Safe Double Blind Studies as a Service

Tyler J. Skluzacek
The University of Chicago
skluzacek@uchicago.edu

Suhail Rehman
The University of Chicago
suhail@uchicago.edu

Ian Foster
The University of Chicago
Argonne National Laboratory
foster@uchicago.edu

Abstract

The emergence of IoT devices is revolutionizing various aspects of human life, including healthcare, where the use of such devices can potentially improve health outcomes for millions. However, the efficacy of treatments and protocols based on IoT devices is measured through the use of rigorous double-blind studies, which can be quite expensive to conduct as they traditionally require a third-party mediator. In this paper, we propose CATnIP, a secure, centralized cloud hub for instrumenting and conducting double-blind studies, with an extended focus on seamless integration with IoT devices. This paper outlines the construction and security considerations of CATnIP, the motivations behind creating such a system, and an evaluation based on the Five Safes and Stakeholder frameworks.

I. Introduction

Given the rise of IoT devices in diverse domains such as medicine, technology, and social science, there arises a need to conduct scientific studies over such channels. Further, many studies have strict privacy requirements, including corporate privacy policies, HIPAA compliance, and other SLAs that govern the use and dissemination of an individual's data (and in many cases, disclosure of such data could lead to fines and prison sentences [1]).

Oftentimes these organizations choose to run double-blind studies—trials with a test group and a control group that are organized such that neither the clinicians (who administer treatment and monitor the study) nor the participants know which participants are in which group, but a single study administrator (who conducts the study and provides data to analysts) does [2] (we call this division in knowledge the Information Obfuscation Line [IOL]). An illustration of such a study is illustrated in Figure 1. But when we implement IoT technologies into studies, such methods for inducing double-blindedness prove vastly inefficient and generally insecure. Further, the large amount of clinician-to-patient contact can introduce non negligible levels of statistical bias [3]. We require a means to conduct a double-blind study that can (1) remove the systemic bias of the clinician by limiting point-contact, (2) automate the collection and security of data from IoT devices, and (3) ultimately reduce the role of the clinician to a study monitor (meaning that in many cases, the clinician could be eliminated altogether when such monitoring is unneeded). To accomplish such restructuring, we present CATnIP (ClinicAl TrIal Platform).

CATnIP has three main tenets to its architecture that allow for the secure automation of double-blind studies: (1) a mediator to shuffle groups, reparameterize a study, and manage data control; (2) an administrative/clinical portal by which administrators can create and edit studies, add patients, and reparameterize the study, and (3) a general IoT REST framework that is easy for intermediate software developers to implement, and for which the complexity is hidden for the participant. We assert our main contributions to be as follows:

- Present an architecture that uses a mediator service as an IOL to assign study groups, respond to event triggers (e.g., emergency shutdowns), and handle reparameterizations of the study.
- Dissect the broader definitions of what it means for a study to truly be double-blind, and how in-place systems fail to provide a number of the inherent requirements of such studies.
- Evaluate such a system in terms of the Five Safes model for disclosure threats as well as both benefits and concerns for all involved stakeholders (and discuss how we can relax such concerns in future work).

The remainder of this paper is organized as follows. Section 2 describes our three-part architecture for CATnIP that ensures safe double-blind study in the cloud. Section III reveals an evaluation of our system, both in terms of the security of the data within, as well as the benefits and concerns facing a stakeholder involved with such a system. Section IV presents an analysis and taxonomy of other work in the trial software space, and explains how CATnIP differentiates itself from extant systems. Finally, Section V provides concluding remarks and discusses how CATnIP can be improved in terms of workflow, study parameterizations, and security as the project progresses.

In the remainder of this paper, we assume a HIPAA-level security study by which patient information is regulated to be confidential. Furthermore, certain measurements that are obtained during the study might be sensitive and may require some level of data protection in order to minimize re-identification risk. We have designed the architecture of our system with these requirements in mind by creating two specific systems that communicate through an authenticated REST-based interface: a mediator service, which handles role-based authentication and subsequent management of the clinical trial data, and a clinician/administrator portal that allows access to study data for setup and analysis during the various phases of the trial. We illustrate this architecture in Figure 2.

II. Architecture

A. Mediator Service

Our mediator service consists of a Django REST Framework (DRF) API mediator that handles the access and manip-

978-1-5386-2686-3/17 $31.00 © 2017 IEEE
DOI 10.1109/eScience.2017.82

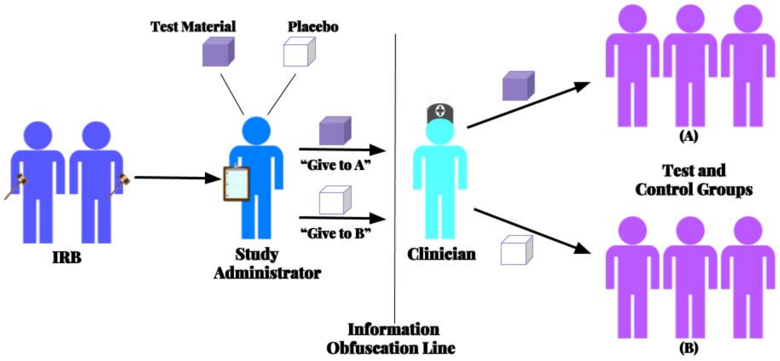

Fig. 1. Depiction of a double-blind study including the IRB, a study administrator, a clinician, and patients.

ulation of all data related to clinical studies. We first describe the data model for the mediator service as well as the role-based access methods for each of the individual data items.

1) Server Data Model: Our core data model consists of **Users** (which have the roles of **Participants**, **Clinicians**, and **Administrators**). Each study is modeled as a **Trial**, which is further divided into **TrialGroups**. Both Trials and TrialGroups can have parameters. The TrialGroup parameters are used to determine controlling variables for the test and placebo groups.

The trial measurements are encoded in the data model in the form of **MeasurementKeys** and **Measurements**. MeasurementKeys consist of individual measurement types (e.g. Heart Rate), and the Measurement type contains the actual measurement along with a capture time as well as transmit time. The MeasurmentKey also contains the encoding for the required data protection mechanism, based on the requirements for the trial as approved by the IRB and study administrators. They can be open disclosure, have aggregated statistics, or have some further form of disclosure protections (k-anonymity or differential privacy).

2) Mediation Mechanisms: The study administrator has the most access in terms of instantiating the clinical trial and its parameters via the clinician/administrator portal (see Section II-B). The core mediation logic of dividing the participants into the test and control groups is automatically determined in the back-end during the trial initialization phase with the available participants. For longer-term studies or studies with open enrollment, the mediation platform can be either configured to assign new participants into a random group, or if required, add them according to group size requirements, as determined by the study requirements.

Clinicians also have access to the clinician/admin portal as described in Section II-B, and can view important information related to the study as long as the data do not violate the data protection requirements as configured by the study administrator.

A participant has only two possible operations: (1) Write out a measurement for a particular trial that the participant is enrolled in, and (2) Pull the trial and/or group parameters for the study. At no point are any data transmitted to the participant regarding the group to which they belong. Since these endpoints are accessible through the authenticated REST endpoints, they can be used with any connected device capable of making HTTPS-based requests.

B. Clinician/Administrator Portal

The clinician and administrator portal provides a means for an administrator to view the current status of their organization's studies, create new studies, and promote clinicians to the study. We use the Flask Principal 0.4.0 library [4] in our server to provide login authentication, Werkzeug 0.12[1] for secure password issuance and storage (passwords use SHA-256 hashing), and UserMixin to create and move tokens between views (and therefore re-authenticate one's role) in the portal. Moreover, over the duration of the study, neither administrators nor clinicians have full access to the study, and all data views are subject to k-anonymity [5] such that the number of people required in a group to view a group-wise aggregate (e.g., mean, median, count) is $k = 3$ or greater (set by the administrator). The remainder of this section explains various attributes of the portal.

1) Creating and Editing a Study: An administrator can create their own study with automatic administrator privileges[2]. One enters a number of parameters to the study, and as of now, one can insert an app-terminating 'trigger' condition that is meant to serve as a safety mechanism. For example, in an application that intervenes when a participant's stress level reaches a certain threshold, if the participant's stress rate reaches some even-higher 'danger' threshold as a result of the intervention (meaning this study could cause more harm than good), the application terminates and must be re-instantiated by the study's administrator. An example of some of the parameterizations for a stress-intervention study is illustrated in Table 1. In this example, the application terminates if it does not hear from the mobile device for 72 hours, the subject's heart rate exceeds 120 bpm, or if 60 days elapse (whichever occurs first). Moreover, 70% of the participant pool is assigned to the Test-Group, and the application is HIPAA-secure (meaning all security measures are activated[3]).

[1] http://werkzeug.pocoo.org/

[2] One should note that this permission can be removed or downgraded at a later time, if necessary.

[3] At the time of this writing, only the HIPAA security level exists, but lesser security measures are straightforward to implement

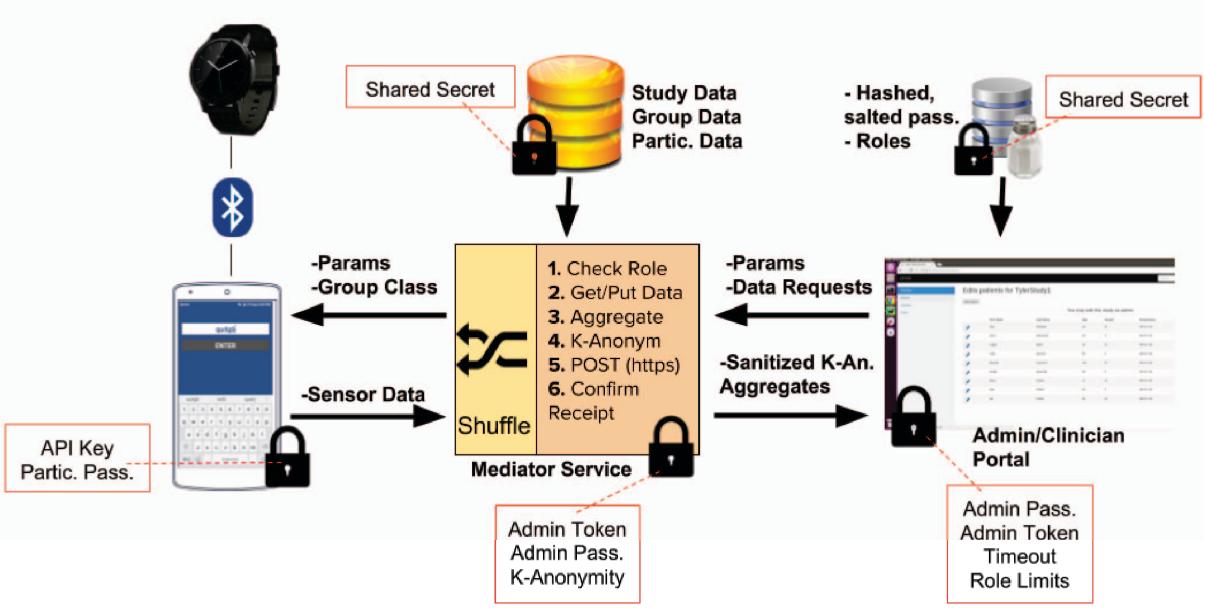

Fig. 2. CATnIP's Architecture and Security Considerations: IoT, Mediator Service, and Admin/Clincian Portal.

An administrator can edit a study after entering their password. The portal has a shared secret with both the mediator service as well as the offsite database containing admin/clincian roles and hashed (SHA-256), salted passwords, as shown in Figure 2. Note that in none of these processes does the administrator have access to subject-level measurement data besides a list of people participating in the study.

2) Adding Patients and Clinicians: When an administrator clicks into the study, they are presented with a view showing the study's participants. Under the HIPAA protocol, only the admin can enter people and see names (clinicians can only see a user ID). Once the participant is added to the study, said participant will receive an email containing the rules for participation in the study and a link to download the application, as mentioned later in Section II-C.

Finally, an administrator can add clinicians to a study, and upgrade them to admin privileges (a decision that should have been previously approved by the IRB). This task also requires a password confirmation and the email approval of the clinician being upgraded.

C. IoT Considerations

We created a sample wearable application that tethers to the mediation service via a smartphone, and allows the study administrators to use the mobile devices' collective sensors, including the Accelerometer (abbreviated in our examples as AC), Gyroscope (GY), and Heart Rate Monitor (HR); and further, one 'computable' variable that a developer can add called Spec that can be calculated from raw sensor data on the device (although a savvy developer can easily add more such variables). At the time of this writing, a software developer needs to publish the application on a download service such as the Google Play Store Beta or TestFlight[4].

After the clinician adds a user to a study, the user is emailed a unique, randomly-generated passcode that can be used to enter the application. In our general application, the user is prompted to hit 'start'. Next, the application measures data from all sensors and sends them to the mediation service for review by confirming the mediator's API key. When a clinician updates the study, the user is sent a brief of the changes made (e.g., "Trigger moved from 70bpm to 60bpm"), and the developer must re-confirm enrollment in the study by entering their passcode into the application once again (as shown in the phone in Figure 2).

This programming model should extend to any application that can connect to the mediator via a REST API, and as more sensor codes are added to CATnIP's study schema, this will become increasingly easy for developers to utilize.

D. Assumptions

After the point of IRB approval, we assume that all administrators are benevolent, and will not maliciously tamper with the study. (However, we still need to guard against malicious actions by participants, clinicians, and hackers.) This assumption should not raise concern due to the legal costs accompanying malicious intent during studies [1]. Further, we assume that our trusted cloud provider (such as Amazon EC2) will not tamper with our virtual machines, which is a fair assumption based on Amazon's HIPAA compliance assurances for PHI stored on Amazon's server [5]. We assume that any relay connections between IoT devices (e.g., the Bluetooth relay between an iPhone and an Apple Watch) is secure. Finally, we

[4] We discuss in Section V how we expect to automate the app packaging and delivery process in the future to improve security

[5] https://aws.amazon.com/compliance/hipaa-compliance/

TABLE I. Study Inputs for Simple CATnIP Intervention Study on Heart Rate

Check-in	Buzz Frequency	Termination Cond.	% Test	Kill Cond.	Pool Parameters	Confidentiality Level
72 hours	Dynamically Adjust	60 Days	0.70	HR >120	{gender=all, race=all}	HIPAA
72 hours	15 Minutes	60 Days	0.30	HR >120	{gender=all, race=all}	HIPAA

assume that participants will not gain access to someone else's device passcode, as participants in the study are supposed to be anonymous.

III. Evaluation

We evaluate our work in two contexts. We first frame our threat model in terms of the Five Safes [6] by looking at how Safe Projects, Outputs, People, Data, and Settings are upheld despite impending threats to patient and organizational data. Second we consider all stakeholders in the model and consider how each explicitly benefits by using a secure double-blind service like CATnIP.

A. The Five Safes

For this analysis we consider the Five Safes model [6] – a framework for evaluating the safety of private or confidential data in the context of a threat model (i.e., what are the threats to the Five Safes, and how do we mitigate them?). In the following, we outline our analysis for each of the Five Safes in a research study run on the CATnIP service:

1) Safe Projects: Safe Projects questions the appropriate use of data in a study in terms of legality, morality, and ethics. We believe this question should be appropriately addressed during the study design and IRB approval phase. With IoT-based clinical trial studies, additional questions are raised in terms of the possibly higher frequency of data collection and the related privacy implications. Although beyond the scope of this paper, we would like to comment that there is an additional onus on the IRB review process to properly vet the proposed data collection mechanisms, end-used IoT application design, and the process of obtaining informed consent from the participants involved. We believe that CATnIP can be used as a tool for finer auditing of IRB-approved studies in order to avoid malpractice and/or general negligence.

2) Safe People: Safe People asks if those with access to both the raw and sanitized datasets can be trusted. We assume that the administrator and clinicians running a study are trusted with following all study-related protocols and have been vetted through the appropriate institutional procedures.

Participants that are involved with clinical trials are a potential hazard in terms of potentially falsifying measurement data or tampering with the actual transmission of the data being sent over to the server. Apart from secure transmission of data using HTTPS, the IoT applications can be digitally signed to ensure that the application binaries are not modified. There are much simpler hazards in terms of device tampering or even simply handing over the IoT devices to someone else, and these are harder to detect or counter against. Such behavior could potentially be mitigated using appropriate EULAs or legal paperwork that might have punitive consequences for such behavior.

Finally, system administrators entrusted with managing the actual server infrastructure and who have root access to the mediation server do have access to raw data. Use of encrypted database techniques such as those employed in systems like CryptDB[7] could mitigate these hazards and can be left to future work.

3) Safe Settings: Safe Settings considers whether the data facilities (in this case, the system) are protected from unauthorized use. This is the strength of CATnIP – until the termination of the study, no parties involved in the study have access to the raw data. Further, even access to k-anonymized aggregate data must be vetted by at least one administrator in the study.

4) Safe Data: Safe Data questions the disclosure risk of the data. Let's assume a case where some malicious entity received access to all of the data in the database. As mentioned in the discussion of Safe People, use of an encrypted database could render the data effectively useless to an external administrator. We believe that the other security mechanisms we have in place, including roles, shared secrets, and off-site passwords will make a full data breach highly unlikely.

5) Safe Output: Safe Output questions whether the statistical outputs from a study can be disclosive of individual participants in the study. CATnIP does not directly create any of its own statistical output, but rather passes the final results to an independent analyst (who may double as a study administrator) who attempts to reach a conclusion regarding some IRB-approved research hypothesis. Protecting against an analysts' leakage of data is beyond the scope of the CATnIP project, but one could theoretically apply differential privacy [8] to the data to create a synthetic dataset.

B. The Stakeholder Model

The stakeholder model [9] presents a means of evaluation for a consumer-level research software system like CATnIP. For this analysis, we consider four groups of stakeholders: Trial Conductors, Test Subjects, Analysts, and The Scientific Community. To be as unbiased as possible, we consider both benefits and costs to each group when using CATnIP.

Trial Conductors. We classify both clinicians and study administrators as Trial Conductors. CATnIP has an intuitive interface, and fewer data linking steps are required to involve participants in trials of all sizes. The identities of trial conductors are protected in our schema, via a combination of shared secrets, tokens, protected passwords, and regular role confirmations. Further, clinicians do not have to worry about data disclosure because they will not see the data before it is sent to the analyst at the conclusion of a trial. (We consider the possibility that an administrator discloses the name of a participant in the study, but it is effectively not worthwhile to run a sensitive study in which the identity of the person is never known by anyone involved with the study.)

Test Subjects. This group refers to all participants in all research studies. CATnIP's IoT programming model hides the complexity of configuring an app. A user receives a link from

CATnIP, downloads an application, enters a passcode, and then follows whatever instructions are provided by the administrators (and can explore the study's privacy parameters before download). This ease of access improves upon current methods in two ways: (1) efficiency—a participant can participate in many studies, minimize technical difficulty, and even minimize the number of visits to, for example, a medical office; and (2) security—CATnIP's data access guarantees acknowledge that data access is limited over the course of the study, which means the identification risks are minimized relative to the clinician typing data into a spreadsheet.

Analysts. Those who analyze resultant trial data are the analysts (although oftentimes analysts can double as study administrators) who turn data into an evaluation of a research hypothesis. The main benefit of CATnIP to the analyst is the clear distinction of variables and the methods by which measurements are collected. There is no room for clerical error, as the entire data collection and delivery process is automated. A primary concern is the possibility that an unsupervised party misuses a measurement instrument (e.g., the *'put the FitBit on the dog'* phenomenon). Such issues can be avoided in the future by periodically checking for data anomalies in raw data, or as previously stated, by issuing punitive EULAs.

The Scientific Community. This community refers to the Institutional Review Board, those who create research, and those who translate and disseminate knowledge to non-scientific communities (e.g., the media). The ease of adding geographically separated participants to a study through an automated system will increase the number of possible studies, and the sample sizes in each of these studies can be increased. Further, the minimization of clinician-to-participant contact effectively reduces the amount of potential systemic bias in a study [3]. Further, as the Five Safes are near-fully upheld in CATnIP, running a study through this system could serve as a 'stamp-of-approval' for ethical institutional research (in addition to the approval provided by the IRB) in the future. One misstep one running a study via CATnIP should avoid is the dilution of a study by relaxing exclusion criteria too far in order to increase sample size.

IV. Related Work

In this section we discuss some of the prior work that relates to CATnIP. Our understanding is that such work can be chiefly categorized into two disparate domains, the first being work that relates to automating the management of clinical trial protocols; and the second being recent innovations in IoT devices and platforms that relate to both personal and provider-based healthcare management. We believe our work is innovative in trying to bridge the gap that exists between these two domains.

A. Experimental Protocols and Design

Blind testing is crucial for reducing or eliminating biases that can affect the outcome of an experiment. In double-blind experiments, both the tester and the subject will have information about the test masked until the outcome of the experiment is known. Double-blind trials are used extensively in medicine to determine the efficacy of treatments and to counter the placebo effect when working with patients, with the modern clinical protocols first established in [10].

In order to further reduce bias, the Randomized Controlled Trial (RCT) [11] is considered to be the gold standard for clinical trials, and can significantly reduce the selection bias. A number of commercial software companies provide software systems that help in the randomized selection process such as [12], [13], [14].

B. Internet-of-Things in Healthcare

The Internet-of-Things (IoT) is the connected network of various devices that can collect and exchange data. The devices themselves can vary from tiny pressure sensors on a ship's sail to entire automotive vehicles. The IoT allows objects to be sensed or controlled remotely across existing network infrastructure, creating opportunities for more direct integration of the physical world into computer-based systems. In addition to reduced human intervention, this results in improved efficiency, accuracy, and economic benefit.

The potential impact of IoT in healthcare has been studied in [15], where they project healthcare to become the largest market sector by 2025. Several start-up companies now operate in this space. NightWare is a wearable smartwatch application that helps in suppressing PTSD symptoms during sleep [16]. Moving Analytics is developing wearables to help clinicians track patients recovering from cardiac ailments [17]. Other examples include Lucid Dreaming Acceleration [18] and the Athlete Performance Wristband [19].

Our preliminary survey has yielded little to no results on existing solutions to develop platforms to manage and connect IoT devices to researchers with the aim of conducting RCTs and interventional studies. TrueVault [20] aims to provide HIPAA compliant cloud storage services for healthcare applications (or as close as one can get to such compliance) when using unrestricted personal devices.

V. Conclusion

We have provided an initial foray towards running secure, double-blind technology trials in the cloud. We presented a unique central service that acts as a mediator serving as the only omniscient party in the trial. Study participants do not need to know the identity of the conductor beyond the organization, and the contact between a clinician and patient is minimized. By decreasing the number of points of contact between participants and study conductors, the clinicians and patients have less space to inject personal bias over the course of a study.

For future work, we consider how to further enhance CATnIP by improving the tightness of the workflow and security of the system. We can enhance the workflow by automating more steps for a safe study, for example, by integrating a consent system for highly-sensitive studies in which the administrator can outline a number of necessary consent-steps to be activated both before and during a study to ensure participants' willingness. There is also space for consent by an auditing party, for example like an IRB representative, to ensure that the study's data safety and ethical requirements are upheld. Finally, in order to improve the security of such a system, we require a means to further authenticate administrators and clinicians logging into the portal. We consider doing this with an AWS-esque key-pair [21] that one must have on a dedicated clinical study machine and/or by using an existing OAuth2 identity provider, such as Google [22]. Further, a complete work in this space would include a setting to automatically build and ship an application by submitting simple computation through the clinician's portal in order to reduce the number of errors (and subsequently, potentially-costly study restarts) and tighten the security of the application (fewer people see the application when it is pre-compiled).

Acknowledgements

We acknowledge Computation Institute Summer Intern Benjamin Glick for sharing his expertise in Flask and the D3.js graphical library. This work was supported in part by NSF grant ACI-1148484 and by the DOE under contract DE-AC02-06CH11357.

References

[1] C. Elliott, "University of minnesota blasted for deadly clinical trial," Apr 2015.

[2] H. Bang, L. Ni, and C. E. Davis, "Assessment of blinding in clinical trials," *Controlled clinical trials*, vol. 25, no. 2, pp. 143–156, 2004.

[3] J. Margraf, A. Ehlers, W. T. Roth, D. B. Clark, J. Sheikh, W. S. Agras, and C. B. Taylor, "How blind are double-blind studies," *J Consult Clin Psychol*, vol. 59, no. 1, pp. 184–87, 1991.

[4] M. Grinberg, *Flask web development: developing web applications with python*. " O'Reilly Media, Inc.", 2014.

[5] S. Gao, J. Ma, C. Sun, and X. Li, "Balancing trajectory privacy and data utility using a personalized anonymization model," *J. Netw. Comput. Appl.*, vol. 38, pp. 125–134, Feb. 2014.

[6] T. Desai, F. Ritchie, and R. Welpton, "Five safes: designing data access for research," 2016.

[7] R. A. Popa, C. Redfield, N. Zeldovich, and H. Balakrishnan, "Cryptdb: protecting confidentiality with encrypted query processing," in *Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles*, pp. 85–100, ACM, 2011.

[8] C. Dwork, "A firm foundation for private data analysis," Association for Computing Machinery, Inc., January 2011.

[9] I. Foster, "Research infrastructures for the safe analysis of sensitive data," *Annals of the American Academy of Political and Social Science*, Jan. 2018.

[10] N. T. K. Harry Gold and H. O. H, "The xanthines (theobromine and aminophylline) in the treatment of cardiac pain," *Journal of the American Medical Association*, vol. 108, no. 26, pp. 2173–2179, 1937.

[11] T. C. Chalmers, H. Smith, B. Blackburn, B. Silverman, B. Schroeder, D. Reitman, and A. Ambroz, "A method for assessing the quality of a randomized control trial," *Controlled Clinical Trials*, vol. 2, no. 1, pp. 31 – 49, 1981.

[12] "Cytel FlexRandomizer." http://www.cytel.com/software/flexrandomizer. Accessed: 2017-04-28.

[13] "Maldaba Randomized Controlled Trial Software." https://www.maldaba.co.uk/products/randomised-controlled-trial-software/. Accessed: 2017-04-28.

[14] "RCT-YES Software." https://www.rct-yes.com/. Accessed: 2017-04-28.

[15] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys Tutorials*, vol. 17, pp. 2347–2376, Fourthquarter 2015.

[16] "NightWare." http://night-ware.com/. Accessed: 2017-04-28.

[17] "Moving Analytics." https://www.movinganalytics.com/. Accessed: 2017-04-28.

[18] "Aurora Dreamband." https://sleepwithaurora.com/. Accessed: 2017-04-28.

[19] "Whoop." http://whoop.com/. Accessed: 2017-04-28.

[20] "TrueVault." http://truevault.com/. Accessed: 2017-04-28.

[21] "Amazon aws ec2 private key pairs," Jun 2017.

[22] Google, "Google oauth2 developer guide," Jun 2017.