

# Plug-and-Play: Framework for Remote Experimentation in Cyber Security

Klaudia Krawiecka  
University of Oxford  
Oxford, United Kingdom  
klaudia.krawiecka@cs.ox.ac.uk

Alina Petrova  
University of Oxford  
Oxford, United Kingdom  
alina.petrova@cs.ox.ac.uk

Jack Sturgess  
University of Oxford  
Oxford, United Kingdom  
jack.sturgess@cs.ox.ac.uk

Ivan Martinovic  
University of Oxford  
Oxford, United Kingdom  
ivan.martinovic@cs.ox.ac.uk

## ABSTRACT

Remote experiments with human participants play an important role across multiple fields of studies, from medical science to engineering, as they allow for better representation of human participants and more realistic experimental environments, and ensure research continuity in exceptional circumstances, such as nationwide lockdowns. Yet cyber security has few standards for conducting experiments with human participants, let alone in a remote setting. In this paper, we introduce an end-to-end framework for remote experimentation in cyber security. This framework systematises design and deployment practices while preserving realistic, reproducible data collection and the safety and privacy of participants. We evaluate our framework using a case study involving Internet-of-Things (IoT) devices deployed at remote locations and analyse the experience from the perspectives of both the researchers and the participants.

## CCS CONCEPTS

• **Security and privacy** → **Social aspects of security and privacy**; **Usability in security and privacy**;

## KEYWORDS

user studies, field studies, cyber security, Internet-of-Things

### ACM Reference Format:

Klaudia Krawiecka, Jack Sturgess, Alina Petrova, and Ivan Martinovic. 2021. Plug-and-Play: Framework for Remote Experimentation in Cyber Security. In *European Symposium on Usable Security 2021 (EuroUSEC '21)*, October 11–12, 2021, Karlsruhe, Germany. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3481357.3481518>

## 1 INTRODUCTION

The outbreak of the COVID-19 pandemic had a significant impact on many research activities, including in-lab experiments and user studies [17]. Many governments enforced national lockdowns,

which prevented direct interaction between researchers and experiment participants. In the field of cyber security, one of the most affected areas was biometric research due to the need to collect samples from human participants.

As the lockdowns highlighted weaknesses in many areas of human activity, they also showed that some well-established methodologies and experimental frameworks are insufficient for comprehensive data collection and analysis. When studying smart homes or smart office ecosystems, lab experimentation yields data that are less realistic than the data that would be obtained from sensors *in situ*, with their background noises and genuine agent behaviour. The processes used to recruit human participants can also restrict the realism of the study. In practice, many studies recruit a homogeneous cohort of university staff and students, many of whom will be inexperienced in the use of the equipment or the ecosystem under test. This also limits the users in the study to a particular place of residence, region, and likely to a certain age group and occupation; other user groups are excluded completely, such as those with disabilities or prohibitively busy schedules.

A possible solution to the problems above is *remote experimentation*. Such experiments might be conducted at scale and enrich the cyber security community with larger, more comprehensive datasets that are currently lacking [15].

*Contributions.* This work contributes the following.

- We propose an end-to-end framework for remote experimentation in cyber security. To the best of our knowledge, this is the first such framework in this domain.
- We provide a comprehensive requirements analysis for remote experimentation in cyber security, and encapsulate this into five framework requirements.
- We evaluate our framework using as a case study an experiment that we conducted under lockdown, and discuss lessons learned incorporating participants feedback.

## 2 BACKGROUND

Modern science is based on empirical evidence. Such evidence can be gathered through unbiased, rigorous observation or experimentation. In fields like computer science, experiments play a key role in validating hypotheses, studying new phenomena, and designing new systems and protocols. The fundamental factors of credible experimentation are randomisation and control [7]; randomisation

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

*EuroUSEC '21, October 11–12, 2021, Karlsruhe, Germany*

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8423-0/21/10.

<https://doi.org/10.1145/3481357.3481518>

ensures true representation and inclusivity in terms of participants and experimental environment, while control enables researchers to isolate variables and components being studied [5].

Remotely controlled experiments have traditionally been used in the fields of physics and engineering [6, 10, 14], in cases when the equipment cannot be accessed during the experiment [20] or as part of distance learning [12]. However, such remote experiments do not typically involve human participants: once the equipment is set up by the researchers, the experiments are conducted and recorded remotely by the same research team. Under lockdown restrictions, many researchers found themselves in a situation in which they needed to conduct an experiment involving human participants, yet they could not be present at the experiment location [16]. With human participants in the loop, the complexity of setting up and conducting remote experiments becomes a new challenge that calls for a new methodological approach.

Unlike other research fields, cyber security has very few formalised frameworks or standards for conducting experiments with human participants [8]. The challenges in developing such uniform standards include (a) ensuring privacy of the participants as they very often share their private information, and their physical safety, (b) accounting for heterogeneous equipment needed to simulate various ecosystems (e.g., smart homes) as well as the maintenance of such infrastructures [1], and (c) creating a realistic setting for an experiment since security controls should be tested in an environment that is similar to the real one.

Constructing a realistic environment during the experimentation phase allows for the examination of participants' natural behaviour, which is crucial for proper validation of the performance of proposed solutions and models [3]. Classical, supervised environments may provide controlled, repeatable data collection; however, as opposed to *field experiments*, they can neither reach user groups that have no access to the lab due to geographic, medical or other reasons, nor account for a full range of factors affecting the system under test [19]. Field experiments not only better capture the natural behaviour of the users and their potential interactions with the system, but allow for a broader set of experimental settings that are not constrained by a given lab [11]. This makes field experiments the default choice when the research focus is on the interaction between users and systems, products, or equipment [2].

We consider an environment to be *remote* if the researcher conducting the experiment has no physical access to it during the experiment. In some cases, a remote environment can be accessed by the researcher before or after the experiment (e.g., a lab or an office space); in other cases it cannot be accessed by the researcher at any time (e.g., a participant's private house). A *remote experiment* is an experiment that (a) is conducted in a remote environment, (b) involves human participants, and (c) uses equipment that records certain actions of the participants (movements, interactions with the equipment, etc.).

### 3 FRAMEWORK

This section describes our proposed framework for remote experimentation and the goals and requirements that drove its development. The framework is divided into four phases, and each phase is broken down into activities that define the experimentation process.

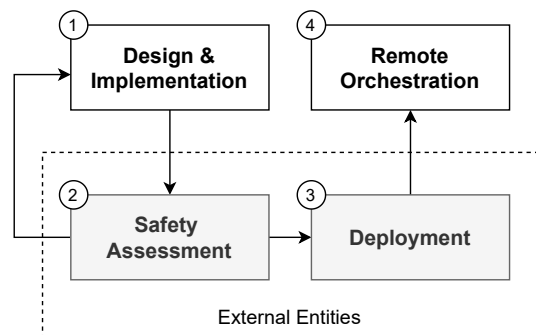
#### 3.1 Goals

Our proposed framework aims to systematise remote experimentation to achieve the following three goals.

- G1 **Systematic Design and Deployment.** Remote experiments are conducted outside of a laboratory, in locations such as people's homes, their workplaces, or anywhere they spend their free time. These are uncontrolled environments, so tests must be structured with respect to which variables can be controlled and which cannot.
- G2 **Realistic, Reproducible Data.** The possibility of conducting experiments in remote environments is important as it allows to study the natural behaviour and reactions of users. Thus, data collected during such experiments reflect reality better than data collected in an artificial setting such as a laboratory. Furthermore, the degree of control introduced by the remote experimentation framework enables reproducible, repeatable data collection [19].
- G3 **No In-person Interaction.** Not undertaking the experiment in a laboratory allows for the recruitment of participants previously excluded from user studies due to their inability to travel, whether due to disability, proximity, or scheduling. While this has improved over the last decade for people with various disabilities in the field of human-computer interaction [18], it is still a fairly neglected topic in other fields of computer science. Furthermore, such an approach can help in more specific scenarios where external conditions, such as a pandemic, limit the possibility of direct interaction with users.

#### 3.2 Overview

Figure 1 shows an overview of our framework. In Phase 1, the researchers refine the experimental design and then implement the tools necessary to conduct the experiment remotely. In the next phase, the risk and safety assessment is carried out by the institutional review board. Feedback from the review process should be incorporated into the design and implementation phases; thus, Phases 1 and 2 cycle until both are satisfied. Once all the requirements are met, the equipment can be transported to a remote site. In Phase 4, the researchers remotely supervise and collect data.



**Figure 1: The four phases of our proposed framework. Phases 2 and 3 are carried out by external entities such as the institutional review board and experiment participants.**

### 3.3 Requirements

Our framework has the following requirements. Each relates to the entire experimentation process, not to a specific phase, and can be adapted to different use cases.

- R1 **Safety.** Ensuring the safety of participants is well established in research facilities, but this is not the case for remote environments. While we cannot influence the general safety conditions of people's private homes, we can identify potential risks and mitigate wherever possible. The instructions that we supply to participants should also cover safety procedures regarding the use of electronic components and activities pertaining to the experiment.
- R2 **Ease of Use.** In contrast to standard experimentation, participants in a remote experiment may have to engage actively in setting up and configuring the testbed. It is important to consider ease of use, both in terms of initial configuration and performing the tasks of the experiment. The equipment given to participants must be easily configurable and durable. The instructions for both setting up and using the equipment must be clear and unambiguous to ensure that the experiment runs smoothly and yields concrete results.
- R3 **Ease of Maintenance.** Without a researcher being present, preparations need to be made for troubleshooting and maintenance. Devices and tools should be set with sensible default configurations such that they start in (and reboot to) a usable state. All equipment should be checked before being shipped, online resources should be available, and participants should have ways to contact the researcher for additional support. Such mechanisms must be comfortable for participants to use.
- R4 **Remote Supervision and Data Collection.** The researchers must be able to monitor the progress of the experiment. In some cases, this will be in real time to provide immediate feedback to the participants; in others, the researchers may just wish to ensure that certain milestones were met during the experiment.
- R5 **Transmission and Data Security.** There is always a risk that equipment in an uncontrolled environment will get damaged, so all collected data must be transmitted to and stored on a separate device. Backing up the data to an online server is also important because data loss endangers the integrity of the experiment. To ensure the privacy and security of the data, all communication between nodes, including backup servers, electronic devices, and other platforms, should be secured by utilising standard data transmission protocols and tools. Moreover, access to the data on the backup server and other storage units should be protected by appropriate protection mechanisms such as strong passwords and two-factor authentication.

### 3.4 Phase 1: Design & Implementation

This phase consists of two main activities: the design activity, which then guides the implementation activity.

*Design.* This phase begins with defining the goals and scope of the experiment. It includes describing the desired results as well as identifying the parameters and limitations of the experiment. These tasks are encapsulated in the design activity to ensure understanding of all key factors that have to be addressed during the

implementation activity. More specifically, this incorporates determining what is to be measured, what is to be controlled, how these things will be done, and what is an expected output. Answering these questions will help to obtain information not only about the data and its format but also about participants and environmental factors that need to be involved in the experiment. Thus, the design activity is concluded by preparing extensive documentation that includes the description of experimental and control groups, test sites, test cases, and physical configuration outlines.

The structure of remote experiments allows for the inclusion of people who would otherwise be marginalised. This includes people who cannot travel to research facilities due to distance or mobility issues. Depending on the assumptions made about the experimental and control groups, standard recruiting approaches such as via newsletters or word-of-mouth may not be sufficient as they limit the reach. Alternatives such as recruiting through social media or through gatekeepers (*i.e.*, people or institutions with access to the target population) should be considered [4].

Choosing the right location for a remote experiment is another key factor. For example, in many types of research, such as biometrics or activity recognition, it is important to represent the participant's behaviour realistically. The research laboratory is an artificial environment that influences the behaviour of the participant, often causing the person not to behave naturally. When conducting a remote experiment, a natural environment in which the user feels comfortable can be selected. Examples of such environments could be users' homes, their workplaces, or places where they spend their free time. In these environments, however, there is a privacy aspect to consider. Participants should be able to decide when data is collected. Controllers, which are the components that allow participants to control the entire process, should be introduced. On the other hand, researchers should be able to monitor the status of the experiment and the equipment. These two aspects need to be carefully balanced to avoid interference in the experiment and invasion of participant privacy.

The next step is to prepare test scenarios that will allow collecting the necessary data from a target group. These scenarios should be straightforward to avoid introducing too much complexity, which could result in the participants' inability to complete the experiment. Due to the fact that such experiments are conducted in uncontrolled environments, all uncontrolled parameters should be considered because they may affect the outcome of the experiment. In addition to considering strategies to avoid noise in the data, strategies for handling incomplete data should also be incorporated into this phase. After designing the test cases, researchers should prepare a user manual that guides participants through the phases of deployment and experimentation. For this task, consider what will be best for the participants to accommodate their needs, such as video and audio tutorials and jargon-free language.

Once all parameters are defined and potential constraints identified, an outline of the physical configuration should be prepared to help guide the construction of software and hardware components required to deploy the experiment at a test site. Such an outline can be created in the form of a list of necessary physical components and a description of their communication channels. It could also be a diagram, or a combination of both. Designing such an outline helps to understand which elements should be given special

attention. Researchers must ensure that the data collected from the participants is transmitted and stored in a secure manner. Many secure communication protocols and tools exist to achieve this goal. In addition, it should be assumed that if the testbed requires internet access, some environments will not have network access points (and some participants will not want to connect to their private network), so alternatives should be considered, such as providing an access point.

*Implementation.* The second activity takes place after all the design aspects are defined. During the implementation activity, the tools to collect data are either created or selected from the pool of existing instruments. First, researchers have to consider which platforms the participants are familiar with. For instance, when building the controller, researchers might consider using a mobile or desktop application that will allow the participants to easily manage data collection. To protect their data, researchers should use secure communication protocols such as virtual private networks and enable data encryption on the nodes that are responsible for storing the data. Moreover, if the testbed consists of distributed components, aspects like temporal synchronisation and traffic balancing have to be addressed. Additionally, the tools should allow both participants and researchers to pinpoint who is currently participating in the experiment to ensure correct data labelling.

Another important aspect is the delivery of the equipment to the test site. Upon delivery, some components may break, so spare components should be included. It is also important to focus on the usability of such tools as it will help to create a suitable environment to conduct remote experiments safely. For example, the equipment should be as self-contained as possible with all software pre-installed, so as to minimise the possibility of errors in setting up.

The last important aspect in this phase is creating a backup. Tools can be programmed to transfer data to a remote server or store them locally. In each of these cases, protecting the data from corruption or transmission issues is crucial. Overall, good design and implementation can help reduce the likelihood of noise being introduced into the data due to external, uncontrolled factors.

### 3.5 Phase 2: Safety Assessment

In this phase, risk and safety assessments are conducted by the responsible institutional review board that reviews the methods proposed for carrying out the experiments. This step should never be neglected as practices deemed to be unethical can result in serious consequences.<sup>1</sup> Each research institution has different procedures to grant formal approval. The purpose is to identify the risks in every aspect of the experiment and to ensure that they are addressed.

*Participants.* It is important that all participants are able to understand fully and to give informed consent to their participation in the experiment. If the study involves any vulnerable people (e.g., based on age, health, or mental impairment), this should be declared and the risks enumerated. In remote experiments, we must also be cognisant that others, including children and pets, may have access to the remote environment, and assess how they might be impacted by

and how they might interact with the equipment (e.g., magnets may need to be shielded so as not to adversely affect medical implants). Depending on the type of experiment, this may necessitate strict eligibility criteria to reject unsuitable environments altogether.

Participants must feel safe and comfortable. Equipment should not pose an unnecessary burden to their everyday life and there should be a mechanism by which they can revoke consent if they so choose. Consideration should be given to how the equipment is shipped to and from participants and how it may be collected early if the participant withdraws from the study.

*Testbed.* The equipment provided to the participants must be safe to set up and use, and must be tested and cleaned in between each deployment to ensure its continued operational safety. Electrical devices should be tested professionally to protect against catastrophic failure (typically, this service can be arranged via an IT department). Does any component pose a choking risk to small children? Are there any sharp edges, open circuits, or loose wires that might cause injury or be pulled by a pet? If such things are necessary for the experiment, and cannot be encased safely, a strategy must be in place to mitigate the risks they pose (such as eligibility criteria, placing objects at height, safety labelling, etc.).

*Data.* The participants should know what data are being collected and how. To protect confidentiality, data should be stored and transmitted encrypted. If stored unencrypted on capture devices, the data should be wiped before the devices are shipped back to the researchers (unless the researchers collect them directly) for protection in transit. To protect integrity, data should be backed up automatically, preferably on an external device or server. The data should be accessible only by authorised users, revocable if necessary, and anonymised before publication.

Care must be taken not to capture more data than specified, such as from other people using the remote environment. This may require that devices be easily turned on and off and that they clearly display their operational state at all times.

### 3.6 Phase 3: Deployment

The main difference from standard experimentation approaches lies in Phase 3. Typically, such experiments are deployed and conducted in a dedicated location within a research facility. This gives researchers complete control over the physical configuration of their hardware or software, and allows them to detect and quickly fix any errors or technical issues. Additionally, they can monitor the course of the experiment, interrupt it or repeat it when necessary. However, in the case of remote experiments, all these tasks are carried out by the participants of the experiment. Even if the researchers can guide them to some extent through a video or audio conversation, they will be unable to physically solve the technical issues.

*Supporting Instructions.* For this reason, the preparation of appropriate instructional materials in the design and implementation phase significantly influences the course of the deployment phase. If the participants cannot follow the instructions and operate the equipment, the experiment cannot continue. The basic assumption should be that the participants do not have any experience in operating the equipment, so the testbed setup should be simplified. Moreover, the deployment should be done within a time

<sup>1</sup>A paper was withdrawn from IEEE S&P 2021 for such a violation, [https://www.ieee-security.org/TC/SP2021/downloads/2021\\_PC\\_Statement.pdf](https://www.ieee-security.org/TC/SP2021/downloads/2021_PC_Statement.pdf)

frame specified in the information sheet given to participants at the beginning of the experiment. Depending on the design of the testbed, its configuration can consume a significant amount of time, and researchers must consider that the deployment takes place in external environments, which may cause problems with the regular use of such spaces. This may cause some participants to withdraw from the experiment, which may impact the outcome and the cost of running such an experiment.

*Physical Setup.* Another important aspect is the unified physical configuration of the testbed. Depending on the type of experiment, a consistent physical arrangement of the equipment across users may affect the outcome. We may want each participant to use the same device configuration and similar locations and orientations to compare readings. The instructions should be sufficiently strict to guide participants during the deployment phase.

After completing the initial testbed deployment, the participants should verify whether the infrastructure is fully operational. One approach to do so could be to equip the controller with an interface that informs participants directly about the status of the software and hardware components of the testbed. Another approach would be for researchers to provide feedback to participants after verifying the state of the test environment using remote supervision tools.

*Testing.* Before starting the experiment, participants should conduct several test trials to become familiar with the equipment and tasks. This will also allow researchers to verify that their supervision tools are working and the system can capture and backup the data. At the end of this phase, each participant involved in testbed deployment should be asked about their experience. This will not only allow for improvements of certain components, but also has the potential to reveal problems that would otherwise be difficult to identify based on the data collected alone. Moreover, the researchers could ask the participants to provide supplementary materials, such as a brief description of the location of individual testbed components or photos. The manner of sharing such materials should continue to preserve privacy.

### 3.7 Phase 4: Remote Orchestration

The final phase covers two main activities, including remote supervision of the experiment and data collection.

*Remote Supervision.* The first activity concerns the ability to monitor the state of the testbed at the test site remotely without violating the privacy of the participants. This includes verifying that all testbed components are properly configured before starting the experiment. This verification should be performed at component level in order to pinpoint the exact source of a possible failure. In addition, researchers should be able to immediately detect problems not only with the infrastructure, but also with the course of the experiment. When the experiment is performed incorrectly, the researchers should be able to interrupt it and give the participants feedback. A dedicated channel should be established between the researchers and participants so that they can communicate their feedback (and any problems) remotely; this could be integrated into the equipment, such as with a controller interface, or it could use third party software such as voice or video calls or a secure messenger app. They should also be able to resume the experiment if any issue occurs. Enabling the participants to stop and restart

the experiment provides time flexibility and a means to handle unexpected interruptions. These are not typical problems when experiments are conducted in a research facility; remote experiments, on the other hand, can be deployed in spaces that can be used by a large number of people, so such flexibility is essential.

*Data Collection.* The second activity in this phase concerns real-time data collection. Data from an experiment can be stored in two ways, locally on the testbed components provided to participants or remotely on a server. We recommend that one of these methods is used as the primary storage point for data and the other as a backup. This is important as data transmission can be unreliable in environments without good network coverage. It should also be assumed that there are environments that do not have internet access points, so researchers should provide the appropriate equipment if the testbed transmits the data to a remote server.

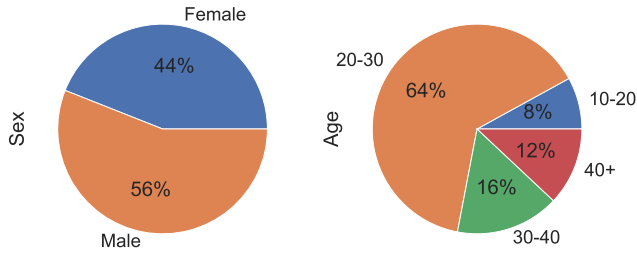
Depending on the nature of the research project, the experiment data can be retained for a certain period of time. As stated in Phase 2, the participants of the experiment should be informed in what form and for how long their data will be stored. They should be provided with information about the anonymisation of their data, any third parties that may use their data, and the possibility of withdrawing from the experiment, which will result in the deletion of their data. The participants should find out when they can withdraw their participation and for how long their data will be kept. Consequently, data management must be part of the data collection activities to ensure that they are stored and processed in line with the information provided to participants. This means not only that the data must be secured during transmission, but also when stored. Overall, this phase helps to avoid incomplete or incorrect sampling as well as procedural errors due to incorrect transmission and storage of data.

## 4 CASE STUDY

In 2020, social interactions and public gatherings were significantly limited. Due to the restrictions, it was not possible to conduct user studies in the laboratory. For this reason, we decided to run our experiment remotely.

We collected samples of behavioural interactions with physical objects. The aim of the experiment was to explore the potential of physical object interaction biometrics for authentication and identification in smart environments. To operate within the lockdown restrictions, we needed the experiment to be performed outside of the lab and without the researcher being there in person (goal G3 of the framework). To achieve this, we had to design the experiment to ensure that every deployment would be consistent (G1) and that the data would be collected in a reproducible manner (G2). Furthermore, this presented us with an opportunity to collect real-world data from apposite users in genuine smart environments (G2).

We ran this experiment from October 2020 to January 2021. During this period, we recruited 25 participants spread across six different households in the United Kingdom and Poland. The recruitment took place through advertisements on social media and a departmental newsletter. The youngest participant of the experiment was 11 years old whereas the oldest was 57. To set up this experiment in a realistic environment, we focused on recruiting



**Figure 2: Review of sex and age distribution of the participants in our remote experiment.**

Household	H1	H2	H3	H4	H5	H6
No. of people	3	2	6	4	3	7
No. of devices	8	10	4	5	4	7

**Table 1: The number of household members and devices in each household that participated in the remote experiment.**

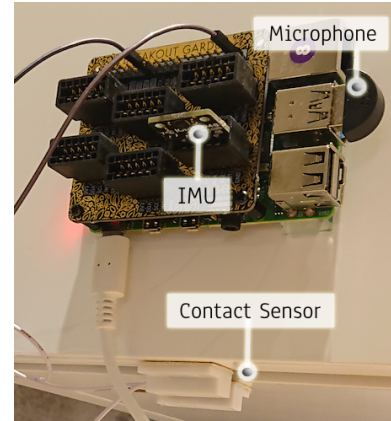
participants of different ages. Figure 2 shows the sex and age distribution among them. Overall, the majority of participants were between the ages of 20 and 30, with 56% male and 44% female.

We delivered ten sensor boards to each household that agreed to participate in the user study. Each sensor board was equipped with an Inertial Measurement Unit (IMU), a microphone, and a magnetic contact sensor, as shown in Figure 3. Participants could place any number of such devices on various household objects inside one of the rooms in their home. Figure 5 shows examples of sensor board deployments in two households. Although we delivered the same amount of equipment to each participant, some spaces were smaller, so we decided to give them flexibility to arrange them as they saw fit. Table 1 shows the number of participants and devices in each of the households. Participants also received a smartphone with a pre-installed mobile application that allowed them to turn data collection on and off by the sensor boards. More about the architecture of this application and other components can be found in Section 4.1.

We prepared a user manual with the information how to proceed with the experiment. We asked each participant to repeat the same sequence of interactions with objects 20 times. However, these 20 trials could be spread over several hours or days depending on users' preferences.

#### 4.1 Architecture

To satisfy all the requirements for the safe conduct of a remote experiment, we created an experimentation kit that consists of three main components, including sensor boards, the controller, and the server-side MQTT broker. Each component communicates with each other using MQTT, a lightweight messaging protocol for IoT devices [9]. Figure 6 shows the messages exchanged between these components during a round of the experiment. The MQTT broker is the key part as it forwards and stores messages sent to specific topics that all nodes subscribe and publish to. To meet requirement



**Figure 3: Raspberry Pis (RPIs) mimicked smart devices. The participants could freely arrange these in the room. Each RPI board contains an IMU sensor that measures acceleration, and has a built-in gyroscope and magnetometer. Each board is equipped with a microphone and a contact sensor.**



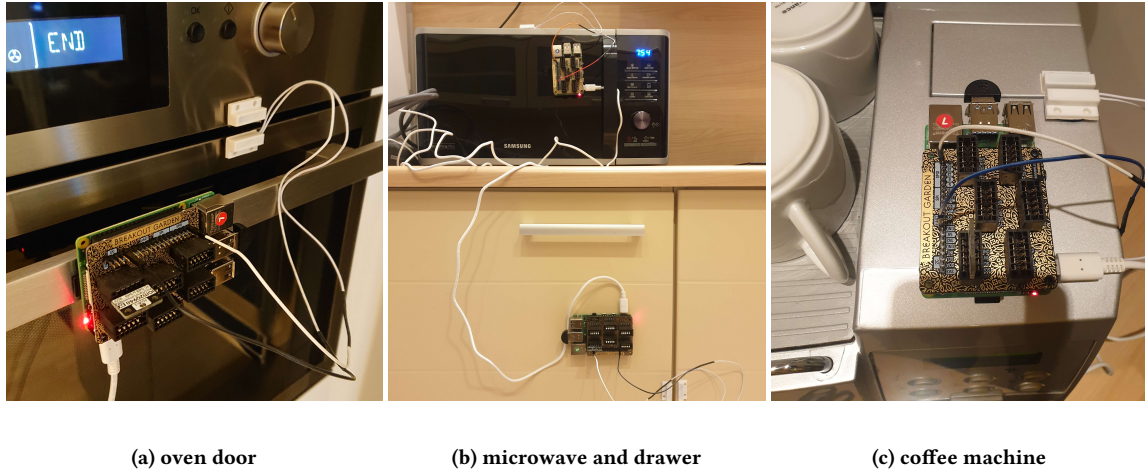
**Figure 4: The left picture shows the main screen of the controller, where the participant enters his name and the number of deployed sensor nodes. The participant is then redirected to the second screen, which displays the current trial number and the start button. Before this button appears, the application checks whether the sensor nodes are ready for data collection.**

R5, we configured the Eclipse Mosquitto MQTT broker [13] on the server in our research lab to use TLS to secure connections between the broker and the clients. Apart from the MQTT broker, all other components implement the MQTT clients.

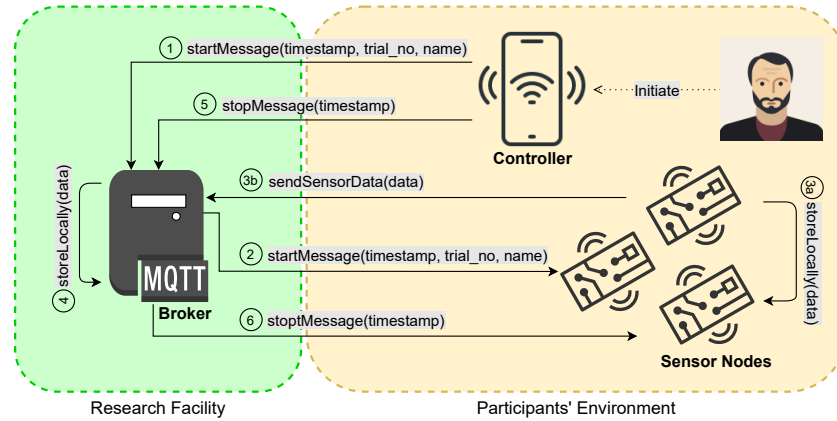
Each Raspberry Pi (RPI) board, in conjunction with a Pimoroni<sup>2</sup> module, formed a sensor node. These devices were chosen specifically for their sufficiently high sampling rates and an easy-to-use interface. We decided to include the Pimoroni module to facilitate

<sup>2</sup><https://shop.pimoroni.com/products/breakout-garden-hat-i2c-spi>





**Figure 5: Participants placed sensor boards on various household objects, including an oven door, a microwave, a drawer, and a coffee maker. These are examples of deployments from two households that participated in our remote experiment.**



**Figure 6: An overview of communication between different nodes during the remote experiment. The MQTT broker that relays messages between the mobile application and sensor devices is configured on the server that is placed in a research lab. The backup of the data is stored on the same server. The mobile device and sensors reside in participants' environments.**

the deployment of the experiment in the participants' environments. Essentially, this module allows various sensors to be easily placed on the RPI board, eliminating the need to solder these components, which satisfies R2. In addition, participants can quickly replace damaged parts with spare plug-and-play sensors included in the experimentation kit. This allows for easy maintenance and therefore complies with R3.

Each RPI board with the Raspbian operating system runs a Python script implemented as a system service that runs on boot and checks network connectivity. The network connection is provided by the controller that acts as a wireless access point. The only steps that participants need to perform are connecting the controller to the local network and enabling a virtual private network (VPN), which is described in the user manual. Next, the Python script examines whether the sensor components are properly installed and mounted in the Pimoroni module. If the script encounters any problems, it tries to relaunch until the system and its components

are fully operational. After connecting to the local network, devices send messages to the server to report their status and identifiers.

Next, the sensor nodes listen for the notifications from the controller. The controller, which in our case is a mobile device with the pre-installed Android application, allows participants to choose the start and end time of each round of the experiment. The sensor nodes do not collect the data until participants have enabled data collection using the controller. The mobile application interface is shown in Figure 4. After the participant enters his name and the number of sensor nodes he deployed, the application sends a message containing a timestamp, trial number and name to the MQTT broker. This message includes a time stamp for the purpose of synchronisation between different components. It is forwarded to all available sensor nodes that respond to the controller to confirm that they are ready to collect the data. The controller will only display a start button if all the deployed sensor nodes are fully operational. After the participants press the start button, these nodes

begin logging and sending data to the server. During each round, the sensor nodes write data locally to the RPI boards as a backup. Participants can easily turn off data collection by pressing the end button on the controller.

We delivered sensor nodes, the controller, spare parts, and the user manual to the participants' homes. During the deployment phase, we also provided them with deployment tutorials and videos. The user manual contained information on the configuration of the controller and other electronic components. To satisfy R1, we have also included safety guidelines for handling electronic components. In addition, the participants were advised to contact us if they encountered any problems. During the experiment, we were able to remotely monitor the status of individual components to check that the data collection was proceeding correctly, which fulfills R4.

Before the experiment, the participants received information sheets containing the description of the experiment and which data would be collected, and a consent form to read and sign.

## 4.2 Survey Results

After conducting our remote experiment, we asked each participant to complete a short questionnaire about their experiences. This questionnaire consisted of ordinal and open-ended questions. Our goal was to investigate what challenges participants encountered during the experiment, how helpful the instruction sheets were, and how they rated the overall usability of the experimentation kit. Below are the questions and the results of the survey.

**Q1. Were you actively involved in setting up the experiment?** The majority of the participants (57%) said that they actively participated in setting up the experiment in their house. Every household had multiple residents, so it was not a requirement that everyone be involved in the set up.

**Q2. Have you read the user manual and information sheet before the experiment?** We asked this question to investigate whether the participants had used the support materials provided. We found that 29% of them did not read these documents and relied on someone else in the household to explain the experiment. Of those who were primarily involved in setting up the equipment, 88% reported reading the documents. In all households, the sensor nodes were deployed and experiment performed correctly.

**Q3. What do you think the experiment was about?** Every participant answered this question correctly, indicating that they have either read the information sheet that came with the experimentation kit or realised it during the experiment. One of the participants explicitly mentioned that despite the lack of understanding of the subject, he was able to deduce from the course of the experiment what its purpose was.

**Q4. How useful was the user manual provided to you?** We asked the participants to rate on a scale of 1 to 5 how useful the user manual was, with number 5 indicating the highest usability. None of the participants rated the user manual below 3. The majority, 42%, chose number 4 on the satisfaction scale. The rest of the participants voted equally between 5 and 3.

**Q5. Were the instructions in the user manual clear?** Next, we wanted to find out if the instructions in the manual were written in a way that was understandable to participants from different professional backgrounds. Again, we used a rating scale from 1 to 5. This time, the participants' votes split evenly between 5 and 3. These options were voted by 36% of the participants, while 28% of them voted for the number 4.

**Q6. How useful was the walk-through video provided to you in the user manual?** We included a video in the user manual that demonstrates an example of how the experiment should be performed. We asked the participants to rate the usefulness of this video in the preparation for the experiment: 43% of them rated its usefulness as 3, 29% as 4, and 21% as 5; 7% of participants voted with 1 which means they did not find it helpful.

**Q7. How did you find the process of setting up the experiment at your house?** The participants' opinions regarding the difficulty of the experiment deployment showed wide discrepancies. In particular, 35% chose a rating of 5, which indicated that setting up this experiment was very easy for them, whereas 42% of the participants voted 4 and 3, and the rest rated it below 2, indicating it was very difficult.

**Q8. What challenges did you face while setting up the experiment?** 21% of the participants of the experiment answered that they did not encounter any difficulties. The others mentioned a number of issues they faced during the deployment process, including troubles with attaching the devices to various types of surfaces. The participants reported that the tape they used came off after some time, which meant that they had to repeat the experiment. Another challenge was the positioning of the contact sensor magnets. If the magnets were too far apart, the sensor could not operate properly. The first problem was solved by using a stronger double-sided tape that did not damage the surface. The second challenge was solved by adding to the user manual the preferred distance between the magnets of the contact sensor.

**Q9. Was the mobile application easy to use?** On a scale of 1 to 5, 72% of participants gave a rating of 5 for the usability of the mobile application that acted as the controller during the remote experiment. 7% of them rated their experience with the app a 4 and 21% a 3. Overall, nobody reported any issues with the mobile application while carrying out the experiment.

**Q10. How satisfied were you with the compensation?** The majority of the participants (72%) were fully satisfied with their compensation (in vouchers), while 7% of them were not satisfied, explaining that the actual deployment time of the experiment was longer than that mentioned in the information sheet, especially while they had to repeat the experiment when the devices dropped.

**Q11. Do you have any suggestions on how to improve the user experience?** Finally, we asked participants for feedback on their entire experience with our remote experiment. One person mentioned that better mounting materials should be used for the



sensor nodes as they are heavy. Another person suggested introducing a gamification element to make the whole process less tedious. The other participants either mentioned that the experiment was a very interesting experience or had no additional comments.

## 5 DISCUSSION

### 5.1 Research Ethics

Before the commencement of any user study, the experimental methodology has to be reviewed and approved by an appropriate ethics committee. A typical application pipeline is straightforward because it is based on established practices and policies. The researchers design their experiments to adhere to the institutional guidelines as well as common safety rules. These documents have been created based on many years of observation and research experience to facilitate the overall application process. However, often such guidelines do not address the cases in which the experiments have to be conducted outside the research facility, including participants' households. To better understand the challenges of conducting experiments remotely, we interviewed a member of our Research Ethics Committee from the Department of Computer Science at the University of Oxford.

Since 2018, this committee has received 121 applications and approved 98 of them. The majority of such applications, namely 68, have been approved after the first round of reviews. The average time to receive comments and detailed feedback from the committee was 12.5 days. While there were individual cases where the overall approval process takes longer, most projects were fully approved in 37.4 days on average.

In 2019, 85 applications were processed. However, in 2020, the committee only received 30 applications due to COVID-19. Because of the imposed restrictions, conducting indoor experiments in research facilities or offices was prohibited. While some user studies easily transitioned to virtual space (e.g., interviews), other types of experiments were postponed. Thus, to tackle this issue, we proposed a non-standard remote experimentation methodology. In our department, we were the first ones to propose such an approach and conduct the experiments remotely in the participants' homes. The common practice is to host such experiments on the premises of the university due to several reasons. This allows the researchers to keep the equipment safe and provide necessary maintenance, and ensures that the equipment is properly configured so that there are no deviations in terms of settings that may impact the final results. Also, hosting user studies locally allows the researchers to supervise and guide the experimentation process and to ensure the safety and privacy of the participants. By conducting experiments in people's private households, the researchers risk accidentally invading their privacy by, for instance, capturing information that they did not wish to share such as their daily habits, schedules, or other types of sensitive information. The committee pointed out that the experiments should be conducted in such a way that will allow the participants to turn off the recording of the data when needed to avoid accidental information leakage. Their other concerns included the physical safety of the participants, so the researchers have to ensure the delivery of instructions on how to safely set up the experiment and take into account other household

members, including children and people with disabilities. Moreover, the researchers have to consider their own safety. While some of these concerns are general and applicable to all types of user experimentation, others are specific to the proposed remote experimentation methodology.

We asked about the procedure of applying to conduct experiments off of university grounds. Normally, the committee looks at the time and place of the experiments. Preferably, such studies should be performed within office hours to support the researchers, if necessary. Depending on the user study, it can take place in public spaces, outside office hours, but the researchers have to provide detailed information on the time and location for safety reasons. However, such an approach is not viable in the case of remote experimentation since the participants are the ones who control the time and place of the experiments. Thus, the committee asked us to submit the detailed instruction sheet and the user manual for review as well as to propose a strategy to safeguard. Furthermore, the risk assessment and relevant mitigation strategies have been supplied along with the application. The researchers had to account for the risks related to COVID-19. To suggest appropriate measures, we looked into guidelines from the British National Health Service.

The committee has a comprehensive and formalised experimentation review structure. First, applicants must complete an online application form, answering specific questions about their experimental methodology. Next, they have to submit relevant documents, including consent forms and information sheets for the participants. While the responsibility to teach the researchers how to prepare such documents relies on their supervisors, often the committee provides such guidance throughout the application process. Since our proposed methodology did not belong to any known experimentation category, it had to be reviewed and approved by additional decision-making bodies, including the departmental Safety Committee and the Research Data Management Team, which significantly prolonged the overall approval process. The main reason is that our application could not be easily processed using standard procedures. Typically, if an accident occurs, the educational institution will take responsibility and face specific consequences. If there is a suspicion of negligence caused by the researchers, their actions are assessed under the academic misconduct document. However, during remote experiments the researchers cannot oversee and supervise the entire process; thus, it is challenging to act upon misconduct. Moreover, in addition to considering the legal obligations, the decision-making bodies had to review the application for compliance with the rules on research grants. Thus, the overall approval process took longer but had drawn attention to important aspects of remote experimentation and helped us to develop a reliable remote experimentation framework.

### 5.2 Lessons Learned

The aim of the experiment was to take realistic samples from participants without the need for direct interaction and to minimise the influence of external factors on the results of the experiment, in line with goals G1 to G3. We designed our user study to meet all framework requirements. Based on feedback from the participants and our observations, we draw the following conclusions with respect to the 5 framework requirements.

R1. Our safety planning proved effective and we experienced no safety-related incidents. While there have been no such incidents, we will consider not using too many cables next time to further reduce the risk of potential safety hazards.

R2. All participants understood the experiment (as evidenced by the answers to Q3) and were able to set up the equipment successfully, indicating that the supplied documentation was sufficient. The answers to Q6 and Q7 showed that the video was well received and that most participants were able to set up the equipment without any trouble, while a small number of less technical participants struggled with some parts, indicating that there is still room for improvement.

R3. We experienced some minor problems with adhesive tape and setting up magnets (Q8) in the early iterations of the experiment. These issues were minor inasmuch as they did not necessitate any modifications to the design or ethical considerations of the experiment and were easily resolved in later runs. We also had some damage caused by falling devices (due to the tape) and broken contact sensors (due to shipping, as they are fragile) that impacted some of the samples collected. To address this, a feature that allows to report the breakages (perhaps integrated into the app) would be useful, especially if done in a way that would absolve the participant of blame and incentivise reporting.

R4. Our app (Q9) and data-handling infrastructure were well designed and facilitated the smooth running of the experiment—the initial effort expended on these systems was worth it. Our remote monitoring provided us with the valuable assurance that the devices had been set up properly.

R5. We handled all data securely and had a full backup available if we had needed it. To the best of our knowledge, we did not experience any data leakage or loss.

During our experiment, we controlled key parameters, such as the correct setting and operation of testbed components, and the time synchronisation of these components. This was possible due to the use of a controller that gave immediate feedback to the participants as well as remote supervision tools. We were able to collect enough samples to build a comprehensive dataset to conduct the analysis of behavioural biometric traits. However, there were aspects that we could not control. Namely, if the device was damaged or the sensors stopped working during the experiment, some participants did not repeat the trials from before the issue was addressed, which resulted in receiving fewer samples from them. This happened only in two cases and did not affect the result of the experiment, but it should be addressed in the user manual. Factors that we could not control also included accidental outside interference, such as household members coming into the shared space during that experiment or outdoor construction work. Participants were advised to conduct the experiment when no-one else was using the shared space—so, when interruptions happened, some of them just repeated the experiment. The construction work was harder to deal with, but we note that this adds realism and can also happen during experiments at a research facility. All in all, our remote experiment was successful and the deployment of this framework allowed us to predict and address potential issues early on at the design stage of the experiment.

## 6 CONCLUSION

In this paper, we introduced a remote experimentation framework that enables the collection of realistic data in the natural environment for the target group and tackles the basic challenges that arise when experimenting in such uncontrolled environments. We evaluated the framework using a case study, in which we applied our approach to a user study on object interactions for biometric authentication and identification in smart environments. Our experiment took place under lockdown conditions and we showed that by deploying this framework, we were able to reach a wider group of participants and effectively create a usable data set without the need for direct interaction with the participants. In general, remote experimentation allows for flexible real-world data collection that involves a diverse set of participants with different needs and backgrounds, and a varied set of locations, each with its unique parameters, including size and layout. Since experiments are no longer bound to one specific location maintained by the researchers, they could be scaled both quantitatively and qualitatively.

The proposed framework could be expanded in several directions. For example, one could further improve the authenticity of the collected data by not only transferring the experiment to a real-world location but also by weaving the data collection into everyday activities by monitoring participants over time or with gamification techniques. Moreover, the framework could be further instantiated to serve particular research subfields such as behavioral biometrics, human-computer interaction, human activity recognition, and beyond. The instantiation could involve such experimental variables as the types of remote locations, equipment used, data collected, or analysis performed and would likely improve the reproducibility of the experiments even further.

## ACKNOWLEDGMENTS

The authors would like to thank Mastercard for financially supporting this work, Katherine Fletcher of our institution's research ethics committee for answering our questions and providing insights about the ethics review process, and the anonymous reviewers for their valuable feedback.

## REFERENCES

- [1] Aditya Ashok, Siddharth Sridhar, Tamara Becejac, Theora Rice, Matt Engels, Scott Harpool, Mark Rice, and Thomas Edgar. 2019. A Multi-level Fidelity Microgrid Testbed Model for Cybersecurity Experimentation. In *12th USENIX Workshop on Cyber Security Experimentation and Test (CSET 19)*.
- [2] Hassan A Aziz. 2017. Comparison between field research and controlled laboratory research. *Archives of Clinical and Biomedical Research* 1, 2 (2017), 101–104.
- [3] David Balenson, Laura Tinnel, and T Benz. 2015. Cybersecurity experimentation of the future (CEF): catalyzing a new generation of experimental cybersecurity research. *SRI International, Tech. Rep.* (2015).
- [4] Heather Becker, Greg Roberts, Janet Morrison, and Julie Silver. 2005. Recruiting People With Disabilities as Research Participants: Challenges and Strategies to Address Them. *Mental retardation* 42 (01 2005), 471–5. [https://doi.org/10.1352/0047-6765\(2004\)42<471:RPWDAR>2.0.CO;2](https://doi.org/10.1352/0047-6765(2004)42<471:RPWDAR>2.0.CO;2)
- [5] Donald T. Campbell and Julian C. Stanley. 2015. *Experimental and quasi-experimental designs for research*. Ravenio Books.
- [6] Dartmouth College. 2020. *Teach Remotely: How to Teach From Anywhere*. <https://sites.dartmouth.edu/teachremotely/remote-lab-activities-and-experiences>
- [7] C Mitchell Dayton. 2002. Some Key Concepts for the Design and Review of Empirical Research. *ERIC Digest*. (2002).
- [8] Carrie Gardner, Abby Waliga, David Thaw, and Sarah Churchman. 2019. Using Camouflaged Cyber Simulations as a Model to Ensure Validity in Cybersecurity Experimentation. *arXiv:1905.07059 [cs.CR]*

- [9] Urs Hunkeler, Hong Linh Truong, and Andy Stanford-Clark. 2008. MQTT-S—A publish/subscribe protocol for Wireless Sensor Networks. In *2008 3rd International Conference on Communication Systems Software and Middleware and Workshops (COMSWARE'08)*. IEEE, 791–798.
- [10] Internet School Experimental System (iSES). 2018. *Collection of examples for remote laboratories*. <https://www.ises.info/index.php/en/laboratory>
- [11] Jesper Kjeldskov and Mikael B Skov. 2014. Was it worth the hassle? Ten years of mobile HCI research discussions on lab and field evaluations. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. 43–52.
- [12] Tomáš Kozík and Marek Šimon. 2012. Preparing and managing the remote experiment in education. In *2012 15th International Conference on Interactive Collaborative Learning (ICL)*. IEEE, 1–4.
- [13] Roger A Light. 2017. Mosquitto: server and client implementation of the MQTT protocol. *Journal of Open Source Software* 2, 13 (2017), 265.
- [14] Milica B Naumovic and Dragan Zivanovic. 2008. Remote Experiments in Control Engineering Education Laboratory. *Int. J. Online Eng.* 4, 2 (2008), 48–53.
- [15] Edgar Padilla, Jaime C. Acosta, and Christopher D. Kiekintveld. 2021. Cybersecurity Methodology for Specialized Behavior Analysis. In *Digital Forensics and Cyber Crime*, Sanjay Goel, Pavel Gladyshev, Daryl Johnson, Makan Pourzandi, and Suryadipta Majumdar (Eds.). Springer International Publishing, Cham, 237–243.
- [16] Kendall Powell. 2020. Science-ing from home. *Nature* 580, 7802 (2020), 419–422.
- [17] The Royal Society. 2020. *Science in Lockdown: the effects of COVID-19 on research and researchers*. <https://royalsociety.org/blog/2020/05/science-in-lockdown-part-one>
- [18] Katta Spiel, Kathrin Gerling, Cynthia L. Bennett, Emeline Brulé, Rua M. Williams, Jennifer Rode, and Jennifer Mankoff. 2020. Nothing About Us Without Us: Investigating the Role of Critical Disability Studies in HCI. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI EA '20*). Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/3334480.3375150>
- [19] Xu Sun and Andrew May. 2013. A comparison of field-based and lab-based experiments to evaluate user experience of personalised mobile devices. *Advances in Human-Computer Interaction* 2013 (2013).
- [20] Christian Thomsen, Harald Scheel, and Sabine Morgner. 2005. Remote experiments in experimental physics. In *Proceedings of the ISPRS E-Learning*, Vol. 2005. Citeseer.