Abstract—Today's embedded and cyber-physical systems are ubiquitous. A large number of critical cyber-physical systems have real-time requirements (e.g., avionics, automobiles, power grids, manufacturing systems, industrial control systems, etc.). The current trend is to connect real-time embedded devices to the Internet. This gives rise to the real-time Internet-of-things (RT-IoT) that promises a better user experience through stronger connectivity and better use of next-generation embedded devices, albeit with safety-critical properties. However RT-IoT are also increasingly becoming targets for cyber-attacks as evident by recent events. This paper gives an introduction to RT-IoT systems, an outlook of current approaches and possible research challenges towards a holistic secure RT-IoT framework.

I. INTRODUCTION

Nowadays smart embedded devices (e.g., surveillance cameras, home automation systems, smart TVs, in-vehicle information systems, etc.) are connected to the Internet—the rise of Internet-of-things (IoT) links together the devices/applications that were previously isolated. On the other hand, embedded devices with real-time properties (e.g., strict timing and safety requirements) require interaction between cyber and physical worlds. These devices are used to monitor and control physical systems and processes in many domains, e.g., manned and unmanned vehicles including aircraft, spacecraft, unmanned aerial vehicles (UAVs), self-driving cars; critical infrastructures; process control systems in industrial plants; smart technologies (e.g., electric vehicles, medical devices, etc.) to name just a few. Given the drive towards remote monitoring and control, these devices are being increasingly interconnected, often via the Internet giving rise to the Real-Time Internet-of-things (RT-IoT). Since many of these systems have to meet stringent safety and timing requirements, any problems that deter from the normal operation of such systems could result in damage to the system, the environment or pose a threat to human safety. The drive towards remote monitoring and control facilitated by the growth of the Internet, the rise in the use of commercial-off-the-shelf (COTS) components, standardized communication protocols and the high value of these systems to adversaries are making cyber-security a design priority. A number of high-profile attacks on real RT-IoT systems, e.g., denial-of-service (DoS) attacks mounted from IoT devices (i.e., cameras) [1], Stuxnet [2], BlackEnergy [3], attack demonstrations by researchers on automobiles [4], [5] and medical devices [6] have shown that the threat is real. Successful cyber attacks against such systems could lead to problems more serious than just loss of data or availability because of their critical nature [5], [7]. Any successful, serious, attacks on one or more of these types of systems can have catastrophic results, leading to loss of injury to humans, negative impacts on the system and even the environment.

Enabling security in RT-IoT is often more challenging than generic IoT due to additional real-time constraints imposed by real-time-enabled systems. The focus of this paper is to introduce the properties/constraints and security threats for RT-IoT (Sections III–IV), summarize security solutions specially designed for such safety-critical domains (Section V) and highlight the research challenges (Section V-A). While there exist some surveys [8]–[12] on security and privacy issues in general-purpose IoT systems, to the best of our knowledge, there is no comprehensive summary in the context of RT-IoT security.

II. REAL-TIME INTERNET-OF-THINGS: AN OVERVIEW

At their core, RT-IoT largely intersect with real-time cyber-physical systems [13]. RT-IoT can be considered as a wide inner-connected network, in which nodes can be connected and controlled remotely. Table I summarizes some of the common properties/assumptions related to RT-IoT systems. In this section, we intend to outline the elements of RT-IoT as well as the scope of security covered in this paper. Figure 1 gives some common scenarios where RT-IoT applications can be implemented.

A. Stringent Timing/Safety Requirements and Resources Constraints

Many RT-IoT devices (e.g., sensors, controllers, UAV, autonomous vehicles, etc.) will have severely limited resources
(e.g., memory, processor, battery, etc.) and often require control tasks to complete within a few milliseconds [14]. RT-IoT nodes apart from a requirement for functional correctness, require that temporal properties be met as well. These temporal properties are often presented in the form of deadlines. The usefulness of results produced by the system drops on the passage of a deadline. If the usefulness drops sharply then we refer to the system as a hard real-time system (e.g., avionics, nuclear power plants, anti-lock braking systems in automobiles, etc.) and if it drops in a more gradual manner then they are referred to as soft real-time systems (e.g., multimedia streaming, automated windshield wipers, etc.).

| TABLE I: PROPERTIES OF MAJORITY RT-IoT NODES |
|------------------------------------------------|
| • Implemented as a system of periodic/sporadic tasks |
| • Stringent timing requirements |
| • Worst-case bounds are known for all loops |
| • No dynamically loaded or self modified codes |
| • Recursion is either not used or statically bounded |
| • Memory and processing power is often limited |
| • Communication flows with mixed timing criticality |

B. Heterogeneous Communication Traffic

Many conventional RTS typically consist of several independently operating nodes with limited or no communication capabilities. However with the emergence of RT-IoT, cyberphysical nodes not only communicate over closed industrial communication networks but are also often connected via the Internet. As an example, an autonomous vehicle is estimated to generate data at a rate of about 1 gigabyte per second while the U.S. smart grid is expected to generate 1000 petabytes of data each year (compared to that 2.4 petabytes of data generated by the U.S. Library of Congress per month) [14]. Since most real-time applications would need to trigger events based on specific data conditions, a real-time communication channel with guaranteed QoS (e.g., throughput and data processing requirements, delay guarantees, etc.) is necessary to support such applications [15, 16].

Another property of RT-IoT is that they often include traffic flows with mixed criticality, i.e., those with varying degrees of timing (and perhaps even bandwidth and availability) requirements: (a) high priority/criticality traffic that is essential for the correct and safe operation of the system; examples could include sensors for closed loop control and actual control commands in avionics, automotive or power grid systems; security systems in home automation (b) medium criticality traffic that is critical to the correct operation of the system, but with some tolerances in delays, packet drops, etc.; for instance, navigation systems in aircraft, system monitoring traffic in power substations, communication messages exchanged between electric vehicles and power grid or home charging station, traffic related to home automation equipments such as water sprinklers, heating, air conditioning, lighting devices, food preparation appliances etc.; (c) low priority traffic – essentially all other traffic in the system that does not really need guarantees on delays or bandwidth such as engineering traffic in power substations, multimedia flows in aircraft, notification messages from smart home equipments, etc. Typically, in many safety-critical RT-IoT, the properties of all high-priority flows are well known, while the number and properties of other flows could be more dynamic (e.g., consider the on-demand video situation where new flows could arise and old ones stop based on the viewing patterns of passengers).

C. Real-Time Scheduling Model

Most RT-IoT systems are implemented using a set of periodic (e.g., fixed temporal septation between consecutive instances) or sporadic (e.g., the tasks that can make an execution request at any time, but with a minimum inter-execution interval) tasks [17, Ch. 1] [18]. For instance, a sensor management task that monitors the conveyor belt in a manufacturing system needs to be periodic but the tasks that monitor the arrival of automated cars at traffic intersections are sporadic. Application tasks in the RT-IoT nodes are often designed based on the Liu and Layland model [19, 20] that contains a set of tasks, \( \Gamma \) where each task \( \tau_i \in \Gamma \) has the parameters: \( (C_i, T_i, D_i) \), where \( C_i \) is the worst-case execution time (WCET), \( T_i \) is the period or minimum inter-arrival time, and \( D_i \) is the deadline, with \( D_i \leq C_i \). Schedulability tests [21–23] are used to determine if all tasks in the system meet their respective deadlines. If they do, then the task set is deemed to be ‘schedulers’ and the system, safe.

D. CPU Architecture and System Development Model

Despite the fact that most RT-IoT applications are designed using platforms equipped with a single-core CPU, the trend towards multi-core systems can be seen as many COTS devices nowadays are built on top of multi-core environment [22]. For some specific applications (e.g., avionics systems), there exist regulations that restrict the use of additional cores. In such cases, the additional cores that do not execute real-time or safety critical tasks can be utilized to provide layers of security to the system. We have leveraged the use of multi-core platforms in the real-time domain and developed security solutions [24–29] as we discussed in Section IV-A.

It is also common that multiple vendors are involved in the development of RT-IoT systems. Such a system is said to be developed under the multi-vendor development model [30]. In this model, each vendor designs/controls several separate tasks. Figure 4 demonstrates an electronic control unit (ECU) for an avionics system that uses the multi-vendor development model. While this demonstrative example focuses on avionics domain other RT-IoT systems (e.g., automotive, home automation, etc.) could also be created using a similar model (albeit loosely defined).

III. SECURITY THREATS FOR RT-IoT

RT-IoT systems face threats in various forms depending on the system and the goals of an adversary. In a system developed using vendor-based model, one of the involved
vendors can act maliciously. This (potentially unverified/untrusted) vendor could embed malicious functions in its tasks. In a system that has network connectivity, the adversary could target the communication interfaces. Due to a lack of authentication many of these systems the communication channels could easily be intercepted and forged.

A. Attacks on RT-IoT

We classified the attack methodologies on RT systems based on the control over computation process and the functional objective of the attack. One way to acquire control over target system could be injecting malicious code (e.g., malware) or by reusing legitimate code for malicious purposes (e.g., code-injection attacks). Besides, since most RT-IoT nodes communicate over unreliable medium such as Internet, the system is also vulnerable to network-level attacks. Other than trying to aggressively crash the system (e.g., using DoS attacks) the adversary may silently lodge itself in the system and extract sensitive information (e.g., side-channel attacks). The side-channel attacks are based on observing properties (e.g., execution time, memory usage patterns, task schedule, power consumption, etc.) of the system. These information may later be used by the attacker to launch further attacks. We summarize the common attack surfaces for RT-IoT systems.

1) Integrity Violation with Malicious Code Injection: An intelligent adversary can get a foothold in the system. For example, an adversary may insert a malicious task that respects the real-time guarantees of the system to avoid immediate detection and/or compromise one or more existing real-time tasks. The attacker may use such a task to manipulate sensor inputs and actuator commands (for instance) and/or modify system behavior in undesirable ways. Integrity violation through code injection attacks conceptually consist of two steps [31]. First, the attacker sends instruction snippets (e.g., a valid machine code program) to the device which is then stored somewhere in memory by the software application receiving it. Such instruction snippets are referred to as gadgets. In the second step, the attacker triggers a vulnerability in the application software, i.e., real-time OS (RTOS) or task codes to divert the control flow. Since the instruction snippets represents a valid machine code program, when the program execution jumps to the start address of the data, the malicious code is executed. As we illustrate in Section IV our recent solutions [24]–[29], [32] can be used to detect integrity violations through various hardware/software mechanisms.

2) Side-Channel Attacks: The adversary may learn important information by side or covert-channel attacks [33] by simply lodging themselves in the system and extracting sensitive information. A side-channel attack manipulates unexpected channels to acquire useful information from the victim. Memory access patterns, cache access time, power consumption traces, schedule preemption, etc. are examples of some typical side-channels used by attackers. These channels as attack surfaces are particularly applicable attacking RT-IoT nodes that execute real-time tasks due to their deterministic behaviors. A demonstrative cache-timing attack is presented in Section III-B2 and Section IV-B1 illustrates our recent approaches [30], [34] to mitigate cache information leakage through side-channel attacks.

3) Attacks on Communication Channels: RT-IoT elevates the Internet as the main communication medium between the physical entities. However, Internet, as an insecure communication medium, introduces a variety of vulnerabilities that may put the security and privacy of RT-IoT systems under risk. Threats to communication includes eavesdropping or interception, man-in-the-middle attacks, falsifying, tampering or repudiation of control/information messages, etc. From the perspective of RT-IoT, defending against communication threats is not an easy task. This is because it is challenging to distinguish rogue traffic from the legitimate traffic (especially for the critical/high-priority flows) without degrading the QoS (e.g., bandwidth and end-to-end delay constraints). Threats to communications are usually dealt by integrating cryptographic protection mechanisms [35], [36]. However this increases the WCET of the real-time tasks and may require modification of existing schedulers. Therefore existing cryptographic approaches may not be a preferable option for many RT-IoT systems. In Section IV-B3 we illustrate a solution to integrate security mechanisms that can also be used for dealing with communication threats but does not require modification of existing real-time tasks.

4) Denial-of-Service (DoS) Attacks: Due to resource constraints (e.g., low memory capabilities, limited computation resources, etc.) and stringent timing requirements the RT-IoT nodes are vulnerable to DoS attacks. The attacker may take control of the real-time task(s) and perform system-level resource (e.g., CPU, disk, memory, etc.) exhaustion. A more severe type of the DoS attack is the distributed denial-of-service (DDoS) attack where a large number of malicious/compromised nodes simultaneously attack the physical plant. In particular, when critical tasks are scheduled to run, an advanced attacker may capture I/O or network ports and perform network-level attacks to tamper with the confidentiality and integrity (viz., safety) of the system. Again the defense mechanisms developed for generic IT or embedded systems do not consider timing, safety and resource constraints.
of RT-IoT and may not be adopted directly without significant modifications. As described in Section IV-A and IV-B, our recent work \cite{29}, \cite{32} may be used to defend against DoS attacks.

But first, in order for those attacks to be successful, reconnaissance is one of the early steps that an attacker needs to do. We illustrate this in the following.

B. Reconnaissance

Reconnaissance, essentially, is the first step for launching other successful attacks and, at the very least, the attacker gains important information about the system’s internals.

1) ScheduLeak: In initial work \cite{27}, we developed an algorithm, “ScheduLeak”, to show the feasibility of a schedule-based side-channel attack targeting real-time embedded systems with a multi-vendor development model introduced in Section II-D. The adversary could be one of the vendors or an attacker who compromises a vendor. The ScheduLeak algorithm utilizes an observer task that has the lowest priority in the victim system to observe busy intervals. A “busy interval” is a block of time when one or more tasks are executing – an adversary cannot determine what tasks are running when by just measuring or observing the busy intervals as they are.

The ScheduLeak algorithm can be represented as a function $R(\Gamma, W) = J$, where $W$ is a set of observed busy intervals and $J$ is the inferred schedule information that can be used to pinpoint the possible start time of any particular victim task. Such a function is illustrated by Fig. 3. By using the ScheduLeak algorithm, an attacker can deconstruct the observed busy periods (with up to 99% success rate if tasks have fixed execution times) into their constituent jobs and precisely pinpoint the instant when a task is scheduled.

![Fig. 3. An example of the schedules produced from a task set of three tasks. The ScheduLeak algorithm can recover the precise schedules from the observed busy intervals.](image)

2) Targeted Attacks: It’s worth mentioning that the effectiveness of side-channel attacks is enlarged when combined with the reconnaissance step we just introduced. For example, in the demonstrative ECU system introduced in Section II-D, let us assume Vendor 2 (as an adversary) would like to identify whether the surveillance camera controlled by the I/O Operation Task is enabled. The attacker can launch a ScheduLeak algorithm to infer exact start times of the I/O Operation Task and carry out a cache-timing attack to gauge cache usage when I/O Operation Task is scheduled. Figure 4 shows the result of such a cache-timing attack – this indicates that the attacker is able to identify the instant when the camera is on (i.e., when a large amount of data is processed by I/O Operation Task).

![Fig. 4. A demonstration of a cache-timing attack. The X-axis is sample points and Y-axis shows both cache usage inference (round dots) and real memory usage amount (the solid line). It shows that a successful cache-timing attack can precisely infer the memory usage of the victim task.](image)

IV. SECURING RT-IOT: HOST-BASED APPROACHES

In what follows we summarize our initial attempts to provide security in the RT-IoT nodes. We refer to these approaches as host-based solutions since they are primarily focusing on securing an individual RT-IoT node. These approaches can be classified into two major classes: (i) solutions that require custom hardware support to provide security and (ii) the solutions at the scheduler/software level that do not require any architectural modifications. Table I summarizes security mechanisms for RT-IoT systems.

A. Security with Hardware Support

The key idea of providing security without compromising the safety of the physical system is built on the Simplex framework \cite{39}. Simplex is a well-known real-time architecture that utilizes a minimal, verified controller (e.g., safety controller) as backup when the complex, high-performance controller (e.g., complex controller) is not available or malfunctioning. The goal of Simplex method is to guarantee that under any behavior of the complex subsystem, the physical system will remain safe. We have used the idea of Simplex in the context of RT-IoT security \cite{24}, \cite{27}, \cite{29}. The key concept of using Simplex-based architecture for security is to use a minimal simple subsystem (say a trusted core) to monitor the properties (i.e., timing behavior \cite{24}, \cite{25}, memory access \cite{26}, system call trace \cite{27}, behavioral anomalies \cite{29}, etc.) of an untrusted entity (e.g., monitored core) that is designed for more complex tasks and/or exposed to less secure medium (e.g., network, Internet, I/O channels, etc.).

1) Secure System Simplex Architecture (S3A): As mentioned in Section I the worst-case, best-case and average-case behaviors for most RT-IoT nodes are calculated ahead of time to ensure that all resource and schedulability requirements will be met during system operation. S3A \cite{24} utilizes this knowledge of deterministic execution profile of the system and use to detect the violation of predicted (e.g., uncompromised) system behavior. S3A is one of our earliest efforts to use another (FPGA-based, in this case) trusted hardware
component that monitors the behavior (e.g., execution time and the period) of a real-time control application running on an untrustworthy main system. The goal of this Simplex-based architecture is to detect an infection as quickly as possible and then ensure that the physical system components always remain safe. Using an FPGA-based implementation and considering inverted pendulum (IP) as the physical plant we demonstrated that S3A can detect intrusions in less than 6 µs without violating safety requirements of the actual plant.

2) SecureCore Framework: As illustrated in Fig. 5 the idea of SecureCore architecture is to utilize a trusted entity (e.g., a secure core) that can continuously monitor the system behavior (e.g., code execution pattern [25], memory usage [26], system call trace [27]) of a real-time application on an untrustworthy entity (e.g., monitored core). The initial SecureCore architecture [25] uses a statistical learning-based mechanism for profiling the correct execution behavior of the target system and careful analysis based on these profiles leads to detect malicious code execution. The SecureCore framework is also extended [26] to profile memory behavior and then detect deviations from the normal memory behavior patterns. We have also designed SecureCore architecture to detect anomalous executions using a distribution of system call frequencies. Specifically we have proposed [27] to use clustering algorithms (e.g., global k-means clustering with the Mahalanobis distance) to learn the legitimate execution contexts (by means of distribution of system call frequencies) of real-time applications and then monitor them at run-time to detect intrusions.

3) Control Flow Monitoring: In literature [28] we propose a hardware-based approach for checking the integrity of code flow of real-time tasks. In particular, we had proposed to add an on-chip control flow monitoring module (OCFMM) with a dedicated memory unit that directly hooks the processor and track the control flow of the tasks. The control flow graph (CFG) of tasks is produced from the program binary and loaded into the OCFMM memory in advance (e.g., during system boot). The detection module inside OCFMM compares the control flow of the running program with the stored one (e.g., CFG profiles that are loaded into the dedicated memory at boot time) during program execution. At run-time (e.g., during execution of a given block) CFG profiles for the next-possible blocks are pre-fetched. The decision module continuously scans the current block and validates the execution flow by comparing the current address of the program counter (PC) against the possible, previously fetched destination addresses. If any mismatch occurs, the detection module raises a detection flag that indicates a possible breach.

4) Security by Platform Reboot: In traditional computing systems (e.g., servers, smart phones, etc.), software problems are often resolved by restarting either the application process or the platform [40], [41]. However, unlike those conventional computing systems restart-based recovery mechanisms are not straightforward in RT-IoT due to the real-time constraints as well as interactions of the control system with the physical world (for example, a UAV can quickly be destabilized if its controller restarts). In initial work [29] we propose a restart-based concept to improve security guarantees for RT-IoT. This Simplex-based framework, that we refer to as ReSecure, is specifically designed to improve security of safety-critical RT-IoT systems. In particular, we propose to restart the

| Reference | Approach | Defense Surface | Overhead/Costs |
|-----------|----------|----------------|---------------|
| [24]–[28] | Use verified/secure hardware module to monitor system behavior (e.g., timing [25] and execution pattern [24], memory access [26], system call usage [27], control flow [28]) | Code injection attacks | Require custom hardware or monitoring unit |
| [29]      | Periodically and/or asynchronously (e.g., upon detection of a malicious activity) restart the platform and load an uncompromised OS image | Code injection, side channel and DoS attacks | Extra hardware to ensure safety during periodic/asynchronous restart events |
| [30]–[34] | Flush the shared medium (e.g., cache) between the consecutive execution of high-priority (security sensitive) and low-priority (potentially vulnerable) tasks | Side-channel (cache) attacks | Overhead of cache flushing reduces task-set schedulability |
| [38]      | Randomize the task execution order (i.e., schedule) to reduce the predictability | Side-channel attacks | Extra context switch |
| [32]      | Execute monitoring/intrusion detection tasks with a priority lower than real-time task to preserve the real-time task parameters (e.g., period, WCET and execution order) | Code injection, side-channel, DoS and/or communication attacks depending on the what monitoring tasks are used | Running security task with lower priority may cause longer detection time due to high interference (e.g., preemption) from real-time tasks |
platform periodically/asynchronously and load a fresh image of the applications and RTOS from a read-only media after each reboot with the objective of wiping out the intruder or malicious entity.

The main idea of restart-based protection is that, if we restart the system frequently enough, it is less likely that the attacker will have time to become re-enter the system and cause meaningful damage to the system. After every restart, there will be a predictable down time (during the system reboots), some operational time (before system is compromised again) and some compromised time (until the compromise is detected or the periodic timer expires). The length of each one of the above intervals depends on the type and configuration of the platform, adversary models, complexity of the exploits, etc.. As a general rule, the effectiveness of the restarting mechanism increases: (i) as the time to re-launch the attacks increases, or (ii) the time to detect attacks and trigger a restart decreases. We develop a probabilistic method to evaluate the expected lack of availability due to restarts and the expected damage from the attacks/exploits given a certain restart configuration. We also show how to find an optimal configuration to balance lack of availability and damage.

B. Security without Architectural Modifications

Despite the fact that architectural modification can improve the security posture of RT-IoT nodes, those approaches require an overall redesign and may not be suitable for systems developed using COTS components. We now review the some of the approaches we recently proposed to enhance security in RT-IoT without any custom hardware support.

1) Dealing with Side-Channel Attacks: As introduced in Section III-B2, we demonstrated that an attacker can carry out a cache-timing attack to indirectly estimate memory usage behavior. It is due to the lack of isolation for shared resources across different tasks in most COTS-based RT-IoT systems. The overlap between tasks happens when the system transitions from one task to another. Therefore, capturing security constraints between tasks becomes essential for preventing side-channel attacks.

In previous work [34], we proposed to integrate security in RT-IoT by introducing techniques to add constraints to tasks scheduled with fixed-priority real-time schedulers. Based on user-defined security levels for each task, the scheduler flushes shared cache when the system is transitioning from a high security task (i.e., a task demanding higher confidentiality) to a low security task (i.e., an insecure task potentially compromised). Let us consider the set of security levels for tasks, $S$, forms a total order. Hence, any two tasks $(\tau_i, \tau_j)$ may have one of the following two relationships when considering their security levels, $s_i, s_j \in S$: (i) $s_i \prec s_j$, meaning that $\tau_i$ has higher security level than $\tau_j$ or (ii) $s_j \prec s_i$.

We proposed the idea of mitigating information leakage among tasks of varying security levels, by transforming security requirements into constraints on scheduling algorithms. The approach of modifying or constraining scheduling algorithms is appealing because, (a) it is a software based approach and hence easier to deploy compared to hardware based approaches and (b) it allows for reconciling the security requirements with real-time or schedulability requirements. Consider a simple case with two periodic tasks, a high priority task $H$ and a low priority task $L$ scheduled by a fixed-priority scheduling policy. Assume that $s_H \prec s_L$: hence, information from $H$ must not leak to $L$. These tasks must be scheduled on a single processor, $P$, so that both deadlines $(D_H, D_L)$ are satisfied. If $L$ (or any part thereof) executes immediately after (any part) or all of $H$, then $L$ could “leak” data from resources recently used by $H$. The main intuition is that a penalty must be paid for each shared resource in the system every time tasks switch between security levels. In this case, the cache must be flushed before a new task is scheduled. Hence, we proposed the use of an independent task, called the Flush Task for this purpose.

In subsequent work [30], we relaxed many of the restrictions (e.g., the requirement of total ordering of security levels) and proposed a new, more general model to capture security constraints between tasks in a real-time system. This includes the analysis for the schedulability conditions with both pre-emptive and non-preemptive tasks. We proposed a constraint named noleak to capture whether unintended information sharing between a pair of tasks must be forbidden. Using this constraint we can prevent the information leakage via implicitly shared resources. For any two tasks $\tau_i$ and $\tau_j$: if noleak($\tau_i, \tau_j$) = True, then information leakage from $\tau_i$ to $\tau_j$ must be prevented; if noleak($\tau_i, \tau_j$) = False, no such constraints need to be enforced. We showed that the system remains schedulable (e.g., all the tasks can meet their deadline) under the proposed constraints without significant performance impact.

2) Schedule Randomization: One way to protect a system from certain attacks (e.g., the schedule-based side-channel attack mentioned in Section III-B1), is to randomize the task schedule to reduce the deterministic nature of periodic RT-IoT applications. By randomizing the task schedules we can enforce non-determinism since every hyper-period will show different order (and timing) of execution for the tasks. Unlike traditional systems, randomizing task schedules in RT-IoT is not straightforward since it leads to priority inversions [42] that, in turn, may cause missed deadlines – hence, putting the safety of the system at risk.

Hence, we proposed TaskShuffler [38], a randomization protocol for fixed-priority scheduling algorithm, to achieve such randomness in task schedule. For instance, by picking a random task instead of the one with the highest-priority at each scheduling point, subject to the deadline constraints. The degree of randomness is flexible in TaskShuffler. Based on the system’s needs, TaskShuffler implements the following randomization schemes:

- Randomization (Task Only): This is the most basic form of randomization in contrast to other schemes introduced
below. We randomly pick a task to execute whenever a task arrives or finishes its job, i.e., at the scheduling points. The effectiveness against the schedule-based side-channel attack is limited since the busy intervals in this scheme remains the same.

- **Randomization with Idle Time Scheduling**: In addition to the randomness provided in the basic scheme, we include the idle task (e.g., the dummy task executed by an RTOS when other real-time tasks are not running) at each scheduling point. It eliminates the periodicity of busy intervals (from hyper-period's point of view). This scheme makes it harder to produce effective results from the schedule-based side-channel attack.

- **Randomization with Idle Time Scheduling and Fine-grained Switching**: To push the randomization to an extreme, one could choose to randomize the schedule every tick. That is, the scheduler will randomly pick a task to execute, subject to the deadline constraints, in every tick interrupt. This way, we gain the most randomness for the schedule. However, it greatly increases the overheads and thus may not be applicable for all use cases.

3) **Integrating Security for Legacy RT-IoT**: As we have described in Section III-B1, an adversary can extract important information while still remaining undetected and it is essential to have a layered defense and integrated resilience against such attacks into the design of RT-IoT. However, any security mechanisms have to co-exist with real-time tasks in the system and have to operate without impacting the timing and safety constraints of the control logic. Besides, the embedded nature of these systems limits the availability of computational power (e.g., memory or processor) required for resource-extensive monitoring mechanisms. This creates an apparent tension between security requirements (e.g., having enough cycles for effective monitoring and detection) and the timing and safety requirements. For example, a critical parameter is to determine how often and how long should a monitoring and intrusion detection task execution to be effective but not interfere with real-time control or other safety-critical tasks. While this tension could potentially be addressed for newer systems at design time, this is especially challenging for retrofitting legacy systems where the control tasks are already in place and perhaps cannot be modified. Any hardware and/or software-level modifications to those legacy system parameters is costly since it will go through several verification and validation steps and may increase system downtime [14].

Recent efforts [35], [36], [43] as well as the frameworks presented above (e.g., [24]–[28], [30], [34], [38]) either require custom hardware [24]–[29], [43], modification of the existing schedulers [35], [36], extra instrumentations [43] or may need to change the tasks parameters (e.g., execution order and/or run-time) [30], [34], [38] and therefore not suitable for legacy systems. Architectural frameworks [24]–[29], [45] that use hardware/software mechanisms to protect against security vulnerabilities assume the availability of hardware-support for monitoring. Integrating monitoring and detection tasks for RT-IoT without custom hardware support is an open problem.

Given the tension between security and timing requirements, while integrating security mechanisms into a practical system, finding the frequency of execution of the monitoring tasks is an important design parameter that trades security requirements with timing constraints. If the interval between consecutive monitoring events is too large, the adversary may harm the system (and remain undetected) between two invocations of the security task. In contrast, if the security tasks are executed very frequently then it may impact the schedulability of the real-time tasks.

In preliminary work [52] we address the problem of determining the frequency of execution (e.g., periods or inter-monitoring interval) of the security tasks. Our approach to integrate security without perturbing real-time scheduling order is to execute security tasks at a lower priority tasks than real-time tasks. We refer this scheme as opportunistic execution since the security tasks are only allowed to execute opportunistically only during slack times when no other real-time tasks are running.

We propose to measure the security of the system by means of the achievable periodic monitoring. Let $T_i$ be the period of the security task that needs to be determined. Since our goal is to minimize the perturbation between the achievable (i.e., unknown) period $T_i$ and the desired period $T_{des,i}$, we define the following metric: $\eta_i = \frac{T_{des,i}}{T_i}$ that denotes the tightness of the frequency of periodic monitoring for the security task $\tau_i$. Thus $\eta = \sum_{\tau_i \in \Gamma_S} \omega_i \eta_i$ denotes the cumulative tightness of the achievable periodic monitoring for a set of security tasks $\Gamma_S$ where $\omega_i$ is the designer provided weighing factor that may reflect criticality or severity of the security tasks. This monitoring frequency metric provides one way to trade-off security with schedulability – if the interval between consecutive monitoring events is too large, an adversary may remain undetected and harm the system between two invocations of the security task, on the other hand, a very frequent execution of security tasks may negatively impact the schedulability of the real-time tasks. Hence, to find the desired $\eta$ of the target system we formulate a constraint optimization problem and also developed a polynomial-time solution that allows us to execute security routines with a frequency closer to the desired values while respecting the temporal constraints of the other real-time tasks.

V. DISCUSSION AND RESEARCH OPPORTUNITIES

A. Securing Legacy RT-IoT Systems

Since most RT-IoT nodes are resource-constrained embedded devices, resource-intensive processing and complex protocols (e.g., heavy cryptographic operations) for securing those systems is unrealistic and may threaten the safety of such systems – for instance a safety-critical task may miss deadline in order to run computation-heavy security tasks. In addition to execution frequency, another important consideration is to determine how quickly can intrusions be detected. Thus responsiveness vs. schedulability of critical tasks is another...
important trade-off. This in itself is a research challenge that needs to be investigated. For many legacy systems, altering the execution order of the real-time tasks is not straightforward due to control/safety requirements. In such cases, a useful solution to provide better responsiveness (against anomalous activities) would allow the security routines to execute in different modes (viz., passive monitoring by opportunistic execution as well as exhaustive checking with higher priority). Hence, security routines can execute opportunistically as before \cite{32} when the system is deemed to be normal (i.e., not compromised). However, if any anomaly or unusual behavior is suspected, the security policy may switch to fine-grained checking mode and execute with higher priority instead of waiting for its next slot. The security routines may go back to normal (e.g., opportunistic) mode if: (i) no anomalous activity is found; or (ii) the intrusion is detected and malicious entities are removed (or an alarm may be triggered if human intervention is required). However, the scheduling policy needs to ensure that the system remains functional and safe even with these mode changes. Also it is necessary to design security mechanisms that can operate in such modes/situations.

B. Security for Multi-Core based RT-IoT Platforms

Most of the work \cite{30,32,34,37} presented so far has been in the context of single core processors – they are the most common types of processors being used in RT-IoT systems. However as mentioned earlier, due to increasing computational demands, multi-core processors are becoming increasingly relevant to real-time systems \cite{22,44}. With the increased number of cores, more computation can be packed into a single chip – thus reducing power and weight requirements – both of which might be relevant to many RT-IoT systems. However multi-core processors can increase attack vectors, especially for side-channel attacks. First, two or more tasks are running together and (most likely) sharing low-level resources (e.g., last level caches). Hence, a task running on one core can snoop on the other – and not just when tasks follow each other. In fact, it has been shown that leakage can occur with a much higher bandwidth in the case of shared resources in multi-core processors \cite{45}. Second, when tasks execute together, a malicious task can increase the “interference” faced by a critical task – for instance, the malicious task can flood the cache/bus with memory references just when an important task (say, one that computes the control loop) is running. This could cause the critical task to get delayed and even miss its deadline. To prevent such problems, designers of the systems need to enforce constraints that protected tasks do not execute simultaneously with unprotected ones on the multi-core chip.

The problem of integrating security tasks into legacy RT-IoT systems is also interesting the multi-core context – perhaps the security tasks always be running (say on one of the dedicated cores) instead of running opportunistically as it is the case for single-core systems. Also it may be possible to take up more cores and execute fine-grained sanity check (e.g., a complete system-wide scan) as it detects malicious activity. Analyzing the impact of integrating security tasks in a multi-core legacy RT-IoT is an open problem worth investigating.

C. Secure Communication with Timing Constraints

With the rise of RT-IoT, the edge devices are more frequently exchanging control messages and data often with unreliable medium like the Internet. Therefore, in addition to the host-based approaches \cite{24,28,30,34,38} described earlier, there is a requirement for securing communication channels to ensure authenticity and integrity of control messages. While some of our previous work \cite{29,32} can also be used to deal with network-level attacks, designing a unified framework to protect edge devices as well as communication messages (given the stringent end-to-end delay requirements for high-critical traffics) is still an open problem.

Most safety-critical RT-IoT systems often have separate networks (hardware and software) for each of the different types of flows for safety (and security) reasons. This leads to significant overheads (equipment, management, weight, etc.) and also potential for errors/faults and even increased attack surface and vectors. Network-level nondeterminism, i.e., unpredictability in sensor reading, packet delivery/forwarding/processing further complicates the management of RT-IoT systems. Existing protocols, e.g., avionics full-duplex switched Ethernet (AFDX) \cite{46}, controller area network (CAN) \cite{47}, etc. that are in use in many of real-time domains are either proprietary, complex, expensive, require custom hardware or they are also exposed to known vulnerabilities \cite{48}.

Given the limitations of existing protocols, leveraging the benefits of software-defined networking (SDN) can also be effective for RT-IoT systems. The advantage of using SDN is that it is compatible with COTS components (and thus suitable for legacy RT-IoT systems) and provides a centralized mechanism for developing and managing the system. The global view is useful to ensure QoS (e.g., bandwidth and delay) and enforce security mechanisms (such as remote attestations, secure key/message exchange, remote monitoring, etc.). However current standards used in traditional SDN (e.g., OpenFlow \cite{49}) do not consider inherent timing and safety-critical nature of the RT-IoT systems. Retrofitting the capabilities of SDN in the RT-IoT domain requires further research.

VI. Related Work

There exists work that has investigated security in real-time systems \cite{35,36,50}. Many researchers have studied this research area from different aspects. Information leakage via side channels has been discussed in many works. Kadooru et al. \cite{51} and Gong et al. \cite{52} introduced analysis and methodology for quantifying side-channel leakage. Kelsey et al. \cite{53}, Osvik et al. \cite{54} and Page et al. \cite{55} demonstrated the usability of cache-based side-channels. Son et al. \cite{56} and Völp et al. \cite{57} examined the exploitation of timing channels in real-time scheduling.

While the work above focuses on exploring vulnerabilities, there exist work that aims to provide security to real-time systems. Ghassami et al. \cite{58} and Völp et al. \cite{59}
proposed techniques to address leakage via shared resources. Xie et al. [35] and Lin et al. [36] presented security in real-time systems by encrypting communication messages. Similar to the hardware-assisted security mechanisms like ours (e.g., S3A, SecureCore, etc.) researchers also propose architectural frameworks [43] that dynamically utilizes slack times (e.g., IoT systems.

infrastructures. Kuusijärvi [44] proposed a security framework that offers security solutions with smart infrastructures. Pacheco et al. [63] introduced a security threat that offers security solutions with smart infrastructures. Kuusijärvi et al. [44] proposed to mitigate IoT security threats with using trusted networks. Those work primarily focuses on generic IoT applications, and do not consider the additional real-time constraints required for RT-IoT systems.

VII. CONCLUSION

The sophistication of recent attacks on RT-IoT requires rethinking of security solutions for such systems. The goal of this paper is to raise the awareness of real-time security and bridge missing gaps in the current IoT context – securing the IoT systems with real-time constraints. The techniques and methodology presented here vary from different perspectives – from hardware-assisted security to scheduler-level as well as those for legacy systems. The designers of the systems and research community will now be able to integrate and develop upon these frameworks required to secure safety-critical RT-IoT systems. We believe that the real-time and IoT worlds are closely connected and will become inseparable in the near future.

ACKNOWLEDGEMENT

The multiple security frameworks presented in this paper is the result of a team effort. The authors would like to acknowledge the contributions from our team members and collaborators: Man-Ki Yoon, Fardin Abdi, Rodolfo Pellizzoni, Rakesh B Bobba, Lui Sha and Marco Caccamo.

REFERENCES

[1] J. Westling, “Future of the Internet of things in mission critical applications,” 2016.
[2] N. Falliere, L. O. Murchu, and E. Chien, “W32.stuxnet dossier,” White paper, Symantec Corp., Security Response, vol. 5, p. 6, 2011.
[3] R. M. Lee, M. J. Assante, and T. Conway, “Analysis of the cyber attack on the ukrainian power grid,” SANS Industrial Control Systems, 2016.
[4] K. Koscher, A. Czeskis, F. Roesner, S. Patel, T. Kohno, S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham et al., “Experimental security analysis of a modern automobile,” in IEEE S&P, 2010, pp. 447–462.
[5] S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, S. Savage, K. Koscher, A. Czeskis, F. Roesner, T. Kohno et al., “Comprehensive experimental analyses of automotive attack surfaces,” in USENIX Sec. Symp., 2011.
[6] S. S. Clark and K. Fu, “Recent results in computer security for medical devices,” in Mobihl, 2011, pp. 111–118.
[7] M. Abrams and J. Weiss, “Malicious control system cyber security attack case study–maroochy water services, australia,” McLean, VA: The MITRE Corporation, 2008.
[8] A.-R. Sadeghi, C. Wachsman, and M. Waidner, “Security and privacy challenges in industrial Internet of things,” in ACM/EDAC/IEEE DAC, 2015, pp. 1–6.
[9] I. B. Ida, A. Jemai, and A. Loukil, “A survey on security of IoT in the context of eHealth and clouds,” in IEEE IDIT, 2016, pp. 25–30.
[10] R. H. Weber, “Internet of things–new security and privacy challenges,” Comp. law & sec. rev., vol. 26, no. 1, pp. 23–30, 2010.
[11] G. A. Fink, D. V. Zarzhitsky, T. E. Carroll, and E. D. Farquhar, “Security and privacy grand challenges for the internet of things,” in IEEE CTS, 2015, pp. 27–34.
[12] S. Kraijak and P. Tuwanut, “A survey on IoT architectures, protocols, applications, security, privacy, real-world implementation and future trends,” in IET WICOM 2015, 2015, pp. 1–6.
[13] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, “Internet of things (iot): A vision, architectural elements, and future directions,” Future gen. comp. syst., vol. 29, no. 7, pp. 1645–1660, 2013.
[14] M. Chiang and T. Zhang, “Fog and iot: An overview of research opportunities,” IEEE IoT Journal, vol. 3, no. 6, pp. 854–864, 2016.
[15] Y. Ge, X. Liang, Y. C. Zhou, Z. Pan, G. T. Zhao, and Y. L. Zheng, “Adaptive analytic service for real-time internet of things applications,” in IEEE ICWS, 2016, pp. 484–491.
[16] J.-E. Kim, T. Abdelraheem, L. Sha, A. Bar-Noy, R. Hobbs, and W. Dron, “On maximizing quality of information for the internet of things: A real-time scheduling perspective,” in IEEE RTCSA, 2016, pp. 202–211.
[17] G. Battazzo, Hard real-time computing systems: predictable scheduling algorithms and applications, 2011, vol. 24.
[18] A. K. Mok, “Fundamental design problems of distributed systems for the hard-real-time environment,” Massachusetts Institute of Technology, Tech. Rep., 1983.
[19] C. L. Liu and J. W. Layland, “Scheduling algorithms for multiprogramming in a hard-real-time environment,” JACM, vol. 20, no. 1, pp. 46–61, 1973.
[20] R. I. Davis, “A review of fixed priority and edf scheduling for hard real-time uniprocessor systems,” ACM SIGBED Review, vol. 11, no. 1, pp. 8–19, 2014.
[21] N. Audsley, A. Burns, M. Richardson, K. Tindell, and A. J. Wellings, “Applying new scheduling theory to static priority pre-emptive scheduling,” SE Journal, vol. 8, no. 5, pp. 284–292, 1993.
[22] R. I. Davis and A. Burns, “A survey of hard real-time scheduling for multiprocessor systems,” ACM CSUR, vol. 43, no. 4, p. 35, 2011.
[23] E. Bini and G. C. Battazzo, “Schedulability analysis of periodic fixed priority systems; IEEE Trans. on Comp., vol. 53, no. 11, pp. 1462–1473, 2004.
[24] S. Mohan, S. Bak, E. Bettii, H. Yun, L. Sha, and M. Caccamo, “S3A: Secure system simplex architecture for enhanced security and robustness of cyber-physical systems,” in ACM HiCoNS, 2016, pp. 65–74.
[25] M.-K. Yoon, S. Mohan, J. Choi, J.-E. Kim, and L. Sha, “SecureCore: A multicore-based intrusion detection architecture for real-time embedded systems,” in IEEE RTAS, 2013, pp. 21–32.
[26] M.-K. Yoon, S. Mohan, J. Choi, and L. Sha, “Memory heat map: anomaly detection in real-time embedded systems using memory behavior,” in ACM/EDAC/IEEE DAC, 2015, pp. 1–6.
[27] M.-K. Yoon, S. Mohan, J. Choi, M. Christodorescu, and L. Sha, “Learning execution contexts from system call distributions for intrusion detection in embedded systems,” in Xiv. preprint arXiv:1501.05963, 2015.
[28] F. A. T. Abad, J. Van Der Woude, Y. Lu, S. Bak, M. Caccamo, L. Sha, R. Mancuso, and S. Mohan, “Onchip control flow integrity check for real time embedded systems,” in IEEE CPS'13, 2013, pp. 26–31.
[29] R. Abdel, M. Hasan, S. Mohan, D. Agarwal, and M. Caccamo, “ReSecure: A restart-based security protocol for tightly actuated hard real-time systems,” in IEEE CERTS, 2016, pp. 47–54.
[30] R. Pellizzoni, N. Paryab, M.-K. Yoon, S. Bak, S. Mohan, and R. B. Bobba, “A generalized model for preventing information leakage in hard real-time systems,” in IEEE RTAS, 2015, pp. 271–282.
[31] A. Francillon and C. Castelluccia, “Code injection attacks on harvardarchitecture devices,” in ACM CCS. New York, NY, USA: ACM, 2008, pp. 15–26.
[32] M. Hasan, S. Mohan, R. B. Bobba, and R. Pellizzoni, “Exploring opportunistic execution for integrating security into legacy hard real-time systems,” in IEEE RTSS, 2016, pp. 123–134.
[33] Y. Zhou and D. Feng, “Side-channel attacks: Ten years after its publication and the impacts on cryptographic module security testing,” IACR Cryptology ePrint Archive, vol. 2005, p. 388, 2005.
10