A Combined Discriminative and Generative Behavior Model for Cyber Physical System Defense

Owen McCusker
Guardian Services
Sonalysts, Inc
Waterford, CT, USA
mccusker@sonalysts.com

Scott Brunza
Guardian Services
Sonalysts, Inc
Waterford, CT, USA
mccusker@sonalysts.com

Marco Carvalho
Associate Professor
Florida Institute of Technology
Melbourne, FL, USA
mcarvalho@cs.fit.edu

Dipankar Dasgupta
Professor of Computer Science
University of Memphis
Memphis, TN, USA
dasgupta@memphis.edu

Setu Vora
Pulmonary Physician
Backus Hospital
Norwich, CT, USA
setu区委@yahoo.com

Abstract— In this position paper we explore the use of behavior models as an enabling methodology in the promotion of a more holistic understanding of CPS that can bridge both cyber and physical domains. Thus, we investigate the use of aggregate behavior analysis techniques combined in both cyber and physical domains. Ultimately, our work focuses on the development of a cyber-physical behavior model that leverages behavior aggregation promoting the creation of a long-view sense-making capability driven by both cyber and physical observations. We look to the use of this approach to establish the ability to anticipate malicious activity in CPS, rather that react.

Keywords— Behavior model, aggregate behaviors, behavior graphs, cyber physical, healthcare, patient safety, machine learning

I. INTRODUCTION

Our perceptions are predominately driven by the present and rooted in short-view data-intensive operational capabilities, establishing a context for our defensive models protecting our nation's infrastructure. As the interconnectedness of our infrastructure increases, these models fail short in their ability to deter an increasingly complex cyber and physical threat. Cyber-Physical Systems (CPS) demand more holistic models in terms of the disparity in functionality that exists between physical and cyber elements of these complex system [1]. A most prevalent case in CPS defense is found in Pervasive Health Monitoring Systems (PHMS), where patient safety is directly linked to cyber defense. Currently, cyber and physical models are not integrated when comprehending patient safety. Our ability to anticipate malicious behaviors in a cyber physical context is limited by the current misuse-based security models that are inherently separated between kinetic and non-kinetic domains. Current trends in the development of security models focus on single ingress points in the establishment of observations and sense-making. Distributed model development is a relatively new area of research and has potential in the protection of CPS. CPS, (e.g. PHMS), are trending to be more distributed and connected, and represent capabilities deployed over geographically dispersed assets. Yet we lack enabling technologies that promote distributed defense within CPS.

Pervasive Healthcare Monitoring Systems (PHMS) are a type of Cyber-Physical System (CPS) that provides a more integrated, agile and responsive ability for healthcare professionals to provide care to their patients. PHMS link patient safety to both cyber and physical models and operations. Yet, current Cyber-Physical Systems (CPS) lack comprehensive detection models that enable integration at the system and the system-of-systems level, as noted in [1], there is a “lack of integrated approaches for securing both the cyber and physical aspects” of these complex systems. New models should integrate both cyber and physical effects, enable distributed defensive strategies, and provide better situation awareness realized through a tighter coupling between physical and cyber system elements. Most importantly these models should provide the inherent ability to scale up to the data volume requirements correlating both cyber and physical observations. This is especially true where task execution time is a matter of correctness in CPS as opposed to a matter of performance in general purpose computing.

CPS requires threat detection models to move past misuse based strategies and integrate anomaly detection ones, integrating effects realized in both cyber and physical domains using hybrid models [4]. Where anomaly detection is based on the desired or positive behavior of system objects (e.g. users, processes, network traffic) we can now move to develop normative specifications in terms of correlated cyber and physical behaviors.

Behavior analysis is a type of anomaly detection strategy. Over the past few years behaviour-based models have emerged to bridge the gap in capability focused on anomaly detection of emergent threats [6], [3], [7]. Current cyber behaviour-based
systems are predominantly event-centric, where behaviours are extracted from event features and aggregated over time in terms of both a source and a destination. Our previous work leverages device-centric aggregate behaviours that facilitate the integration and correlation of cyber physical events and scaling to meet the data handling requirements of CPS [9].

In the recent past, the healthcare industry has incorporated the use of pervasive technologies facilitating the management of patients [13], specifically, pervasive healthcare monitoring systems (PHMS) are used to provide continuous monitoring of a patient's health. PHMS' gather health data using a complex and heterogeneous network of medical and environmental sensors/devices worn by the patients and connected to backend servers for storage and analysis. PHMS are promoting the bridging of the CPS through complex networking and such integration has a direct effect on patient safety. Yet, there are fundamental differences between the design approaches between cyber and physical components of such systems. Embedded developers in the past would purchase batches of processors to ensure that the supply would last for years, implicitly locking themselves into security constraints inherent in the devices [14].

The cyber portion of CPS has different developmental lifecycles and security patch cycles; when vulnerabilities are found in software patches are delivered to systems for updates. Bridging these profoundly different development paradigms without a common defensive model leaves this system vulnerable, and unlike information systems, the impact of attacks on physical systems can be instantaneous.

In the past [9] we have introduced a discriminative model that establishes the ability to detect behavior primitives. Our goal in this research is to establish a rich set of behavior primitives, providing a basis for the realization of a threat language in terms of behavior narratives. The behavior narratives are then realized in terms of probabilistic graphs (Figure 1, Layer 4 Host/Network-based Narratives/Motifs).

Figure 1 illustrates a layered detection methodology broken down into two sections: Generative Models and Discriminative Models. This perspective promotes the development of a set of models that work independently at different abstraction layers and yet cohesively within the overall problem domain. Research is being done today that combines probabilistic networks aligning observations from multiple domains [17]. Our work would first focus on transforming observations into a behavior feature space and then align.

We envision multiple parallel algorithms being leveraged in the layered model. The approach we take in this research is to establish a rich set of behavior primitives that facilitates the correlation of behavior between cyber and physical components of a CPS. These behavior primitives represent the base constructs of a behavioral language, where behavior graphs can establish threat narratives. Narratives are viewed as graphs of behavior primitives that capture aggregate description of threat agents. These threat narratives can represent social relationships and/or characteristics that are shared between groups of devices within a CPS. Narratives can be shared as actionable intelligence throughout a trusted community of cyber defenders.

In this current work we aim to expand on that model establishing: 1) generative models that are used to realize emergent behavior graphs, attack models and normal behaviors, 2) mature discriminative models that provide a set of behavior primitives that are used as building blocks in understanding normal and abnormal behaviors in CPS, 3) provide a model that realizes aggregate behaviors in both cyber and physical domains and 4) integrate the two models creating a generative discriminative framework that establishes a model to defend CPS.

II. RELATED WORK

In this section, we review related behavior analysis work. We then review related behavior modeling research leading to the development of a hierarchical model comprised of discriminative and generative models. We then present important works on cyber physical systems (CPS) and their application to PHMS, trust models and probabilistic graph theory as it pertains to CPS defense. This section contrasts the CPS needs with the models that were used in establishing existing defensive capabilities.

The dynamic nature of threat agents represents a challenge in developing feasible classification methodologies. Supervised learning provides the basis for training a classifier using a labeled data set, albeit the establishment of ground truth is costly. Classifiers tend to be less accurate when trained using unlabeled data using unsupervised methods. In our approach we want to leverage the combination of both labeled and unlabeled data to improve the accuracy of our classifiers creating a behavior model that combines both discriminative and generative models. We posit that such a model can better adapt to the dynamic nature of threat agents.
There are a number of behavior models that have been identified that create a behavior interpretation of network behaviors and user behaviors. Some of these models are discriminative, yet others are generative. Often, these models function off of raw data (See Figure 1) and may be difficult to scale with both cyber and physical observables. Our approach differs by taking a device-centric and aggregate view of behaviors.

Probabilistic graphical models, a type of generative model, are common in language processing and bioinformatics. These models have not yet been applied to the classification of threat agents in terms of aggregate behaviors. In the subsections we review related work in: A) Behavior Models, B) Generative and Discriminative Approaches in Intrusion Detection Models and C) CPS Modeling Challenges.

A. Behavior Models

Behavior analysis, and their supporting models, provides unique insight into making sense of threats in CPS environments. This approach offers the ability to further integrate cyber and physical observations supporting a more coherent model. The models supporting behavior analysis have been used to establish trust [6] [10] [11] [19], provide generative models to identify insider threats [7], and establishing metrics for behavioral trust in social networks [20]. These approaches when combined with behavior aggregation can provide a basis for hierarchical behavior model to protect CPS.

Rehak et al has created a agent-based behavior model that that establishes metrics for distributed trust [6]. In this approach classifiers agents are used in scoring events as legitimate or malicious [6]. This approach uses a self-adaptive technique based on the anomaly detection paradigm. Essentially, a generative model is used to process challenges that allow user agents to identify the threats, which achieves the best separation of the challenges that represent known instances of legitimate behavior from the challenges that represent known malicious behavior. Trust is then measured in terms of the distance between normal and abnormal behaviors established by the system.

In [7] a behavior is defined in terms of an “observable action”, where a cyber behavior is defined in terms of activities and a statistical characterization of those activities. The statistical characterization can be seen as a discriminative modeling of cyber behaviors. These behaviors can be realized as activities such as: web browsing at a high level that can be decomposed into an activity tree. Instead of combining both a generative and discriminative model, the approach applies concepts of topic maps to the emergence of activity trees creating a purely generative model.

In [3], Czejdo, et al, use aggregate event graphs to make-sense of behaviours, obtained from sensors. The event takes into account both the source and destination providing a connection, or edge in the graph. Aggregation operators are then used to transform the data, and enable its exploration by security analysts. This approach is contrasted with our approach in that it is event-centric, where ours is device-centric. Device-centric aggregation can lead to further reduction in device behaviours, and provide correlation opportunities in CPS.

Lastly, LLNL created a system SETAC that uses a distributed agent-based model to detect both local and global anomalous behaviors within their networks [8] [15]. It is stated that by LLNL the three differentiating characteristics of SETAC are distributed decision making, behavior modeling using machine learning, and real-time analysis and detection.

B. Generative and Discriminative Approaches in Intrusion Detection Models

Discriminative and generative models have both been applied to complex security problems to identify network threats. Examples of discriminant models include linear discriminant analysis, random forests, support vector machines, boosting and neural networks. Discriminant models are inherently supervised. In [16], both learning-based and specification-based behavior models are reviewed as applied to intrusion detection. This paper looks at the learning-based approach in terms taxonomy defined as: rule-based, model-based, and statistical based. Based on the greater availability of data computational power approaches have been slowly migrating from rule-based models to probabilistic data-driven models.

Generative models have been used as anomaly detection classifiers [5]. In this paper a graphical model is used to arrange the coverage of multiple anomaly detectors running over network traffic specifically looking at IP addresses and ports. Their approach uses a topic model training technique, specifically Latent Dirichlet Allocation (LDA). LDA is a type of generative model originally developed to identify unobserved topics within a set of text documents. The analogy used is that IP src-dst pairs represent a document, where each to-port represents a word, and vocabulary of the corpus is defined in terms of all the unique ports used by an IP pair. Our approach differs where all the behavior primitives used by a single host define our vocabulary.

Robinson has used generative models in the context of topic maps to create a behavior model for a user’s cyber behaviors that can identify threat agents such as insider threats [7]. Their approach is used to “characterize, predict, and detect change” in both individual and group behaviors. They feel that topics are similar to behaviors and build of the use of topic models as is used in the application of natural language processing to documents [2].

In the context of this paper we are establishing a model that is used to create a behavior language to describe an object we are tracking (e.g. host, device, user, threat agent). After we define the language we then use that language to determine if the object exists within the sample space.

C. CPS Modeling Challenges

In [4], Derler et al, states the “intrinsic heterogeneity, concurrency, and sensitivity to timing of ” CPS have created many modeling challenges. CPSs have unique problems that general purpose computing domains. An example provided is that the time it takes for a task to complete within the general purpose computing is a matter of performance and not
correctness. Yet within CPSs, such as PHMS, task execution time is critical to safety and has a direct impact on risk.

D. Pervasive Healthcare Monitoring Systems and CPS Modeling Integration

Patient safety is directly linked to both cyber and physical changes within PHMS, yet most models, and the realization of these models, separate the handling of these domains. In this section we will identify various operational areas that support healthcare services in terms of telemedicine, mHealth, and Telemonitoring. We will review these areas in terms of a growing methodology called Quantified Self. The goal of this section is provide a understanding of how a physical healthcare model can be created and integrated to a cyber model.

mHealth is a fast growing segment in Healthcare Information Technology (HIT) and is defined by the NIH as the use of mobile and wireless devices to improve health outcomes, healthcare services, and healthcare research. At the latest conference of Healthcare Information and Management Systems Society - HIMSS 2013, with increasing numbers of smart phones and other gadgets used by both patients and providers, mobile security was another big topic, particularly with regard to concerns about how to keep medical information secure, even while making it more available through mobile applications.

Tele-monitoring is a medical practice that involves remotely monitoring patients at distributed locations. Cafazzo et al [18] have demonstrated advances in information and communication technology and tele-monitoring that provide the capabilities of enhancing patient self-management, such as the use of Bluetooth-enabled medical devices to collect vital signs in the home. Such technologies allow the information gathered by patients to be made available to healthcare providers at the time of office visits or even earlier in a format that aids clinical-decision making.

The tremendous proliferation of HIT in the form of EMR, mHealth, telemedicine, remote patient monitoring (RPM), and the quantified self movement has created a complex Cyber Physical System with great potential to improve healthcare delivery and outcomes. The growing dependency on these CPS also exposes healthcare systems and ultimately the patients to the inherent risks of technology such as unscheduled downtime, outages, denial of services, malware, data breach, data corruption, and espionage.

III. ESTABLISHING A COMBINED GENERATIVE- DISCRIMINATIVE BEHAVIOR MODEL FROM AGGREGATE BEHAVIORS FOR CPS

Semantically aligning observations from heterogeneous CPS elements is a challenge, along with choosing what layers in our model to best align them [Figure 2]. There has been recent work that focuses on the concept of probabilistic networks networking described as a network of networks [17]. Our work would first focus on transforming observations into a behavior feature space and then align.

A. Discriminative Behavior Model

In the past [9], we have show a methodology supported by a discriminative model that establishes the creation of behavior primitives. In this work we build off of this work to establish a model that facilitates the creation of behavior graphs.

![Figure 2. Notionally Combining Behaviors Primitives transformed from Observations from Two Different domains using aggregate primitives](image)

The overall system [12] (See Figure 2) focuses on the notion of tracking various network objects, $O$, e.g. hosts, host-groups, and networks, and determining if they are threats. Tracking these objects involves collecting events and data from a number of different network sensors, e.g., network flow, NIDS, honeypots, and creating a sample space. Layers 1-3 are focused on the use of discriminative supervised/semi-supervised models to identify behavior primitives.

In our current data fusion system, network flow data and alerts generated by network sensors reflect the totality of information and model’s sample space, $S$, available to the detection system regarding the objects to be analyzed.

To utilize this data, it is first normalized and transformed into a representation that is conducive to algorithmic processing. The fusion engine operates over a sample space denoted as $S$ representing sensor data. This fusion operation is represented by an behavioral analysis function, $B$.

A function measuring the aggregate behavior of the sample for a specific object $O = f(S_O)$, produces a feature characteristic, or behavior, for that object denoted by $F_O$ accumulated within a set time window $F_{tw,O}$. The sample space, $S$, is then transformed into an aggregate feature space $F_S$. The Time window, $tw$, consists of periods such as hour, day, month, or year. $F_O$ is represented by a n-tuple, or n-gram, of individual time-based features, for example $F_{month,O} = \langle f_1,f_2,...,f_n \rangle$, describes $O$ over a period of a month. These features consist of structural, behavioral, and/or application specific properties of $O$ over a given time period.

1 http://obssr.od.nih.gov/scientific_areas/methodology/mhealth/
A behavior primitive for Object \( O \) is denoted by \( B_o \) and is represented by a single feature characteristic, or an \( n \)-tuple of feature characteristics.

B. A Generative Behavior Model

We assume that any given networked object \( O \) (e.g. host) is represented by behavior graphs \( G \) containing one or more sub-graphs, where \( G = \langle g_1, g_2, ..., g_n \rangle \). Each behavior graph is represented by a set of behavior primitives’ \( B \).

We will use the Distributed Hash Table protocol (DHT) in the context of a P2P architecture as a method of reference to facilitate the description of model development (Wiley, 2003). We choose this example because this protocol is used in both Bittorrent, and also the Storm botnet, where it manages hosts joining an peer-to-peer network. When a DHT Command and Control (C2) primitives are discovered on a network they may mean either that it is normal behavior activity of a Bittorrent peer, or that there are infected workstations with a botnet. Another discriminating set of primitives, \( G \), are then needed to discern between normal and abnormal activity of the object.

In the Chord implementation of DHT each node keeps a reference to the next and previous nodes in a list, and the addresses of other peer nodes in a list. The ordering in the list is defined by the next machine smallest distance clockwise away. We have measured DHT C2 of Bittorrent (Figure 3), and would posit a common pattern of behavior in communication in like DHT implementations.

In Figure 3 we propose two behavior graphs that can be used to notionally discern between Bittorrent and STORM. We highlight the use of the following behavior primitives: data upload, data download, DHT C2, and Unique Web POST found within the STORM botnet after a machine has become infected.

We will now formalize our classification methodology before we describe the generative behavior model.

We are given a set of behavior primitives \( B = \{b_1, b_2, ..., b_l\} \) where \( b_i \) represents a single primitive for a given object (e.g. host), and a set of corresponding class labels \( C_x = \{c_1, c_2, ..., c_l\} \) where each \( c_i \) represents a particular labeled behavior. The suffix numbers, \( 1,2,3 \ldots L \), correspond to the passage of time and other similarity measures used to assemble the graph (e.g. # of flows, repeats visits of host to network). The primitives were identified using various discriminative methods described in previous section. Moreover, we assume there is a set \( Y = \{x_{L+1}, x_{L+2}, ..., x_{L+U}\} \) of unlabeled behavior primitives. Each primitive, \( x_i \) consists of various measurements with \( x_{ij} \) denoting the \( j \)th feature of behavior primitive \( i \).

Our goal is to define various classifiers \( f(\cdot) \) using the sets \( B, Y, C_x \) such that for any feature \( x^* \) we can predict a good class label \( f(x^*) \). The classifier would provide insight into its behavior of CPS elements. Changes in behaviors can then be realized as either normal or abnormal behaviors using a probabilistic approach.

Using a graphical model formalism it is represented by: first defining a probability distribution over all known and unknown variables in our sample space and, second, we specify a rule which says how the probability distribution relates to our decision rule \( f \).

C. Behavior Graph Features

This section describes how the features of a behavior graph represent characteristics of a network object, \( O \), representing both normal and abnormal behaviors. Graph theory provides a notion of centrality of a vertex as a measure of importance. We will briefly review how In-degree/out-degree of nodes, path and cycle measurements can be used to characterize a behavior graph, and its sub-graphs.

One of the simple measurements of a behavior graph is on degree centrality. The in-degree and out-degree are measurements of the number of links within a behavior graph. The in-degree of a behavior graphs is count of links going into a behavior primitive, and out-degree is the link count of connection to other nodes from a single node. The degree centrality measures provide our work with insight into common behaviors of CPS element.
Figure 5. Behavior graph of primitives identified within a PHMS that are associated with a specific patient. The blue primitives are cyber behavior primitives, green are physical behavior primitives associated with monitoring and control devices, and the orange are health status primitives.

The path within a behavior graph provides the narrative, or topic map, of behavior primitives for a CPS element. These behavior paths provide insight into the common behavior subgraphs found within the element. One interesting feature within the graph is the existence of behavior cycles. These cycles represent behavior patterns that are repetitive within the behavior graph. Cycles can be reduced into a single node for further aggregation.

IV. SUMMARY AND FUTURE WORK

We proposed a novel combined determinative and generative approach for modeling a CPS. We claim that such an approach enables efficient modeling of interactions between heterogeneous entities and yet allows us to leverage on the generative models like graphical analysis. We presented a unique multi-layered model to represent our approach. Building a framework to extract behavior primitives from multiple domains is the next logical step for our work. Discovering dependencies across these domains, developing a methodology that is used to construct meaningful graphs, and probabilistic models that are used to identify threats in CPS, specifically PHMS are candidates for our future work.

Acknowledgment

Sonalysts would like to acknowledge support from of the Cyber Security Program Area of the Command, Control and Interoperability Division within the Science and Technology Directorate of the U.S. Department of Homeland Security, especially the support from Dr. Douglas Maughan.

REFERENCES