University of Salerno

Department of Computer Science

Dottorato di Ricerca in Informatica
Curriculum Internet of Things and Smart Technologies
XXXV Ciclo

Tesi di Dottorato / Ph.D. Thesis

Security and Privacy Concerns within Smart Environments: User-Centered Approaches for Defining Secure Trigger-Action Programs

Bernardo BREVE

Supervisor: **Prof. Vincenzo DEUFEMIA**

PhD Program Director: **Prof. Andrea DE LUCIA**

A.A 2021/2022

ABSTRACT

The advent of the Internet of Things (IoT) paradigm has launched a new world of opportunities, bringing with it a new understanding of objects and technology that, today, is all around us. In fact, especially in domestic environments, there is a real digital overhaul of the objects that users interact with on a daily basis. Thus, devices such as lights, TVs, cameras, locks, and electrical outlets are expanding their set of physical and technical characteristics expected from such devices with the addition of "smart" functionality, made possible by providing such devices with Internet connectivity. The activation and management of these devices, therefore, can be coordinated remotely, via smartphone, or through voice assistants such as Amazon Alexa or Google Home. Being devices eternally connected to the Internet, they never stop collecting, processing and sending data of the environment in which they are placed. For example, a temperature sensor can at any time send updates on the state of the environment, directly to the user's smartphone, or a camera can provide, at any time, a live image of the apartment. In addition, such devices can be configured to provide true automation that is triggered in response to the occurrence of certain conditions. This aspect is precisely among the most important of those offered by IoT technology; in fact, the interoperability that can be established between different IoT devices and/or communication with different web services is capable of radically simplifying the everyday life of users who make use of them.

Recently, we have witnessed the emergence of several platforms specifically designed to simplify the definition, by the end-users, of automation, based on the interaction between devices and services in a smart environment. Among these platforms, the most popular ones are those that provide a reinterpretation of the Trigger-Action paradigm, i.e., the ability to define automatisms by specifying the event (or trigger) and the action. The first component establishes the type of event for which the automation is triggered if conditions are met. Instead, the second

component refers to the operation to be performed to complete the automation. Such platforms take the name of Trigger-Action Platforms (TAPs).

The wide spread of IFTTT and other TAPs raised the question of the enormous security risks to smart environments and the privacy of end users interacting with the platforms, which might be caused by the defined behaviors. This is partly due to the high level of abstraction that TAPs provide and that very often gives little emphasis to security and privacy issues. In addition, the low level of technical knowledge that the average TAP user has, however, would not allow them to approach such issues and understand their severity.

This thesis aims to address a common problem in trigger-action platforms (TAPs) - the inability for users to create behaviors that actively protect their smart environment. To solve this issue, we propose integrating the capabilities of a newly created IoT device, called the *Intrusion Defender* (ID), into an existing TAP. The ID is capable of monitoring network traffic throughout the smart environment for unusual patterns that may indicate a cyber attack. Additionally, by providing an appropriate level of abstraction, the events detected by the ID are presented to the user in a more understandable format, allowing them to create rules that respond to, for example, a Denial-of-Service attack.

Additionally, we propose an NLP-based solution that can automatically identify any potential security and privacy risks associated with the trigger-action rules defined by the user. We achieved this by utilizing the capabilities of transfer learning models that incorporate the transformer architecture to achieve precise outcomes even with a limited amount of data for training. Specifically, we employed the Bidirectional Encoder Representations from Transformers (BERT) model developed by Google for identifying risks. The results from our experiments demonstrated the reliability and precision of the model we trained in classifying risks into three different damage classes, particularly when compared to other similar methods in the field.

To further reinforce the understanding of these risks, we use natural language models to generate example scenarios. In particular, our risk identification model is paired with a component that explains why the rule is activated through a possible scenario. We

fine-tune the model through prompts, which are virtual tokens embedded in a continuous space, to aid the model in understanding the rule's purpose. Our evaluations of the trained models prove that our approach is effective in producing plausible and contextually appropriate justifications.

ABSTRACT

L'avvento del paradigma dell'Internet of Things (IoT) ha dato vita ad un nuovo mondo di opportunità, portando con sé una nuova comprensione degli oggetti e della tecnologia che, oggi, ci circonda. In effetti, soprattutto negli ambienti domestici, c'è una vera e propria rivoluzione digitale degli oggetti con cui gli utenti interagiscono quotidianamente. Pertanto, dispositivi come luci, televisori, telecamere, serrature e prese elettriche stanno espandendo il loro insieme di caratteristiche fisiche e tecniche attese da tali dispositivi con l'aggiunta di funzionalità "intelligenti", rese possibili dalla fornitura di tali dispositivi con la connettività Internet. L'attivazione e la gestione di questi dispositivi, quindi, possono essere coordinate in modo remoto, tramite smartphone o tramite assistenti vocali come Amazon Alexa o Google Home. Essendo i dispositivi eternamente connessi a Internet, non smettono mai di raccogliere, elaborare e inviare dati dell'ambiente in cui sono collocati. Ad esempio, un sensore di temperatura può inviare in qualsiasi momento aggiornamenti sullo stato dell'ambiente, direttamente sul smartphone dell'utente, oppure una telecamera può fornire, in qualsiasi momento, un'immagine in diretta dell'appartamento. Inoltre, tali dispositivi possono essere configurati per fornire un'automazione vera e propria che viene attivata in risposta all'occorrenza di determinate condizioni. Questo aspetto è proprio tra quelli più importanti offerti dalla tecnologia IoT; infatti, l'interoperabilità che può essere stabilita tra diversi dispositivi IoT e/o la comunicazione con diversi servizi web è in grado di semplificare radicalmente la vita quotidiana degli utenti che ne fanno uso.

Recentemente, abbiamo assistito alla comparsa di diverse piattaforme specificamente progettate per semplificare la definizione, da parte degli utenti finali, dell'automazione, basata sull'interazione tra dispositivi e servizi in un ambiente intelligente. Tra queste piattaforme, quelle più popolari sono quelle che forniscono una rielaborazione del paradigma Trigger-Action, ovvero la capacità di definire automatismi specificando l'evento (o trigger) e

l'azione. Il primo componente stabilisce il tipo di evento per il quale l'automazione viene attivata se le condizioni sono soddisfatte. Invece, il secondo componente si riferisce all'operazione da eseguire per completare l'automazione. Tali piattaforme prendono il nome di piattaforme Trigger-Action (TAP).

La diffusione capillare di IFTTT e altre piattaforme TAP ha sollevato il problema dei rischi di sicurezza enormi per gli ambienti smart e la privacy degli utenti finali che interagiscono con le piattaforme, che potrebbero essere causati dai comportamenti definiti. Ciò è dovuto in parte al alto livello di astrazione che le TAP forniscono e che molto spesso dà poca importanza alle questioni di sicurezza e privacy. Inoltre, il basso livello di conoscenza tecnica che ha in media un utente TAP, tuttavia, non gli permetterebbe di affrontare tali questioni e di comprenderne la gravità.

In questa tesi, proponiamo una soluzione ad un comune difetto presente in tutte le piattaforme trigger-action (TAP), ovvero la mancanza di capacità per gli utenti di definire comportamenti che difendono attivamente il loro ambiente smart. Per risolvere questo problema, integriamo la funzionalità offerta da un dispositivo IoT creato ad hoc, chiamato Intrusion Defender (ID), all'interno di una TAP esistente. L'ID è in grado di monitorare il traffico di rete in tutto l'ambiente smart per rilevare schemi anomali che possono indicare un attacco di sicurezza informatica in corso, e attraverso l'introduzione di un adeguato livello di astrazione, gli eventi catturati dall'ID vengono presentati all'utente finale in modo più comprensibile, al fine di consentirgli di definire una regola le cui azioni si verificano in risposta ad un attacco di negazione del servizio, ad esempio.

Inoltre, proponiamo una soluzione basata su NLP in grado di identificare automaticamente eventuali rischi di sicurezza e privacy associati alle regole trigger-action definite dall'utente. Abbiamo fatto ciò sfruttando il potenziale dei modelli di transfer learning che sfruttano l'architettura transformer per ottenere risultati precisi anche con un insieme di dati limitato disponibile per la fase di addestramento. In particolare, abbiamo utilizzato il modello Bidirectional Encoder Representations from Transformers (BERT) presentato da Google per eseguire i compiti di identificazione dei rischi. Gli esperimenti condotti hanno di-

mostrato la affidabilità e la precisione del modello addestrato nell'identificazione dei rischi, classificandoli in 3 diverse classi di danni, soprattutto quando confrontati con soluzioni simili presenti in letteratura.

Infine, per rafforzare ulteriormente la comprensione di questi rischi, utilizziamo i modelli di linguaggio naturale per generare scenari di esempio. Abbiamo abbinato il modello per l'identificazione dei rischi con un componente di spiegabilità, che presenta uno scenario plausibile per l'attivazione della regola. Abbiamo utilizzato il fine-tuning basato su prompt, con prompt morbidi costituiti da token virtuali incorporati in uno spazio continuo, per guidare il modello nella deduzione del contesto di esecuzione della regola. Le valutazioni condotte sui modelli addestrati hanno dimostrato l'efficacia della nostra soluzione nella generazione di giustificazioni plausibili e adeguate al contesto.