a reasonable time for problems of a useful size. Fortunately, the search is less complex than (unordered) subgraph matching, which is NP-complete. The temporal constraints on the transactions allow pruning of subgraph searches, greatly reducing the search depth. The complexity of the search is therefore a function of the time-density of observed transactions relative to the timing constraints, as well as a function of the number of actions and transactions in the specification.

A recent test of query speed was performed against databases of transactions extracted from our VBS2 simulation. The first database contains 899 transactions, with an average of one transaction every 6.4 s performed among 439 individuals. The second database is derived from the first, duplicating records and changing times and person identifiers, resulting in twice the transactions and individuals with the same transaction

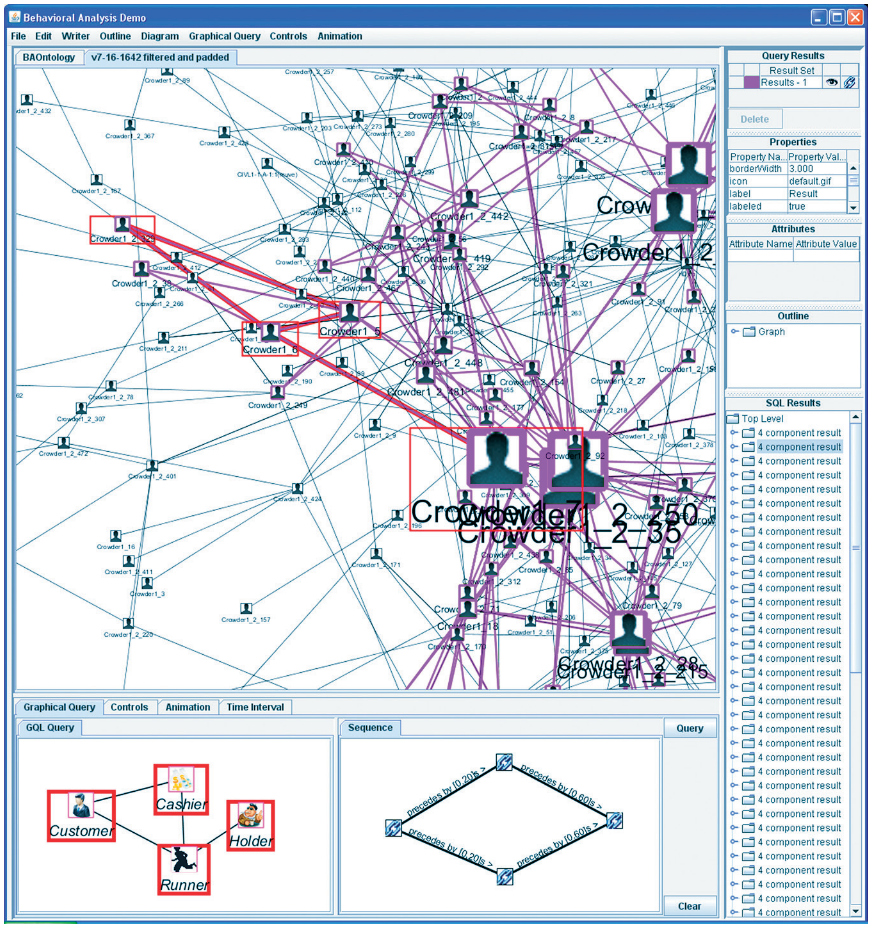


**Figure 6.** GANA user interface showing the specified group activity defined by the user as a transaction network (lower left) and STN (lower right). The results are shown in the upper window, with all matching subgraphs highlighted in purple and the one selected by the user highlighted in red.

rate. The third database similarly doubles the size of the second database. The drug-deal query shown in Fig. 4 was executed against the three databases using the H2 relational database engine and a desktop computer. The average of 10 database engine execution trials is shown in Fig. 7. The scaling of the query is close to linear, as shown by the comparison to the appropriate multiples of the time for the first database. Although this speed may be sufficient for many applications, we are investigating both graph analysis[5] and database-optimization techniques for increasing the scale of the problems addressable by this approach.

Knowing the ground truth for the simulated activities in the database, we were able to calculate performance metrics for the executed queries. We created a larger database with 1163 drug deals, 1903 flower sales and deliveries, and 8019 hot dog sales. The simulated hot dog sales and flower sales and deliveries are designed to generate activity patterns similar to drug deals. The precision of the query is the fraction of returned subgraphs that are actually drug deals, and recall of the query is the fraction of drug deals that were correctly returned as subgraphs. The detection performance of the drug-deal query for several different temporal constraints is shown in Fig. 8. The shorter times on the left result in high precision, with few false detections but a relatively low recall, as one-quarter of the drug deals are missed. As the temporal constraints are relaxed, the recall rate increases, but the precision falls as more random transactions are mistaken for a drug deal.

## ROUTINE DISCOVERY AND CHARACTERIZATION

As a complement to the model-driven specification of targeted group activity, we are investigating data-driven approaches for discovering routine activities. An understanding of the activity pattern of a person or population helps to identify interesting activity, either because it does not fit into a known pattern, a pattern has evolved, or a new pattern has emerged. In addition, a detected instance of a routine activity can be included as part of a larger group activity specification.

Although many of the human activities in the physical world can be casually described as routines, identifying these patterns of unknown structures in time and space is a challenge because of the patterns being embedded among unrelated data sequences and the data streams having timing behavior spanning multiple spatiotemporal scales.

We have investigated approaches to identify human routines by using location data extracted from camera network test beds.[6] The test bed was developed for research on monitoring the elderly and those in assisted living. We observed that recurring human routines tend to happen inside periodic time windows (i.e., hourly, daily, weekly, etc.). The routines themselves were not periodic in the strict sense but they occurred within time intervals that are periodic.

Using data from a live test bed, we performed the data abstraction steps discussed previously to produce a database of activities. We used privacy-preserving imaging sensors in our house test bed, and there was typically only one individual in the house. We therefore had minimal data abstraction requirements. Given our low-

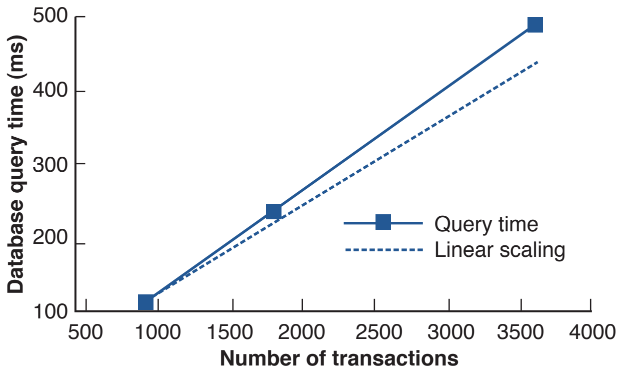


**Figure 7.** Average query time for 10 trials of the drug-deal query as a function of number of transactions in the database. The dashed line linearly scales the times for the smallest database.

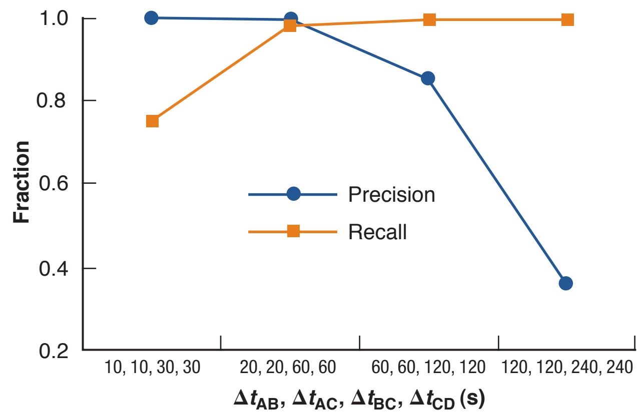


**Figure 8.** Recall and precision metrics for drug-deal queries with different transaction delay constraints.

## BOX 3. ANALYSIS OF ROUTINE ACTIVITIES

The algorithm for detecting a human behavior routine in a sequence of events evaluates candidate periods, $l$, and finds the smallest time envelope in which a given event satisfies the desired frequency and consistency parameters. This step is needed because the occurrences of events are not periodic in the strict sense, but they do occur within time envelopes that are periodic. The challenge in detecting these routines is to simultaneously identify the period of the routine envelope and determine which events occur persistently within the discovered time envelope.

The algorithm for determining whether a set of events is a routine with a candidate period $l$ is based on a sliding window sequence approach. Suppose the event type "kitchen visits that last approximately an hour and occur between noon and 5:00 p.m. every day" is a routine. To help visualize the basic approach, Fig. 9a shows these events on a time line, which with inspection shows that there is a sequence of contiguous time intervals, each of length $l$ = 24 h, such that each 5-h envelope in the routine belongs to one of the intervals, and no two envelopes are in the same interval.

We determine whether the events are part of a routine by analyzing each candidate interval $l$, from smallest to largest. We have developed an efficient algorithm[7] to determine the set of all possible intervals. If $L$ is the length of the entire interval of observation, and $t_0$ is the first time point on the interval, we can construct $W$, a sequence of contiguous $\left\lfloor \frac{L}{l} \right\rfloor + 1$ time intervals each of length $l$,

$$W = [t_0 - l, t_0], [t_0, t_1], [t_1, t_2], \cdots, \left[t_{\left\lfloor \frac{L}{l} - 1 \right\rfloor}, t_{\left\lfloor \frac{L}{l} \right\rfloor}\right],$$

as seen in Fig. 9b. Let δ denote the distance between the first event and the left endpoint of the time interval $W$ containing $t$. If we slide the entire sequence of time intervals in $W$ to the right by δ (Fig. 9c), we will discover a set of envelopes [of events with the same type as (*kitchen, 60 min*)] that make up a temporal property of a routine [(*kitchen, 60 min*)] with period $l$. Because δ is at most $l$, we will, after at most $l$ time units, find that (*kitchen, 60 min*) is a routine of period $l$ with a frequency of 4, a minimum consecutive repetition of 2, and with events in 66% of the observed time intervals. The time envelope of the routine is found by reversing the slide of $W$ until events no longer are in separate intervals.

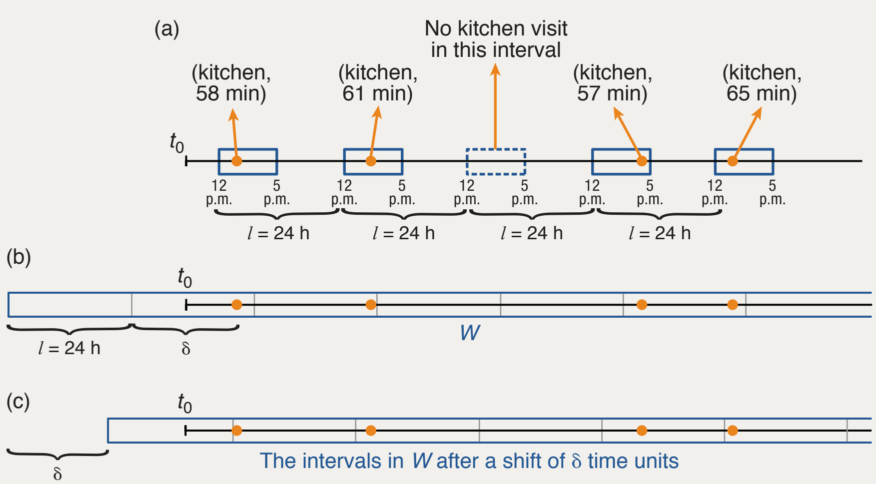


**Figure 9.** (a) Shown is a set of four approximately hour-long kitchen events and the targeted characterization of a 5-h time envelope (denoted by blue rectangles) and a 24-h periodic interval, $l$. (b) Given a candidate interval $l$ = 24 h (as part of a series of candidate intervals), construct a sequence $W$ of the intervals. (c) Shifting $W$ by increments up to δ will find that $l$ is a periodic interval for the events.

resolution sensors, we directly interpreted the presence of an individual at a specific location in their home as an activity of that individual. For example, presence in the dining room was interpreted as a dining activity. Location and activity can be separately recorded as a natural extension of this work. In addition to the activity classification, we recorded the time and duration of that activity for each instance. This processing allowed us to construct an activity database for the resident of the test bed.

To discover routine behavior, we aimed to find all spatially tagged activities with approximately the same start time and duration within periodic time intervals of interest. We have developed efficient algorithms (see Box 3 and Fig. 9) to detect and characterize routines for each activity type across a range of periodic time intervals. The strength of each routine is a measure of the consistency with which the activity is observed as part of the routine. This approach is easily extensible to other applications with multiple individuals and more complex activities derived from more informative sensors.

The spatiotemporal characterization of activity routines allows a more powerful encoding of activity that includes the temporal context of the activity. The same activity may have a different meaning at different times of day. For example, a 7- to 9-h presence in the bedroom at night can be interpreted as sleeping, whereas a 1- to 2-h presence in the bedroom during the day can be interpreted as napping. With the activities clustered into spatiotemporal events, traditional data mining techniques can now be used to discover correlations between events and build spatiotemporal models of the observed data.

A model of individual activity routine derived from 30 days of data from the test bed is shown in Fig. 10. The model is represented similarly to an FSM, with the activities represented as nodes and the probability of transition to the next activity represented by a labeled edge. The labels for the nodes are user-specified interpretations of the spatiotemporal events. The labels are provided for convenience and illustration but are unnecessary because the nodes are explicitly defined by the location, start time, and duration of the event. Varying levels of modeling resolution can be obtained by varying the threshold strength for the activity routines. The circuit of orange edges in Fig. 10 represents the sequence of activities in a normal day, as defined by the most probable path through the model.

This general approach to routine discovery and modeling forms the basis for the general spatiotemporal analysis of routine activities of multiple entities in

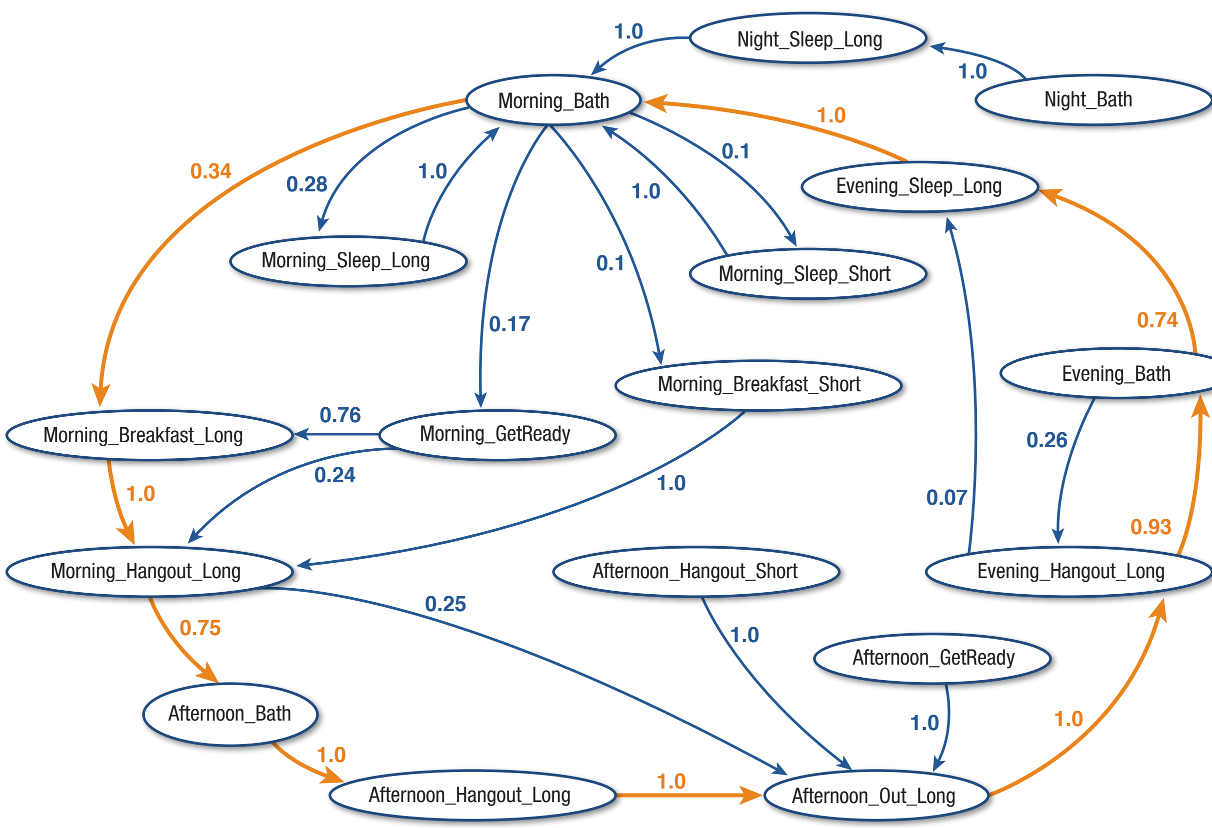


**Figure 10.** A model of routine activities for a 1-day time window derived from an instrumented house. Activities are clustered based on when, where, and how often they are performed. Activities that fit a spatiotemporal profile are modeled as an FSM with probabilities derived from observations, yielding a predictive model of routine behavior.

a persistent surveillance context. Data-driven models of routine activity enable novel capabilities for the spatiotemporal analysis of surveillance data. The presumably large number of routine activities can be separated from those activities that are not routine. First, the stable and strong routines can be analyzed to understand and characterize a large fraction of everyday activities forming the background activity "noise" against which one is seeking to identify threats. Second, a shift in the activity from that predicted by the model may indicate that the population knows of an unseen threat. Lastly, with the routine activities removed, the burden of examining the remaining activities is reduced for alternative analysis such as for the detection of targeted group activities.

## CONCLUSIONS

The challenges of understanding the coordinated activities of more than one individual monitored by persistent surveillance systems are numerous. To efficiently and accurately extract the salient information from the raw data, many technologies must be tailored to the particular sensor suite and desired system goals. We are investigating analysis approaches and tools that can be shared across many of these systems. With a focus on developing scalable approaches useful in real-world applications, we are leveraging expertise spanning several technical fields and two institutions.

For the detection of specified group activities, we have developed general and powerful visual representations of both the query and database returns, connected by automated and efficient database searches, to enable rapid screening of large databases and iterative hypothesis generation and evaluation. The next step is to implement, test, and refine strategies for more robustly specifying group activities and to validate these approaches by adding human players and enhanced sensor-error models to our simulations.

We have also developed novel, efficient approaches for the detection and characterization of routine activities. These approaches have been tested on real-world test beds by using video and Global Positioning System sensors. We will continue the validation of routine detection and characterization on increasingly complex real-world data.

**ACKNOWLEDGMENTS:** We are grateful for the significant insights and contributions of Athanasios Bamis, Jia Fang, Dimitrios Lymberopoulos, Nathan Bos, Russell Turner, John Gersh, and George Cancro. This material is based on work supported by the National Science Foundation under Grant IIS-0715180. In addition, portions of this effort were funded by the APL Science and Technology Business Area.

### REFERENCES

[1]Hongeng, S., and Nevatia, R., "Multi-Agent Event Recognition," in *Proc. Eighth IEEE International Conf. on Computer Vision*, Vancouver, Canada, Vol. 2, pp. 84–91 (2001).

[2]Albanese, M., Moscato, V., Chellappa, R., Picariello, A., Subrahmanian, V. S., and Udrea, O., "A Constrained Probabilistic Petri-net Framework for Human Activity Detection in Video," *IEEE Trans. Multimedia*, **10**(6), 982–996 (2008).

[3]Dechter, R., Meiri, I., and Pearl, J., "Temporal Constraint Networks," *Artificial Intell.* **49**(1–3), 61–95 (1991).

[4]Allen, J. F., and Ferguson, G., "Actions and Events in Temporal Logic," *J. Logic. Comput.* **4**(5), 531–579 (1994).

[5]Bamis, A., Fang, J., and Savvides, A., "Detecting Interleaved Sequences and Groups in Camera Streams for Human Behavior Sensing," in *Proc. Third ACM/IEEE International Conf. on Distributed Smart Cameras (ICDSC)*, Como, Italy, pp. 1–8 (2009).

[6]Lymberopoulos, D., Bamis, A., and Savvides, A., "A Methodology for Extracting Temporal Properties from Sensor Data Streams," in *Proc. 7th Annual International Conf. on Mobile Systems, Applications and Services (MobiSys 2009)*, Krakow, Poland, pp. 193–206 (2009).

[7]Bamis, A., Fang, J., and Savvides, A., "A Method for Discovering Components of Human Rituals from Streams of Sensor Data," in *Proc. 19th ACM International Conf. on Information and Knowledge Management (CIKM '10)*, Toronto, Canada, pp. 779–788 (2010).

# The Authors

Jeffrey S. Lin

Ariel M. Greenberg

Clayton R. Fink

Wayne L. Bethea

Andreas G. Savvides

**Jeffrey S. Lin**, **Ariel M. Greenberg**, **Clayton R. Fink**, and **Wayne L. Bethea** are APL staff members in the System and Information Sciences Group of the Milton S. Eisenhower Research Center. Mr. Lin is a member of the Principal Professional Staff, and his current research interests are the analysis and modeling of interactive behavior and biological systems and processes, including proteomics, genomics, and biomechanics. The current research focus of Mr. Greenberg, a member of the Associate Staff, is complex systems modeling, particularly in the domains of human behavior and molecular biology. Mr. Fink is a Senior Software Engineer whose current research interests are in developing approaches for analyzing online user-generated text for understanding psychological, social, and cultural phenomena. Dr. Bethea is a member of the Senior Professional Staff, and his recent work is in the areas of data modeling, data representation, and data management up through the processing chain to information management, knowledge management, and pragmatics. His current research interests focus on semantic technology, including (but not limited to) ontology, ontology mapping, semantic Web, semantic Web technologies, knowledge representation, semantic expressivity, knowledge management, and semantic discovery. **Andreas G. Savvides** is the Barton L. Weller Associate Professor of Electrical Engineering and Computer Science at Yale University. He joined Yale in 2003, where he leads the Embedded Networks and Applications Laboratory (ENALAB). Dr. Savvides's current research interests include spatiotemporal sensor data processing for the analysis of human behavior by using sensors, macroscopic sensor composition from simpler sensors, networked systems for sensing humans and their application in energy systems, elder monitoring, and aging-in-place applications and security. For further information on the work reported here, contact Jeffrey Lin. His e-mail address is jeff.lin@jhuapl.edu.

The *Johns Hopkins APL Technical Digest* can be accessed electronically at **www.jhuapl.edu/techdigest**.