Volume I: Technical and Management Proposal

Cover Page

BAA Number	DARPA-BAA-15-29
Technical Area	(2) Human Data Interaction (HDI)
Proposal Title	**Sunlight is the Best Disinfectant: Increasing Privacy through Awareness with the Hubble Scalable Web Transparency Infrastructure**
Lead Organization	The Trustees of Columbia University in the City of New York
Type of Business	Other Educational
Contractor's Reference Number	RASCAL PT-AABL4408
Technical Point of Contact	Prof. Roxana Geambasu Department of Computer Science, Mail Code 0401 Columbia University, 1214 Amsterdam Avenue New York, NY 10027-7003 212-939-7099 (v) roxana@cs.columbia.edu
Administrative Point of Contact	Kammy Lou Cabral Director Sponsored Projects Administration 615 West 131st Street, Room 254, Mail Code 8725 New York, NY 10027-7003 (212) 854-6851 (v) ms-grants-office@columbia.edu
Subcontractor Information Technical Point of Contact	University of Washington Prof. Franziska Roesner Department of Computer Science & Engineering, Box 352350 University of Washington, Paul Allen Center, 185 Stevens Way Seattle, WA 98195-2350 206-221-8248 (v) 206-543-2969 (f) franzi@cs.washington.edu
Administrative Point of Contact	Lynette Arias Director of Sponsored Research University of Washington, Office of Sponsored Programs 4333 Brooklyn Ave. NE Seattle, WA 98195 206-543-4043 (v) 206-685-1732 (f) osp@u.washington.edu
Award Instrument Requested	Grant
Period of Performance	09/01/2015 – 08/31/2019
Places of Performance	New York, NY; Seattle, WA
Proposal Validity Period	120 days
Prime DUNS Number	049179401
Prime TIN	13-5598093
Prime CAGE Code	1B053

Contents

1 Executive Summary

Motivation and Goal: Today's web services – such as Google, Amazon, and Facebook, as well as third-party advertisers less visible to users – collect and leverage user data for varied purposes, including personalizing recommendations, targeting advertisements, and adjusting prices. At present, users have little insight into how their data is being collected or used and how that affects them. This lack of awareness prevents them from making informed choices about the services they use, what they should be revealing to those services and what not, or what protection tools they should use to prevent misuse. Our goal is to develop *user awareness tools* that will help users gain a better understanding of the implications of their online actions by revealing to them concretely how their data is being collected and used by the services to target them. For example, one user awareness tool could reveal what specific data within a user's profile – such as emails, prior browsing behaviors, etc. – are being targeted by each advertisement they receive. Another tool could reveal to a user that she is seeing a differentiated price, and specifically which data within her profile triggered that differentiation. In support of such tools, we propose to build *Hubble*, an extensible, generic, and scalable infrastructure that provides the necessary scientific methods and programming abstractions to facilitate the building of many such user awareness tools. Using Hubble, we will develop and evaluate several user awareness tools, and will study how transparency and awareness can help shape user actions and enable them to better manage their online privacy. Our effort targets *Technical Area #2* (Human Data Interaction) and is *fundamental research*.

Key Technical Challenges: Constructing user awareness tools raises significant and unresolved challenges. First, once data is given out to a service, how can one still track its use? Tracking data in a controlled environment, such as a modified operating system, language, or runtime, is an old problem with a well-known solution: taint tracking systems [19,27,32,68]. However, is it possible to track data in an uncontrolled environment, such as the Web? Can robust, generic mechanisms assist in doing so? What kinds of data uses are trackable and what are not? How would the mechanisms scale with the amount of data being tracked? Second, constructing user awareness tools that do not themselves create new privacy challenges is a difficult challenge. Intuitively, to reveal how data is being used, a user awareness tool needs to monitor that user's data, and perhaps share it with a third party that aggregates data from multiple users. Why should the users trust those tools and the third-party that run them, and how can we minimize that trust? Third, quantifying the effect of user awareness tools on the end-users is an open question. For user awareness tools to be effective, they must not only help educate users – and watchdog organizations like the Federal Trade Commission (FTC) or the Electronic Frontier Foundation (EFF) – about data collection and use, but they must provide useful and auditable actions that users can take to manage the privacy of their data.

Review of Proposed Technologies: Hubble will develop both the tools and the necessary building blocks to increase users' awareness over what happens with their data once they share it with web services. The key intuition is to XXX. Doing so at scale, generically, and with privacy-preserving properties is challenging. **[Write this after we develop the proposal further.]** xxx

Current Approaches and Limitations: Our project will create *robust, generic user awareness tools to track the use of personal data at fine granularity (e.g., individual emails, photos, or visited websites) within and across arbitrary Web services.* At present, hardly any such tools exist, and the science of tracking the use of personal Web data at scale and at a fine granularity is extremely limited. Our own recent system, XRay [41], includes some preliminary results that transparency at fine granularity is possible, but does not address any of the significant scaling, privacy, and usability challenges defined above. We have also previously developed TrackingObserver [57] to detect third-party trackers on the web, but it remains limited in terms of the types of data collection it can detect (notable, omitting fingerprint-based trackers) and does not provide information directly useful to end users. Other transparency systems, such as Bobble [64], AdFisher [**?**], and OpenWPM [28], are either not generic (e.g., Bobble reveals personalization of news and search results on based on a few user attributes but would be hard to extend to other use cases) or operate at small scale [**?**, 64].

Expected Impact: The greatest impact of our work will be to increase user awareness about the implications of their online actions and to provide building blocks for tools that empower users to manage the information that they reveal. We believe that a vital part of protecting private data that users knowingly provide to third parties is to enable non-expert users to *know more*, *take action*, and *verify the results of their actions*. Moreover, we believe that by empowering users, as well as third-party privacy watchdogs, with transparency tools we will help transition the web services world toward a more privacy-aware future. In Louis Brandeis' own words – "Sunlight is said to be the best of disinfectants; electric light the most efficient policeman" [**?**]. Hubble will help bring a new level of oversight and accountability into a very obscure Web world, thereby putting pressure on web services to be more privacy aware. Finally, while this proposal focuses on awareness tools and building blocks for the Web, we believe that our technologies will be applicable more broadly to track how sensitive information – be it users' personal data, proprietary enterprise information, or classified defense data – is being used (or abused) by the parties that obtain it (such as web services, partner enterprises, or foreign governments). We thus expect that extended versions of Hubble could be applicable to use cases of national importance beyond protecting and increasing end-user privacy on the web. To engender this level of impact, we will make all our source code available open source and will make explicit efforts to transition our technology into users hands and/or commercial products, as we have done in the past with other technologies.

Cost, Duration, and Team: Our proposed effort will last 4.5 years (starting on 09/01/2015), with a total cost of $3,960,419. The team members are from Columbia University (Geambasu, Chaintreau, Hsu) and University of Washington (Roesner). Their expertise spans *systems* (Geambasu), *theory and social networks* (Chaintreau), *statistics and machine learning* (Hsu), and *security and human factors* (Roesner) – all areas required for a successful project. All PIs have long-term expertise and achievements in privacy, and several PIs have a history of transitioning their systems into industry.

2 Goals and Impact

Many of today's pervasive practices that collect and use user data are invisible, or at best unexpected, to users. For example, web and mobile applications commonly collect and aggregate information about users (including browsing behaviors, location, and unique identifiers) for the purposes of targeted advertising or other types of personalization [41, 54]. Many of today's users have some notion that this data collection is happening (e.g., through extensive media reporting on the topic [60]) and that they are exchanging some amount of private information for the use of free services (email, search, social media). Indeed, these practices are typically disclosed in terms of service agreements, to which users must technically agree to use an application or service. However, users' understanding of the extent of this data collection, as well as its use and implications, remains limited and abstract [62]. **Thus, a necessary goal on the path to protecting private data that users knowingly provide to third parties is to help non-expert users *know more*, *take mitigating actions*, and *verify the results of their actions*.**

To this end, we propose the design, development, and evaluation of a new generation of **user awareness tools** that help non-expert users better understand and monitor the data collected about them and how it is (or might be) used. We identify a set of goals for effective user awareness tools:

1. *Actionability:* Beyond just displaying information about private data collection and use to users, an effective user awareness tool must be actionable — that is, users must be able to do something with the information they learn. Though it can be useful to simply inform users about the amount of data invisibly collected about them to build support for broader efforts to manage such collection, such solutions have limited effect on individual end users at present.
2. *Auditable results:* Once a user takes an action to mitigate data collection or use based on increased awareness, it is important that the user be able to audit the results of his or her action. In other words, users should be able to answer the question: "Are my tools, actions, and mitigation strategies actually doing what I expect?"
3. *Attribution:* An effective user awareness tool should allow users to attribute data collection and use to the specific entities responsible. For example, when multiple third-party trackers are loaded on a web page, an effective tool would allows users not just to identify their presence but to trace back particular page content (e.g., ads) to the responsible third party. This attribution helps with both actionability and auditable results, as it helps users understand who is (or is not) doing what.
4. *Awareness about use, not just collection:* We must help users understand not just what data is collected about them but also the potential uses of that data. We cannot expect that non-expert users will be able to extrapolate all possible implications of revealing or allowing certain data to be collected, particularly when multiple third parties collecting data interact in unexpected ways. Thus, our user awareness tools must help users understand and anticipate these implications in order to help them make informed decisions about which data they are willing to share with whom.

Myriad of aspects are interesting to reveal about personal data on the Web. For example: can we build tools to reveal to users how their data is leveraged to target advertisements or recommendations, whether shopping or mortgage sites are using their browsing histories or Facebook profiles to adjust their prices, whether their purportedly encrypted email service is actually decrypting their emails and using the data for its marketing purposes, or whether a service shares their data with third parties – and then how those third parties use the data? For each case, can we reveal exactly which specific data item (or items) that they share with their services – such as emails, documents, locations, or previously visited websites – trigger the specific ads, recommendations or prices? Such visibility, we believe, would be beneficial to the end users to better understand the implications of their online actions, as well as to assess the effectiveness of any defenses they apply.

Unfortunately, constructing user awareness tools that reveal these and many other potentially interesting aspects about the data's journey on the web is extremely difficult today, due to a lack of scientific methods to both *detect* data collection and use and *surface* it to end users in effective and actionable ways. For example, a number of tools exist that visualize third-party web trackers (e.g., Ghostery [31], Lightbeam [46]). While these tools can help users understand how many trackers they encounter in their browsing experience, and allow users to block individual trackers, they lack desirable properties including attribution — that is, users may know that a tracker is present on a webpage, but not which parts of the page were affected, e.g., which ads were placed by that tracker. The lack of attribution also limits the auditability of effectiveness, as it can be hard for non-expert users to verify that anything is different when a tracker is reported blocked. Finally, hardly any tools exist today, which can expose to the users how their data is being used by the services that collect it. A few efforts have recently been made (e.g., AdReveal [**?**], Bobble [63], AdFisher [**?**], and our own XRay system [41]), but they are all primitive in both detecting data use by Web services and effectively surfacing that information to the end users.

Thus, **our specific goal is to develop not only the first *effective and actionable user awareness tools* that reveal specific aspects of personal data collection and use on the web, but also *the science and infrastructural support* for building many such tools in the future**. More specifically, as part of this program, we will develop *Hubble*, an extensible, generic, and scalable infrastructure that will provide the necessary scientific methods and programming abstractions to facilitate the building of a new generation of user awareness tools for for the web. Hubble's two main scientific contributions are: (1) providing an *extensible, scalable, and dynamic architecture* leverages statistical methods in unique ways to accurately detect tracking, targeting, personalization, and discrimination in black-box services based on observations of differentiated user profiles, and (2) providing primitives for *effectively surfacing to end users* information about detected data collection and use.

To drive Hubble's design, we will develop and evaluate at least four user awareness tools, which leverage and inform Hubble's programming abstractions to detect and visualize various aspects about online data collection and use for targeting, personalization, and discrimination. The specific tools are: (1) *CollectionObservatory*, a tool that detects and visualizes third-party web content that invisible collects information about users' browsing behaviors; (2) *AdObservatory*, a tool that detects and visualizes how third-party web trackers leverage the information they collect about the users – such as visited pages, Facebook Likes, or explicitly shared information – to target ads at them; (3) *DiscriminationObservatory*, a tool that detects and visualizes personalized content present on arbitrary websites, with a particular focus on personalized prices or offers on ecommerce, lending, and mortgage websites; and (4) *LocationObservatory*, a tool that detects and visualizes **[XXX]**. xxx

If successful, our work will lay the first scientific foundations and technology for comprehensive tracking of data collection and use within and across the Web. We foresee multiple domains of impact for our technology. First, by increasing user awareness of how their data is being used on the Web, we hope to make users more mindful of service selection and usage. Second, by enabling robust and scalable transparency tools, we can empower privacy watchdogs – such as journalists, Federal Trade Commission (FTC) investigators, consumer protection agencies, or internet freedom groups (e.g., EFF) – to keep this giant, complex, and ever-changing Web in constant check to discover any abuses. Third, by enabling transparency at scale and from the exterior, we hope to usher in a new era of voluntary transparency and responsible data behaviors by the web services themselves. Fourth, we believe that our work can integrate well with personal data protection technologies that will be developed as part of the Brandeis program, including TA1 and TA2 technologies. We discuss our vision of such integration in Section 3. Finally, we believe that our technologies will be applicable more broadly to track how sensitive information – be it users' personal data, proprietary enterprise information, or classified defense data – is being used (or abused) by the parties that obtain it (such as web services, partner enterprises, or foreign governments). We thus expect that extended versions of Hubble could be applicable to use cases of national importance beyond protecting and increasing end-user privacy on the Web.

3 Collaborative Research Team Concept

[Guys: I need your help on this section! I am unsure what DARPA wants to see here. I think we need to talk about how we'll be working with TA3, but what I see as obvious is a relationship with TA1 and TA2 technologies (see below a description of the relationship I envision). Please tell me what you think and how we should improve.] xxx

We foresee significant opportunities for integration with other TA1 (Privacy-preserving Computation) and TA2 (Human Data Interaction (HDI) technologies. Specifically, transparency and awareness tools (which focus on revealing data (mis)use and are this proposal's focus) are complementary to protection tools (which focus on preventing data misuse and will likely be the focus of other TA1 and TA2 proposals). First, transparency and user awareness tools can help incentivize adoption of protection tools. Users are known to have poor and misaligned models of digital threats, including threats to their privacy. As a result, they are often considered incapable or unwilling to adopt protection mechanisms, which inevitably, result in some level of inconvenience or slowdown. By revealing specific implications of how they are being targeted – e.g., that they are being targeted by advertisements because they are gay or lesbian,[1] or that they are being offered poorer insurance offers because they have "Liked" a bungee jumping group on Facebook – we hope to instill in users a greater sense of urgency when it comes to protecting their online privacy. A recent paper [49] provides initial support for our hope: "Awareness of the potential consequences of data aggregation, such as Facebook or Google knowing what other websites one visits or one's political party affiliation, was associated with greater likelihood of reporting concern about unwanted access." This finding suggests that greater transparency for users is critical – and effective – to helping them shape accurate opinions about the collection and (mis)use of their data.

Second, transparency tools can enable *auditing* of the effectiveness of protection tools. For example, consider a TA2 project that develops a novel interface for expressing high-level human intentions (of the form "I don't want my data to be used in such and such a way") and which, perhaps with help from a TA1 project, enforces those intentions. How does a user know that she has configured her protection tool correctly, or that the protection scheme truly enforces her intentions? As another example, consider a TA1 project that provides novel encrypted-email technology (e.g., with spam detection and search enabled). How does a user know that the service adopting that technology does not actually insert a back-door that lets it decrypt the data and use it for other purposes? Transparency tools like those we propose – either used by end users or by privacy watchdogs such as the FTC or investigative journalists – could reveal such abuses.

To facilitate collaboration with TA1, TA2, and TA3 projects, we have dedicated a specific task to integration and evaluation of our technologies as part of the collaborative team (see Section 7). We envision integrating with TA3 (Experimental Systems) to (1) evaluate the effectiveness of our approaches at detecting data (mis)uses that might be inserted in such systems and (2) revealing

[1] In preliminary experience, we have found that advertisers do target such aspects as homosexuality, race, religion, and challenging health or financial conditions [41].

the effectiveness (or ineffectiveness) of TA1 and TA2 protection tools protecting user data on such systems, ultimately helping iterate on and improve those designs. **[I don't know what I'm saying. Guys – help!]** xxx

4 Technical Plan

Our project will develop both the first *tools* and the first *extensible, scalable, and robust infrastructure* needed to track data use in the uncontrolled Web. While others have previously studied various specific data uses (e.g., price discrimination in Orbitz [24, 34], coarse-grained ad targeting studies [12, 43], or advertising discrimination studies [58]), to the best of our knowledge we are the first to actively seek generic, accurate, and scalable systems to track personal data use on the Web.

Our specific plan involves efforts in three thrusts, which we will execute in parallel. First, we will develop the *Hubble infrastructure and programming abstractions*; it will provide a set of highly reusable and scalable components that will facilitate the building of transparency and user awareness tools to lift the curtain on how personal data is being used. Second, we will build a set of robust, scalable, and usable *transparency and user awareness tools* that leverage those abstractions and enable users, journalists, and investigators to obtain visibility into Web services' data uses. Third, we will leverage these tools to run *measurement studies* of various data-driven platforms, such as targeted advertising ecosystems, online trackers, and online price discrimination. These studies will increase awareness, and perhaps help uncover examples of data mistreatments, which will provide the grounds for an informed societal argument on the need for increased voluntary transparency by the services. Moreover, our studies will evaluate the impact of transparency tools on end-user mental models of the privacy implications of online data sharing.

Our approach is to make progress in all three thrusts in parallel. The process will be inherently iterative, with the design and implementation of each abstraction in Hubble being informed by prior tool instantiations and measurement studies. The remainder of this section describes our specific plans for each of these thrusts.

4.1 Thrust 1: The Hubble Transparency Infrastructure and Abstractions

Hubble and its abstractions support the development of transparency and user awareness tools that reveal aspects about data use on the web. More specifically, Hubble will support the development of any tool that aims to reveal which specific data *inputs* – such as emails, documents, Facebook Like's, or previously visited websites – are being used to target which specific service *outputs*, such as advertisements, recommendations, or prices. Hubble offers an infrastructure and programming abstractions that vastly simplify the building of such tools. Examples of tools that could benefit from Hubble support include: CollectionObservatory, AdObservatory, DiscriminationObservatory, LocationObservatory (all tools we propose), as well as a number of web transparency tools that others have recently built (e.g., Bobble, AdFisher, as well as tools used for measurements of personalization on the Web in the past [**?**, **?**, 33, 34, 45, 64], all appear amenable for Hubble).

To support the needs of such tools, we formulate the following specific design goals for Hubble:

1. *Generic abstractions for non-expert transparency developers.* We envision regular developers leveraging the Hubble infrastructure and abstractions to build new transparency tools that reveal various aspects about personal data on the web. While our project will create a few such tools – as examples and driving forces for Hubble's design – we anticipate a great need

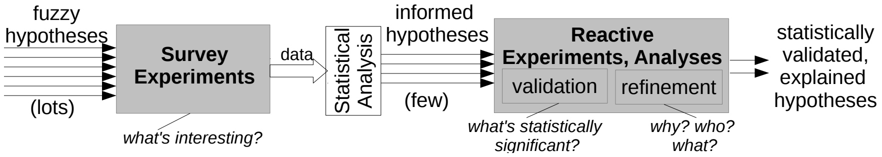


Fig. 1: **Example Experiment Workflow in Hubble-based Transparency Tool.**

for many more such tools that reveal many different aspects about personal data use on and across the vast and diverse World-Wide Web. Our goal in designing Hubble is thus provide generic, robust, and broadly applicable abstractions so developers other than us can continue and scale up our web transparency effort. We target developers with reasonable programming skills but who are not experts in large-scale system design, sophisticated statistical methods, or privacy-preserving protocols. (Section 4.1.1)

2. *Statistically sound inferences and auditable justification.* Hubble must provide statistical justification for its targeting hypotheses. An important Hubble design decision is to leverage well-established and well-understood statistical methods, but leave open the possibility of tapping into other methods in future extensions. This requirement represents a departure from most prior work in web transparency [**?**, **?**, **?**, 33, 34, 45, 64], which rely upon ad-hoc, difficult to extend and enhance solutions for targeting inference. Any validations necessary to support or enhance statistical justification for an inference must be run in real-time, immediately after Hubble initially makes that inference, so that the necessary ephemeral evidence is collected before it disappears from the rapidly changing web. Where possible Hubble seeks to make *causal inferences*, which we believe are easier to grasp than correlations for regular users and developers. (Section 4.1.2)
3. *Privacy-preserving transparency protocols.* Hubble must be ready to allow transparency and user awareness tools to operate with real user data. Protecting the privacy of this information raises significant challenges to targeting inference. Hubble must thus incorporate privacy-preserving protocols and targeting inference algorithms that transparency developers can use to guarantee privacy to their users. (Section 4.1.3)

4.1.1 The Hubble System

[This will change to match abstractions for end-user tools. It will also be shorter.] xxx

In support of such tools, Hubble offers one core abstraction: the notion of an *experiment*. An experiment is specified by a set of *inputs* that the developer hypothesizes might be used for targeting of a particular type of *outputs* (e.g., the websites in a user's history might be used to target ads on the web). In practice, transparency and user awareness tools combine multiple experiments into more complex workflows to reveal a desired aspect. While Hubble imposes no particular structure on these workloads, we find one design pattern particularly useful in practice.

Hubble implements three functions for experiments: (1) support for *XXX experiments* that efficiently survey many hypotheses at once, (2) primitives for *statistical analysis* of the data collected from those experiments to provide statistical inference and confidence levels despite noise, and (3) a *reactive architecture* that lets developers compose experiments – large and small – into workflows in which experiments build upon previous findings to further investigate and validate them. After a brief overview, we describe each core function in turn.

Fig.1 shows this pattern. The developer first runs a large-scale survey experiment to determine interesting hypotheses from a sea of possibilities. Hubble's survey experiment abstraction lets the developer simultaneously evaluate many possible targeting hypotheses, using powerful ideas from compressed sensing [15,23] to minimize the experimental costs and maximize statistical efficiency. Data from the experiment (reports of output observations in particular profiles) feeds into the statistical analysis engine, which yields a set of *plausible, informed hypotheses* (specific inputs or sets of inputs that appear targeted by specific outputs). Several of these hypotheses may have confidence scores that are high enough to suggest some effect but perhaps not high enough to be useful to an developer. These plausible hypotheses are used to trigger a set of follow-up experiments, called *reactive experiments*, that focus on specific hypotheses and attempt to either boost their confidence (*validation experiments*) or provide a more detailed investigation (*attribution experiments*).

Taking advantage of the adaptive nature of Hubble's reactive architecture, validation experiments can typically be less statistically complex than survey experiments and thus afford more statistical power. For example, a validation experiment may focus on just the specific inputs that have are believed to have triggered the output, and this number may be far fewer than the original number of inputs from the survey experiment. Refinement experiments are also informed by survey results, and typically pipelined after the validations, and they attempt to pose more in-depth questions about the plausible hypothesis. For example, Bob from the preceding example may create an initial experiment (survey) that collects information about a large number of sites that are suspected of having targeted ads and then a series of attribution experiments that determine what sites and which specific trackers ads target. All of these experiments are defined by the developer up-front by implementing Hubble's API. Hubble executes the workflow in real-time according to the developer's specification, and returns a set of statistically validated, explained hypotheses.

To launch experiments in Hubble, a developer registers the first experiment in her workflow with the Controller by calling `registerExperiment` in the Hubble API (Fig. **??**(b)). This registration requires four main parameters on top of a unique ID for that experiment. First, she declares a set of profiles that will be populated with varying inputs to detect targeting. Profiles can be either soft profiles (represented by cookies and other browser state, requiring no *a priori* setup) or accounts (such as Google accounts). Second, she declares the set of inputs on which she wants to detect targeting, as well as the type of these inputs. Inputs may be categorical (e.g., gender) or binary (e.g., inclusion indicators for subsets of input webpages or e-mails). Third, the developer declares the uncontrolled variables. Uncontrolled variables are inputs that may influence the targeting, but over which Hubble has no real control in this setting. This can include the IP address from which a profile was used, or the time-of-day when the data was collected. These variables are included in the analysis to keep them from polluting targeting inferences, but will not be varied in a controlled way to detect targeting. Last is the data collection procedure to invoke for that experiment; it is passed the experiment ID and the ID of the profile to exercise.

The Controller (Fig. **??**(a)) assigns the input values to the different profiles. These values are determined independently, and (by default) chosen uniformly at random. The Controller then saves the mapping profile-to-inputs in the `Experiments` table, and queues a data collection job for each profile in a reliable job queue. Each profile will be exercised by a data collection worker, which runs the collection procedure to populate it with the specified inputs (e.g., visits the set set of input webpages), and then collects the service outputs offered to its profile (e.g., the ads shown on the visited pages). The data collector reports any observed outputs as well as the values of uncontrolled variables to Hubble via the `addObservation` function in the Hubble API. The function persists information about the context of the observation into the `Observations` table in the reliable database for subsequent analysis.

Timeliness is vital for effective investigations of the ever-changing web. A key feature in Hubble is to both identify plausible targeting hypotheses, and validate and refine them in as close to real time as possible. Hubble monitors the number of `Observations` for each output. When sufficient data is available for a particular output O (e.g., when an ad is observed in the context of a sufficient number of differentiated profiles), the DB triggers a notification to the Controller, which launches a data analytics job for that particular output O in an attempt to determine the inputs that it is targeting. The analytics job is picked up by an analytics worker, which leverages our *statistical correlation engine*, described at length in §**??**, to identify whether any subset of the inputs strongly correlate with the output, and if so which. In addition, the statistical methods also yield a *confidence score* that measures the statistical significance of the inferred correlation. All data needed to do the correlation is in the `Observations` and `Experiment` tables.

For example, using the information about the profiles in which ads were seen, statistical correlation may find that an ad O is often seen in profiles that include websites $I1$ and/or $I2$ in their histories, and never in profiles missing one or both of these websites. In such a situation, statistical correlation will conclude that O targets $\{I1, I2\}$ with high confidence (e.g., $.99$). This association, along with its confidence, will be added to the `Hypotheses` table in the DB. If later on, more observations of the ad are amassed through data collection, then the correlation job will run again, which may result in a higher confidence hypothesis. Whenever a new targeting hypothesis with some minimal confidence is added to th `Hypotheses` table, the Controller is notified and invokes an developer-provided callback, `onNewHypothesisNotification`, which determines the next steps to follow. This is where an developer can register any validation and/or attribution experiments in her workflow, which focuses on the newly discovered targeting hypothesis and either gathers more data to further increase the confidence or asks a different question (e.g., which specific tracker was responsible for targeting ads against webmd.com). To register a new experiment, the developer will use again the `registerExperiment` method in the Hubble API, and Hubble will launch that new experiment (or experiments if there are multiple) similarly to the starting experiment.

Hubble critically supports *reactive experiments* that are automatically triggered and conducted on-the-fly in response to targeting hypotheses generated by earlier experiments. These facilities enable developers to automatically gather additional evidence or study additional aspects of targeting that may not be available from an initial survey experiment.

For example, as in the design from Fig.**??**, we may follow an initial survey experiment with *attribution experiments* that considers different trackers that may be responsible for ad targeting. Specifically, for each plausible targeting hypothesis corresponding to an ad and a pair of input/output websites, we launch an reactive experiment where the inputs are the trackers appearing on either the input or output website, and the outputs are the ads on the output website (just as in the survey experiment). These experiments then yield targeting hypotheses about trackers (rather than just website), along with associated p-values.

Both initial survey experiments as well as reactive experiments may generate targeting hypotheses that are promising but not quite at the level of being declared statistically significant. Therefore, we may also use Hubble's reactive experiments facilities to conduct follow-up *validation experiments* for these hypotheses in the hopes of more definitive confirmation. These experiments may limit the inputs to a much smaller set (e.g., just the few input websites suggested as targeted for a given output site/ad), but use many more (soft) profiles to increase the statistical power to reject the null hypothesis (if it is indeed false) in the statistical analysis.

We note that all of these reactive experiments are automatically and promptly triggered in Hubble, which is critical because the dynamic nature of the web may alter or suppress evidence of targeting over time. Hubble therefore enables developers to efficiently collect the necessary evidence to support targeting hypotheses in a highly adaptive manner.

4.1.2 Statistical Correlation and Causal Inference

The proposed Hubble infrastructure requires mechanisms for both generating and validating plausible targeting hypotheses. The possible causes for ad targeting and tracking in a given system are myriad, and it is intractable—for both human users and computational methods—to exhaustively consider all of the possibilities. Therefore, it is critical to identify and develop methods that efficiently search for the most likely causes, properly evaluate these potential causes, and then succinctly report reliable results in an interpretable fashion. While there are many existing techniques designed specifically for finding causes of ad targeting in various settings (e.g., [22,41,64]), they are generally fragile, inflexible, and do not scale with the large numbers of potential targeting hypotheses (contrary to claims).

As part of Hubble, we will develop a rigorous and scalable statistical methodology for generating and testing targeting hypotheses based on Hubble's primitives for conducting randomized experiments based on synthetic user profiles, which permit strong causal findings of targeting. Such causal findings can then be used to inform users of the privacy implications of exposing sensitive information to online systems and trackers. We will also develop methods for testing hypotheses based on real user profiles from a trusted and secure peer-to-peer network. While such observational data cannot provide strong causal findings without further assumptions, they can still be informative for an end-user if presented with the proper context.

Basic approach to generating targeting hypotheses. We will first develop a method based on linear regression to discover putative targeting hypotheses from experimental data collected by Hubble. A linear regression model posits that a real-valued *output variable* y is determined by a linear combination of p *input variables* $\mathbf{x} := (x_1, x_2, \ldots, x_p)$, plus a random mean-zero noise ε. (Categorical variables are expanded using "dummy variables": a variable that takes r possible values expands to r mutually exclusive $\{0,1\}$-valued variables.) The linear model is written as $y = \sum_{i=1}^{p} w_i x_i + \varepsilon$, where $\mathbf{w} := (w_1, w_2, \ldots, w_p)$ is the *regression coefficient* vector. Given n vectors of input values $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$ together with their corresponding output values $y^{(1)}, \ldots, y^{(n)}$, the goal of linear regression is to estimate the regression coefficients $\mathbf{w}$.

In a basic Hubble application for a particular service (e.g., ads targeting), each vector of input values corresponds to a user profile. The variables in $\mathbf{x}$ correspond to either $\{0,1\}$-valued indicators for the possible targeting inputs—such as the websites that the user has visited that might be used to target ads—or uncontrolled variables associated with a user profile (e.g., time-of-day of experiment, IP address of client used). For each user profile created, Hubble randomly and independently assigns a value to each targeting input, and also records the value of the uncontrolled variables. The output variable y encodes a particular measured output of an online service or system: for instance, it may represent the number of times a particular ad was displayed to the user, or it may be some aggregate function of all ads displayed to the user. The regression coefficients $\mathbf{w}$ are used to screen the inputs and generating plausible targeting hypotheses based on the coefficients' magnitudes; the uncontrolled variables are accounted for in the regression and hence may help suppress correlations between the output and irrelevant targeting inputs that would otherwise arise. The targeting inputs associated with large regression coefficients are then be, in some combination, hypothesized to be responsible for differences in the observed output. Such a hypothesis will then be vetted using a valid statistical test in a subsequent stage (again, discussed later).

Using sparse linear regression. As described so far, this regression approach for generating targeting hypotheses is not scalable because there are likely many possible targeting inputs—and combinations thereof—that may *a priori* have causal effect on the output. Yet using standard linear regression approaches will require at least as many user profiles as there are possible inputs (i.e., $n \geq p$), regardless of how many of these inputs are actually responsible for the targeting output. Such experiments would be very costly and time-consuming, hence will severely limit the utility of Hubble-based applications.

We propose to use *sparse linear regression* methods, which are effective at estimating $\mathbf{w}$ even when $p \gg n$, as long as $\mathbf{w}$ is sparse—i.e., has only a few non-zero entries. This sparsity assumption entails that only a few input values are, in combination, correlated with the output. A well-established method for sparse linear regression is Lasso [61]. Under certain conditions on the n input vectors, which we ensure are likely to be satisfied *by construction* of our profiles, Lasso accurately estimates $\mathbf{w}$ as long as $n \geq O(k \log p)$, where k is the number of non-zero entries in $\mathbf{w}$ [14]—i.e., the number of input variables potentially correlated with the output. In fact, this collection of $O(k \log p)$ input vectors supports the *simultaneous* estimation of multiple coefficient vectors for different outputs (e.g., different ads); this remarkable phenomenon (related to compressed sensing [15, 23]) enables highly scalable experiments for generating targeting hypotheses.

Validating targeting hypotheses. To verify whether a targeting hypothesis is valid, we propose to use a two-stage protocol commonplace in statistics and machine learning. We employ two groups of user profiles: the first group ("training set") is used for generating plausible targeting hypotheses, and the second group ("test set") is used solely for testing the hypotheses. We will use tests that provide measures of statistical signifiance in the form of *p-values*. Each targeting hypothesis will be formalized as a function f of the targeting inputs associated with a user profile, and the function f is hypothesized to be correlated with or causally related to the associated output y. There are numerous statistical tests that may be appropriate for this validation task; we will asssess these tests based on the statistical assumptions they require for soundness (e.g., independence of the outputs associated with each user profile), and for the conditions under which they are able to positively validate hypotheses. We note that when the input values relevant to a hypothesis function f are randomly assigned to the user profiles in the test set, then we are able to assess the *causal effects* of these inputs on the output.

Complex targeting hypotheses. The sparse linear regression approach described above most naturally generates targeting hypotheses based on (thresholded) linear functions of the inputs. Such functions include as subclasses conjunctions and disjunctions of boolean input variables, which may be sufficient for a substantial number of cases. However, we anticipate that targeting hypotheses may admit additional structure that can be used to improve scalability. First, the targeting inputs may naturally partition into semantically meaningful groups (e.g., health websites, travel websites) that are targeted as a group rather than as individual inputs. If these groups were known, we could bet on group-level sparsity in the sparse regression approach to discovering targeting hypotheses, and require fewer user profiles to accurately estimate regression coefficients [39]. We may try to discover these groups by applying methods for clustering the inputs using correlation metrics, such as those employed in [8]. Secondly, we may also seek out higher-order combinations (e.g., conjunctions) of inputs that are potentially relevant, and include these combinations in the regression [6]. This would ultimately expand the class of targeting hypotheses that are considered (e.g., disjunctions of conjunctions). To support these forms of complex hypotheses, as well as others that we may happen upon, we propose to use a multi-stage approach whereby groups or higher-order inputs are constructed in a first stage, targeting hypotheses are generated in a second stage, and finally hypothesis testing is conducted in the final stage.

Targeting hypotheses from observational data. Thus far, we have discussed approaches to generating and validating causal targeting hypotheses. However, these methods are based on synthesizing user profiles that may be far removed from any given real user's profile. Therefore, we believe it will be beneficial to also consider targeting hypotheses based solely on actual users' profiles, as obtained from a trusted peer-to-peer network. Due to the lack of direct interventions and randomization, it is generally not possible to obtain strong causal findings from these data without strong modeling assumptions. Nevertheless, such users may be able to more closely related to these findings than the ones based on synthesized profiles. In the statistical parlance, these data are regarded as *observational data*, and there is a vast body of literature in statistics and economics on methods that attempt to make causal inferences (under various modeling assumptions) from these data (e.g., [48]). We will explore and evaluate techniques for estimating causal effects from observational based on an assumed casual model, as well as techniques for estimating this causal structure from observational data. It is not clear whether the necessary assumptions for these techniques will be met in a real application, so we will also pursue non-causal hypotheses (e.g., hypotheses of correlations or other measures of associations) that may simply be annotated with a familiar disclaimer ("correlation does not imply causation"). For such hypotheses, we will aim to also discover latent factors such as population stratification structure that could induce or mask correlations between targeting inputs and outputs. Accounting for these latent factors may produce more reliable findings, as well as increase the statistical power to discover targeting behavior that only manifest in subpopulations.

4.1.3 Privacy-Preserving Transparency Protocols

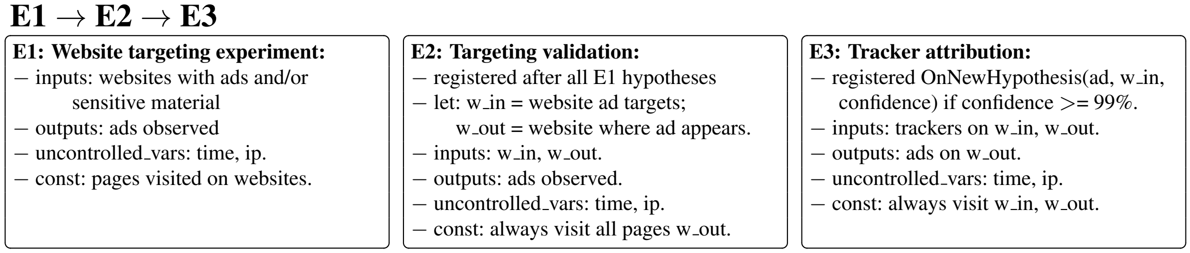
4.2 Thrust 2: Transparency and User Awareness Tools

[Need a segway here, b/c CollectionObservatory doesn't use Hubble.] xxx

A second major direction of our research will be to build a powerful new set of transparency and user awareness tools. Some of these tools will be directly built atop Hubble's abstractions for detecting data use for targeting, personalization, and discrimination; others inform users users about. The tools build upon Hubble, as well as upon each other, and we suspect will integrate nicely with TA1 and TA2 data protection technologies to audit the effectiveness of those technologies. We describe the proposed tools in turn.

4.2.1 AdObservatory: Revealing Targeting in Online Advertising

The first tool we propose to build atop Hubble is *AdObservatory*, which leverages Hubble's abstractions to reveal to the users how they are being targeted by online advertisers. For each ad that a user encounters while surfing the Web, AdObservatory, a browser plugin, will tag the ad with two pieces of information: (1) which specific website(s) in the user's browsing history that caused that ad to be shown and (2) which specific tracker(s) that witnessed the users' visits caused the ad to be shown. These pieces of information correspond directly to the goals we established for user awareness tools (see Section 2): (1) enables *targeting awareness* and (2) provides *attribution*.



E1 → E2 → E3

E1: Website targeting experiment:
– inputs: websites with ads and/or sensitive material
– outputs: ads observed
– uncontrolled_vars: time, ip.
– const: pages visited on websites.

E2: Targeting validation:
– registered after all E1 hypotheses
– let: w_in = website ad targets;
w_out = website where ad appears.
– inputs: w_in, w_out.
– outputs: ads observed.
– uncontrolled_vars: time, ip.
– const: always visit all pages w_out.

E3: Tracker attribution:
– registered OnNewHypothesis(ad, w_in, confidence) if confidence >= 99%.
– inputs: trackers on w_in, w_out.
– outputs: ads on w_out.
– uncontrolled_vars: time, ip.
– const: always visit w_in, w_out.

Table 1: **AdObservatory Experiment Design.** First conducts a large scale survey for cross website ads, then runs validation experiments to improve confidence on the targeted ads, and a tracker stage to attribute ads to specific trackers.

Fig.1 shows an experiment workflow that we might use for AdObservatory. It consists of three Hubble experiments: E_1 is a broad survey experiment to formulate rough targeting hypotheses for each ad; E_2 is a smaller experiment to validate the targeting hypotheses generated by the survey and prune out ad erroneously labeled as targeted; E_3 is an attribution experiment to determine which specific trackers contributed to the targeting of cross-domain ads discovered by E_1 and confirmed by E_2. In more detail, E_1 aims to identify what ads are targeting from a huge range of possibilities. In E_1 each input is one of hundreds or thousands of websites in a user's web history. The data collection is either an automated browser (e.g., using Selenium) in the case of controlled experiments using AdObservatory, or a plugin running in users' browser. Regardless, the data collection procedure records as Hubble outputs all display ads observed while visiting web pages. To detect ads on arbitrary pages, we will modify AdBlock to report any identified ad but does not disable it [5]. In addition to collecting display ads, the data collection also records all trackers detected on each site for use in future experiments in the workflow. To detect trackers, AdObservatory will leverage the CollectionObservatory tool we will develop as part of this project. AdObservatory will use Hubble's statistical correlation and causation methods (§4.1.2) to identify which output display ads target which input sites.

E_2 aims to validate targeting hypotheses from E_1 in a more rigorous and controlled fashion. E_2 creates a group of n profiles, half of which get assigned the targeted website. In addition, all of the profiles get assigned the websites on which the ad appeared; excluding the targeted website if the ad appeared there too. Since, our input is the presence of the targeted website in a profile, E_2 is restricted to ads that appear on at least one website other than the targeted. The data collection and analysis follow the same procedure as E_1. The ads and their respective groups of site validated as targeted in E_2 will be used in E_3.

E_3 is similar to the first two experiments and uses the same groups of sites as E_2 but uses trackers collected in E_1 as inputs rather than sites. Using the standard Hubble assignment mechanism each tracker is randomly assigned to half of the accounts. The data collection worker used in E_3 drives the profile to all sites in the assigned group blocking trackers all trackers no allocated to that profile. AdObservatory uses the Hubble's default statistical methods to determine which ads target which trackers.

4.2.2 DiscriminationObservatory: Revealing Online Price Discrimination

A second tool that we propose building is *DiscriminationObservatory*, a tool that leverages Hubble to reveal to the users how arbitrary websites are personalizing their content based on their personal data. A specific use case we are aiming to support is to reveal price or offer differentiation on eCommerce, mortgage, and loan websites based on users' personal information, such as Facebook or Google+ profile information, or web histories purchased from trackers. Many websites today leverage the single-signon capabilities of a handful of giant-scale services, such as Facebook Connect and Google OAuth, to authenticate their users. Upon authenticating a user through Facebook or Google, the websites obtain access to various aspects of a user's profile on these services, and the level of access depends upon the permission level that the website asks for. These pieces of information are often used by the websites to personalize content. For example, Pinterest leverages Facebook Connect to authenticate its users; it requests access to the friend list of a user (among other things) and uses that list to personalize the content it recommends its users. Many other websites do this, and there has been speculation in the media recently that mortgage, loan, and other eCommerce websites might soon start using Facebook Likes and other social information to present the users with differentiated quotes on their websites [2]. At present, no one knows whether any such websites apply such differential treatment, and (worse) no one can find out, because there are no scalable, robust, and generic tools that can identify this kind of behavior in the wild.

Our goal in DiscriminationObservatory is to *detect* such differential treatment on arbitrary websites and *surface* sufficient information to the end-users and privacy watchdogs. End users may leverage this information to inform their decisions about the offers they receive. Privacy watchdogs can use DiscriminationObservatory to search for websites on the web that discriminate against protected user categories, such as specific races, ethnic groups, or genders. To first order, DiscriminationObservatory will leverage Hubble to obtain and analyze the contents of websites of interest (the Document Object Model, or DOM) from the vantage points of users with differentiated profiles. It will compare the contents of the websites at DOM tree level to identify differences that are consistent with differences in the user profiles. Finally, it will highlight visually any DOM portions that receive differential treatment based on various aspects available in social profiles (e.g., gender, relationship, Likes, friend list, etc.).

A key challenge, and technical contribution, in DiscriminationObservatory will be to identify differences between the DOM versions of a particular page that are consistent with differences in user profiles. Comparing unordered tree-structured data is known to be an NP-Hard problem (MAX SNP-hard class to be precise) [67]. KF-Diff+ is an algorithm developed for comparing XML trees (ordered and unordered) in linear time, works around this problem by assigning unique IDs to nodes with the same label and father [65]. A similar ID-based work-around is leveraged by the Facebook React framework [1] to identify changes within a DOM after page refresh so as to perform rendering optimizations. Unfortunately, we find that web programmers often fail to assign unique IDs to their elements, although most web programming frameworks do so. Additionally, approximate algorithms exist, such as RWS-Diff, a log-linear algorithm that provides an approximate solution for differentiating ordered and unordered trees [29]. We will investigate and evaluate these algorithms' performance and accuracy on real DOM trees at large scale, and infuse them with robust heuristics from our application domain that will help guide them toward more efficient solutions. Our extremely preliminary algorithm currently identifies differentiation on several websites (including Pinterest, Indieogo, and Techcrunch) without requiring any customizations specific to these sites.

4.2.3 CollectionObservatory: Revealing Third-Party Content and Tracking

A third tool we propose to build is *CollectionObservatory*, a comprehensive data collection auditing tool that reveals who tracks. CollectionObservatory is valuable as a tool for the end users as well as to support XXX. CollectionObservatory will offer three functions for the first time:

1. *Comprehensive web tracking detection:* CollectionObservatory should automatically detect a wide range of web tracking behaviors, including not only well-understood trackers based on browser cookies but also stateless fingerprint-based trackers, which use browser and machine fingerprints to re-identify users, and more esoteric tracking mechanisms (e.g., cache-based, Flash cookies, etc. [40]).
2. *User-facing visualization and awareness:* CollectionObservatory should visualize third-party content and data collection for users in a way that is effective and actionable, XXXXXXXXXXXhelping users take control of the collection and use of their web browsing behaviors.
3. *Research platform:* CollectionObservatory is intended as a platform, allowing other researchers to adapt and build upon our tracking detection capabilities, visualization primitives, and user-facing actions to build additional web privacy user awareness tools.

In prior work we developed TrackingObserver [54,57], a more limited browser-based web tracking detection and measurement platform which detects only cookie-based tracking. CollectionObservatory will build on TrackingObserver but move significantly beyond it to (1) detect web tracking behaviors of much more diverse and subtle types and (2) provide useful user-facing visualizations of the observed behaviors.

Detecting fingerprint-based trackers. Our previous work, TrackingObserver [57], detects primarily cookie-based tracking that explicitly store state in the user's browser. In CollectionObservatory, we will extend the scope of this automatic detection to include additional tracking behaviors, primarily *fingerprint-based trackers*. Fingerprinting-based trackers re-identify users based on unique combinations of attributes such as IP address, user agent, installed fonts and plugins, etc [25]. While researchers have explored how fingerprinting works and conducted limited measurement studies of specific fingerprinting techniques or known fingerprinting libraries (e.g., [3,4,47,66]), there has been no extensive non-blacklist-based study of fingerprinting in the wild nor a user-facing tool to detect these behaviors. Implementing fingerprint-based tracking detection in CollectionObservatory, e.g., via hooks on the JavaScript APIs commonly used to generate fingerprints, would allow us to perform a similar study for these trackers. We will conduct a measurement study of tracking on a large number of popular and less popular websites, including from different vantage points (e.g., from different geographic locations). Ultimately, these findings will inform a user awareness tool for web tracking, described below.

Revealing third-party web content. A number of tools exist to reveal which third-party web trackers are loaded on a given web page, but (as described above) none of these tools localize those trackers on the page. That is, a user can learn that `doubleclick.net` was contacted as the page was loaded, but not which, if any, ads on the page were served by `doubleclick.net`. Similarly, a user cannot easily answer the question "where did this ad come from?" for a given ad, since even ads loaded from a particular domain may have been placed there by a different third-party (typically an advertising network) [54]. Indeed, some ads might not even have been intended by the web page developer, such as those injected by malicious browser extensions [7]. We propose a tool to identify third-party content on a page and attribute it to its source; achieving this requires addressing a number of technical challenges, including identifying content modifications on the first-party page that are the result of third-party scripts. We plan to integrate this tool with CollectionObservatory, and envision that it can be used to bootstrap both a user study of attitudes towards and expectations surrounding web tracking (see Section 4.3) as well as a measurement study of third-party content on the web.

Full-fledged web tracking transparency tool. Building on the above and on other aspects of the Hubble infrastructure, and informed by the user studies we describe in Section 4.3, we will ultimately extend CollectionObservatory into a full-fledged web tracking transparency tool for end users. In addition to providing useful visualizations to users about how their information is collected and used as they browse the web, this tool will provide useful, actionable, and verifiable changes that users can make to improve their privacy. We will release this final version of CollectionObservatory as open source, and we will deploy the tool publicly, ideally as part of an existing tool (e.g., as part of the Electronic Frontier Foundation's Privacy Badger tool [26]), as we have done with ShareMeNot [56]) in the past. This deployment will serve as a field study of the tool, which in turn will inform additional iteration on the tool itself.

4.2.4 LocationObservatory: Revealing Privacy Implications of Location Tracking

[augustin] xxx

4.3 Thrust 3: User Studies and Transparency Tool Measurements

[Augustin: Add IRB note.] xxx

To maximize the effectiveness of the transparency infrastructure and the user awareness tools that we build, it is critical that we understand users themselves. To this end, our proposed work will include user studies of two types: (1) user studies to help us understand *users' existing mental models and attitudes*, and (2) user studies to help us *evaluate the effects of our tools*. We will work with our institutions' human subject review boards to obtain IRB approval before conducting any studies involving human subjects.

User Studies for Existing User Mental Models and Attitudes. Our transparency and user awareness tools aim to close the gap between users' existing mental models and attitudes with respect to the privacy of their data and the reality of what today's applications and services collect and use. To achieve this, we must first understand what users already know or believe about the collection and use of their private data. Prior work has studied users' mental models and attitudes in contexts such as targeted advertising (e.g., [42, 44, 49, 62]); we propose to extend that work here, and to update the findings for current users and systems.

Example 1: Reactions to Ad Targeting. As one example, we detail a user study intended to help inform our transparency and user awareness tools for web tracking and targeted advertising. We ask: what are users' mental models about ad targeting? How will they react upon learning that a particular ad is targeted at them? To explore this question, we will design a study in which we post ads (e.g., on Facebook or via Google ads) targeted at specific—possibly sensitive—keywords. The content of our ads will inform the person viewing them about the targeting, e.g., by revealing the keyword that was used to target that particular ad. Clicking on the ad will direct the participant to a page with additional information about targeted advertising and about our study, including several survey questions to help us evaluate the participants' reactions to (1) learning about the targeting as well as to (2) the targeting itself. By understanding and comparing participants' reactions to

Component	Sub-tasks	Responsible PI(s)
Hubble infrastructure	1.1, 2.1, 3.1	Geambasu
Statistical correlation and causation	1.3, 1.4, 2.3, 2.4, 3.3, 3.4	Hsu
Privacy-preserving transparency	1.5, 2.5, 3.5	Chaintreau
CollectionObservatory	1.7, 2.7, 3.7	Roesner
AdObservatory	1.2, 2.2	Geambasu
DiscriminationObservatory	2.2, 3.2	Geambasu
LocationObservatory	1.6, 2.6, 3.6	Chaintreau
User studies	1.8, 2.8, 3.2, 3.6, 3.8	Roesner, Chaintreau, Geambasu
Integration, Evaluation on TA3	1.9, 2.9, 3.9	All PIs

Table 2: **Team member responsibilities (research areas and subtasks).**

different targeting keywords, our results can help motivate and inform our transparency tools, which may in turn motivate changes within targeting systems themselves. For example, if we find that users are comfortable with ads targeted at debt-related keywords but not cancer-related keywords, we might recommend that ad targeting companies stop targeting cancer, or to offer an opt-in to such "sensitive" topics. More broadly, studies such as this one will help us understand the notion of "sensitivity" — how much does it depend on the user, what kinds of things are uniformly "sensitive," etc.? These findings will ultimately inform our transparency and user awareness tools as well as others working in this space.

Example 2: Blah blah. **[Something from Augustin somewhere around here?]** xxx

User Studies to Evaluate our Tools. In addition to user studies aimed to teach us about users in general, we must also evaluate the effectiveness of our transparency and user awareness tools with real users. These studies will take different forms through the design of a tool, beginning with limited usability studies of preliminary designs, followed by more in-depth studies to evaluate the effectiveness of our tools to improve user comprehension and to positively affect user behaviors, culminating in full-fledged beta-tests with real user populations. For example, co-PI Roesner has previously released a user-facing anti-web tracking tool (originally called ShareMeNot [56]) as part of the Electronic Frontier Foundation's Privacy Badger tool [26]. We will use connections like these to iteratively beta-test our tools with large numbers of real users in real contexts.

Web Targeting and Discrimination Measurements. We will leverage AdObservatory and DiscriminationObservatory to run the first large-scale measurement studies of tracker-based ad targeting and online discrimination/personalization. We will .

5 Personnel and Management Plan

The team members are faculty at two institutions: Columbia University and University of Washington. Columbia University will be the Prime Contractor for the project, with University of Washington acting as a subcontractor; the formal agreements are already in place for this project. Roxana Geambasu will be the overall project PI, responsible for general technical direction, coordination and reporting (in addition to conducting a portion of the research). Each co-PI will be responsible for one or more component and associated sub-tasks, as identified in Table 2. Each faculty member will be responsible for supervising Ph.D. Graduate Research Assistants (GRAs). Each faculty member will dedicate a significant amount of their time to this project, as identified in Table 3.

The management structure is relatively flat, with Geambasu the lead PI and everyone else working with each other and under the general guidance of Geambasu. The PIs already have a history of collaboration with each other and are co-advising students. For example, Chaintreau and Geam-

Key Individual	2015	2016	2017	2018	2019
Geambasu	53 h	160 h	160 h	160 h	160 h
Chaintreau	53 h	160 h	160 h	160 h	160 h
Hsu	53 h	160 h	160 h	160 h	160 h
Roesner	53 h	160 h	160 h	160 h	160 h

Table 3: **Team member commitments.**

basu co-authored the XRay paper [**?**] and are co-advising a Ph.D. student, the paper's first author. Chaintreau, Geambasu, and Hsu have been working on follow-on technology and are now writing a joint paper for CCS'15 on a related topic. Geambasu and Roesner were colleagues at the University of Washington and share a Ph.D. advisor; they have already started a collaboration in the space of user awareness studies. The Columbia Co-PIs meet face-to-face almost on a daily basis. To facilitate collaboration with the UW Co-PI, we will have regular meetings over Skype or other technology. We will also organize two physical meetings per year, hosted on a rotating basis among the institutions and/or co-located with the program PI meetings. We will use a wiki and Github for coordination and record keeping. We will organize a website to make all of our findings publicly available. **All code resulting from this program will be released open-source.**

5.1 Personnel

The PIs span a broad range of expertise: *systems* (Geambasu), *security and human factors* (Roesner), *theory and social networks* (Chaintreau), and *machine learning and statistics* (Hsu). We will combine this broad expertise in a close collaboration to produce the first scalable infrastructure for transparency and the first valuable tools for end-user privacy awareness. Following are the biographies of each participant. Section 6 describes the team's relevant expertise and joint projects.

Roxana Geambasu Dr. Roxana Geambasu is an Assistant Professor of Computer Science at Columbia University. She has made research contributions in software systems across a broad range of areas, research revolves around broad systems topics, including operating systems, distributed systems, and security and privacy. One over-arching theme of her research relates to increasing privacy in today's data-driven world by developing transparency, fairness, and data management tools for both programmers and privacy watchdogs, as well as the end-users. A list of her publications is available at: `www.cs.columbia.edu/~roxana`. Prof. Geambasu is a member of the Information Science and Technology (ISAT) focus group, having been appointed in 2014 to serve for a period of three years. For her work in privacy, Prof. Geambasu received a Microsoft Research Faculty Fellowship, a "Brillint 10" Popular Science listing, an NSF CAREER award (all in 2014), an Honorable Mention for the inaugural Dennis M. Ritchie Doctoral Dissertation Award in 2013, a William Chan Dissertation Award in 2012, two best paper awards at top systems conferences (USENIX Security and EuroSys), and the first Google Ph.D. Fellowship in Cloud Computing. Geambasu's second-year Ph.D. student has won the 2015 Google Ph.D. Fellowship in Privacy for work under her supervision. Prof. Geambasu's research has been featured in multiple articles in New York Times, The Economist, NPR, and others.

Augustin Chaintreau

Daniel Hsu Dr. Daniel Hsu is an Assistant Professor of Computer Science at Columbia University, and is a member of the Data Science Institute, also at Columbia. His research interests are in algorithmic statistics, machine learning, and privacy-preserving data analysis. His work on interactive and unsupervised learning have yielded the first computationally efficient and statistically consistent algorithms for a number of core estimation and learning problems that were only previously tackled using heuristics or suboptimal methods. Much of his current research focuses on developing scalable and statistically sound learning algorithms for discovering hidden structure in massive data, as well as on the interaction between statistical inference and privacy. He was the organizer of several workshops and tutorials on algorithms for learning hidden variable models at premier machine learning venues (ICML, NIPS); he received a Yahoo Academic Career Enhancement Award in 2014 and the UC San Diego Departmental Dissertation Award in 2010.

Franziska Roesner Dr. Franziska Roesner is an Assistant Professor of Computer Science and Engineering at the University of Washington. She has made research contributions in computer security and privacy, spanning broadly from systems to human factors. Her work involves designing and building systems that address security and privacy challenges faced by end users of existing and emerging technologies. For example, she has made contributions in computer security and privacy in the contexts of third-party web tracking, permission granting in modern operating systems (such as smartphones), secure embedded user interfaces, and emerging augmented reality platforms. A list of her publications is available at: `http://www.franziroesner.com`. Her work on web privacy included the development of ShareMeNot, a defense for one type of web tracker, which was incorporated into the Electronic Frontier Foundation's Privacy Badger tool in 2014. For her work in security and privacy, Prof. Roesner received the William Chan Memorial Dissertation Award in 2014, the IEEE Symposium on Security and Privacy Best Practical Paper Award in 2012, a NSF Graduate Research Fellowship, and a Microsoft Research PhD Fellowship.

5.2 Integration and Evaluation

Although each component is led by a particular team member, the PIs will work together as part of a unified team and will integrate all of their components to produce one coherent system and a useful set of tools, but more important, a robust infrastructure that can be leveraged by other researchers and developers to reveal other aspects of personal data treatment on the web.

We will evaluate our prototypes.

6 Capabilities

Our proposed work will leverage expertise, techniques and tools that we developed in a number of past and concurrent projects. Some of these techniques are in the process of being patented; the US Government has unlimited use rights to these.

PI Geambasu has been working on increasing privacy and transparency in computer systems for multiple years. As part of a DARPA MRC project (MEERKATS), she has CleanOS, a mobile operating system designed with privacy and transparency in mind [**?**, 59]. Unlike existing mobile OSes, CleanOS manages users' data carefully so as to (1) minimize exposure of users' personal data at any time in anticipation of attack and (2) provides visibility post-attack into what specific data might have been compromised. Pebbles, CleanOS's follow-on [**?**], provides users and auditors with meaningful levels of abstraction at which to audit data compromises post-attack. CleanOS is now being considered for transition into production.

Geambasu and Chaintreau have recently developed *XRay*, a preliminary transparency infrastructure that reveals data targeting in Web services [**?**]. The system, which is our preliminary foray into the topic of Web transparency and our inspiration for Hubble, is the very first to accurately reverse targeting in multiple services, including Gmail, Amazon, Youtube, and Google Search. The system, however, is limited in scale, applicability, features, and our personal experience with it. We are planning to address these and more limitations in Hubble, and develop the very first scalable, extensible, and robust transparency infrastructure.

Geambasu, whose core expertise lies in building scalable, extensible, and robust distributed systems [**?**, **?**, 30], has a track record of transition into practice of the systems she builds. For example, Synapse [**?**], a scalable, heterogeneous-database replication system, has been deployed at Crowdtap, a data-driven marketing startup in NYC, which has been running it in production for about a year with great success. The system vastly improves the way that company manages their highly heterogeneous databases, and the engineers have claimed that it increases their company's agility to develop new features on top of diverse data. As another example, Geambasu deployed the first security measures in a commercial, giant-scale distributed hash table (DHT) with millions of users. Her defenses alleviated the potential for certain Sybil attacks on that DHT, which had been wide open to attack [**?**].

PI Chaintreau

PI Hsu works on algorithmic statistics, machine learning, and privacy-preserving data analysis. He has developed several foundational algorithms in the areas of active learning and unsupervised learning. For active learning, he has developed the first computationally efficient and noise-tolerant methods for active learning [13,21], as well as the first statistically-consistent active learning methods for exploiting population stratification structure [20]. These methods provide an algorithmic basis for adaptive experimental design in classification problems, which we hope to employ in Hubble for scalability. In the context of unsupervised learning, Hsu has developed the first computationally efficient algorithms for learning a host of latent variable models that had only previously been tackled using local search or sampling heuristics [9–11, 35, 37, 38]. Latent variable models are probabilistic models that capture hidden structure in data; this structure can be leveraged to improve the statistical power for finding significant correlations and causal relationships.

In addition to work on active and unsupervised learning, Hsu has studied implications of privacy constraints on statistical machine learning and data analysis, and he has experience in developing scalable learning algorithms for complex regression problems. His work has revealed practically-relevant limitations of requiring differential privacy guarantees on learning algorithms and statistical estimators [16, 17], and also developed new methods that exploit conditions that are favorable for learning when it is available [18]. He also developed highly scalable algorithms for discovering complex variable interactions that are useful in regression [6], as well as methods for exploiting sparse linear regression in the context of complex output prediction [36].

PI Roesner has worked on web privacy topics for several years, focusing on (1) studying and measuring the existing state of web privacy, (2) building tools to enable measurement and other follow-on work, and (3) providing users with visibility into and control over their privacy on the web. Her 2012 taxonomy and measurement study of third-party web tracking in the wild [54] was among the first efforts to deeply understand the web tracking space. As part of this work, Roesner developed *ShareMeNot*, a defense for social media web trackers (such as the Facebook "Like" button). ShareMeNot's techniques were adopted by Ghostery [31], a popular anti-tracking browser add-on, and ShareMeNot's code itself was incorporated into the Electronic Frontier Foundation's Privacy Badger [26] web privacy tool in 2014. Roesner's work has also focused on ensuring that the security and privacy properties of systems match users' expectations in other contexts. For example, she developed *user-driven access control* [53] as a new approach for permission granting in modern operating systems (such as smartphones), by which the operating system is able to extract a user's permission granting intent from the way he or she naturally interacts with any application. Roesner implemented user-driven access control in *LayerCake*, a modified version of Android that provides security for embedded user interfaces [50,51]. Her work has also focused on emerging security and privacy challenges in emerging augmented reality and continuous sensing platforms [52,55].

7 Statement of Work

Our effort is composed of one overall task, aimed at developing a complete and demonstrable Hubble prototype and tools. We define a number of subtasks that partition the effort into smaller, easily manageable components that can be separately developed and evaluated prior to integration.

TASK: **Objective**: Investigate, develop, and experimentally evaluate a Hubble prototype; develop and evaluate user awareness tools built upon its primitives.
General Description: This is our main goal and high-level task, around which a number of smaller tasks (broken down by phase) are organized. We will develop and integrate the individual components, and evaluate the integrated architecture across the full duration of the project.
Responsible Organization and Location: Columbia University (NYC), University of Washington (Seattle).
Exit Criteria: An extensible, scalable, and robust infrastructure system for building transparency tools to increase users' awareness of how their data is being collected, used, and exchanged by online services. A greatly improved understanding of how such tools can help change user perceptions of the risks involved and improve their mental models of protection techniques that exist or are being developed as part of the Brandeis program. Evaluation in terms of accuracy, scale, performance, and lightweightness are successful on real systems as well as systems developed by TA3 researchers.
Deliverables: Prototype implementation of Hubble and tools, including documentation and the final project report, quarterly technical progress reports, slide presentations, evaluation data, and other reports per requirements. All source code for Hubble and tools will be released publicly on Github.

Our goal is to develop this task in the course of the program, with core milestones that match the program's phases. We start from a basic infrastructure system and user awareness tools (Phase 1), after which we enhance to add more abstractions and features (Phase 2), and we finalize our technologies and integrate them with each other and with other Ta1 and TA2 technologies as appropriate (Phase 3). For each phase, we dedicate one subtask to integration aned evaluation with systems developed by TA3 performers. Specific subtasks per phase follow.

7.1 Phase 1 (Months 1-18)

TASK 1.1: **Objective**: Design and implement basic Hubble infrastructure and tool development API.
General Description: Design an early version of Hubble's architecture and developer APIs. The architecture will support single-stage experiments (no validations or refinements). Implement this early architecture, use the most basic statistical correlation engine available, and stub any other components yet unavailable (e.g., privacy-preserving protocol, causal inference, etc.). Focus on controlled-input use cases.
Responsible Organization and Location: Columbia University (NYC)
Exit Criteria: Infrastructure that reveals input/output targeting by measuring correlation on differentiated profiles. Supports 10s-100 inputs and has precision/recall for detecting targeting of 70-90%.
Deliverables: Early software prototype and design documents.

TASK 1.2: **Objective**: Design and implement basic AdObserver tool.
General Description: Implement a basic version of the AdObserver tool to exercise Hubble's architecture and APIs. Use AdBlocker to identify ads on arbitrary pages.
Responsible Organization and Location: Columbia University (NYC)
Exit Criteria: Tool that can reveal ads targeted on previously visited websites or other data.
Deliverables: Software prototype and design documents.

TASK 1.3: **Objective**: Develop basic statistical methodology for testing targeting hypotheses.

General Description: Developing a formal specification for targeting hypotheses as generated by Hubble, together with a methodology for reliable testing of the hypotheses.

Responsible Organization and Location: Columbia University (NYC)

Exit Criteria: Concrete specification of targeting hypotheses, and software tool that computes valid statistical tests at any specified level, incorporated into Hubble.

Deliverables: Software prototype and design documents.

TASK 1.4: **Objective**: Apply scalable sparse linear regression methods to generation of targeting hypotheses.

General Description: Develop techniques based on sparse linear regression to infer putative targeting hypotheses from data collected by Hubble. Evaluate scalability using simulated targeting mechanisms as well real data collected by Hubble.

Responsible Organization and Location: Columbia University (NYC)

Exit Criteria: Scalable and empirically-validated implementation of sparse regression methods, incorporated into Hubble pipeline.

Deliverables: Software prototype and design documents.

TASK 1.5: **Objective**: Design and implement basic privacy-preserving transparency protocol.

General Description: XXX. Relies on central collection service.

Responsible Organization and Location: Columbia University (NYC)

Exit Criteria: XXX.

Deliverables: XXX.

TASK 1.6: **Objective**: Design and implement basic LocationObserver to reveal information that can be inferred from location.

General Description: XXX describe this task. Evaluate privacy-preserving protocol against alternative designs.

Responsible Organization and Location: Columbia University (NYC)

Exit Criteria: XXX.

Deliverables: XXX.

TASK 1.7: **Objective**: Implement fingerprint tracking detection infrastructure in CollectionObservatory.

General Description: We will build on TrackingObserver, our prior web tracking detection and measurement platform, to begin developing CollectionObservatory. First, we will implement detection of fingerprint-based web trackers that use browser and machine fingerprinting techniques to re-identify users. Rather than using a known list of fingerprinting scripts, we will detect fingerprinting behavior using a measurement of entropy extracted by a potential tracker's JavaScript API accesses.

Responsible Organization and Location: University of Washington (Seattle, WA)

Exit Criteria: Initial version of CollectionObservatory that successfully detects a large fraction of fingerprint-based trackers, evaluated by a comparison with blacklist-based tracking detection tools.

Deliverables: Initial version of CollectionObservatory that detects fingerprint-based trackers.

TASK 1.8: **Objective**: Conduct user study of attitudes towards targeting.

General Description: We will conduct a user study to better understand users' attitudes towards targeted advertising. We will target ads using a variety of keywords (including sensitive keywords) and inform users about the targeting in the content of the ads. For participants who click on the ad, we will debrief them about the study and ask addition survey questions.

Responsible Organization and Location: University of Washington (Seattle, WA)

Exit Criteria: Sufficient participation in the user study to draw statistically significant conclusions.

Deliverables: Conclusions drawn from user study results.

TASK 1.9: **Objective**: Demonstrate our TA2 technology on a TA3 Research System.

General Description: Integrate basic implementation of Hubble and transparency tools with the Research System(s) implemented by TA3 researchers. Our tools should be able to determine data uses or other privacy attacks built within those systems.

Responsible Organization and Location: Columbia University (New York), University of Washington (Seattle)

Exit Criteria: Successful detection of data use in TA3 Research System.

Deliverables: Software prototypes, design documents, and results from evaluation.

7.2 Phase 2 (Months 19-36)

***TASK 2.1:* Objective**: Extend Hubble and APIs for multi-stage transparency tool designs.

General Description: Incorporate support for multi-stage transparency tools. Support validation and refinement as abstractions for multi-stage tools. Develop API for such tools. Incorporate causal inference building block into into Hubble as part of the validation phase. Continue to focus on controlled-input use cases.

Responsible Organization and Location: Columbia University (NYC)

Exit Criteria: A system capable of validating and explaining its own targeting assessments. Where possible, the system will make causal inferences. Its evaluated scale will be in the range of 100s-1000s inputs, but we expect its recall/precision to grow thanks to validations.

Deliverables: Software prototype and design documents.

***TASK 2.2:* Objective**: Extend AdObserver and DiscriminationObserver tools to leverage Hubble's multi-stage architecture.

General Description: Design and implement using Hubble's APIs validation and refinement stages for each tool. Run experiments to test and evaluate.

Responsible Organization and Location: Columbia University (NYC)

Exit Criteria: Tools that both scale and validate/explain their own assessments to the users.

Deliverables: Software prototype and design documents.

***TASK 2.3:* Objective**: Develop and evaluate methodology for generating and testing targeting hypotheses from observational data.

General Description: Explore and evaluate techniques for estimating causal effects from observational based on an assumed casual model, as well as techniques for estimating this causal structure from observational data. Also develop and evaluate correlation hypotheses that do not assert causal implications.

Responsible Organization and Location: Columbia University (NYC)

Exit Criteria: Software tool for hypothesis generation and testing using observational data.

Deliverables: Software prototype and design documents.

***TASK 2.4:* Objective**: Extend sparse linear regression methodology to support complex targeting hypotheses.

General Description: Develop two-phase methodology to support testing of complex targeting hypotheses: in the first phase, we generate putative input combinations based on correlations in an initial set of data; in the second phase, we test introduce new input combinations from the first phase. Evaluate this strategy using simulated data and real data collected by Hubble.

Responsible Organization and Location: Columbia University (NYC)

Exit Criteria: Scalable and empirically-validated implementation linear regression approach using higher-order inputs, incorporated into Hubble pipeline.

Deliverables: Software prototype and design documents.

***TASK 2.5:* Objective**: Extend privacy-preserving transparency to avoid trust in a central point.

General Description: XXX.

Responsible Organization and Location: Columbia University (NYC)

Exit Criteria: XXX.

Deliverables: XXX.

***TASK 2.6:* Objective**: Extend LocationObserver to integrate privacy-preserving techniques.

General Description: XXX.

Responsible Organization and Location: Columbia University (NYC)

Exit Criteria: XXX.

Deliverables: XXX.

TASK 2.7: **Objective**: Measurement study with CollectionObservatory.

General Description: With the fingerprint-based tracking detection implemented in CollectionObservatory (in addition to existing tracking detection capabilities in TrackingObserver from prior work), we will conduct a large-scale measurement study of tracking on the web.

Responsible Organization and Location: University of Washington (Seattle, WA)

Exit Criteria: Conduct a measurement study of tracking on a large number of popular and less popular websites, including from different vantage points (e.g., from different geographic locations).

Deliverables: Measurement study results, including the prevalence and effectiveness of fingerprint-based trackers, a comparison with previous measurement results, etc.

TASK 2.8: **Objective**: Small-scope user awareness tool that visualizes third-party content.

General Description: We will develop an initial user awareness tool for web tracking that identifies third-party content on a webpage and visualizes it for the user. This tool, combined with CollectionObservatory, will serve as a building block for our later, more full-fledged web tracking user awareness tool.

Responsible Organization and Location: University of Washington (Seattle, WA)

Exit Criteria: Software prototype that identifies and visualized third-party content on a webpage.

Deliverables: Software prototype and design documents.

TASK 2.9: **Objective**: Demonstrate our enhanced TA2 technology on a TA3 Research System. Initial trial of demonstration on a TA3 Existing System.

General Description: Integrate enhanced implementation of Hubble and transparency tools with the Research System(s) implemented by TA3 researchers. Begin integration of our tools with Ta3 Existing System(s), as well as TA1 and TA2 protection-oriented technologies to enable auditing of the effectiveness of their protection.

Responsible Organization and Location: Columbia University (New York), University of Washington (Seattle)

Exit Criteria: Successful detection of data use in TA3 Research System. Our tools should detect data uses in TA3 Research systems.

Deliverables: Software prototypes, design documents, and results from evaluation.

7.3 Phase 3 (Months 37-54)

TASK 3.1: **Objective**: Extend Hubble to support collaborative transparency scenarios.

General Description: Incorporate statistical correlation building block for uncontrolled inputs to support end-user scenarios. Also incorporate privacy-preserving building block to limit the need for users to trust Hubble. Run experiments with simulated users to evaluate.

Responsible Organization and Location: Columbia University (NYC)

Exit Criteria: A privacy-preserving collaborative transparency system where users can submit their inputs/outputs partially and retrieve targeting assessments.

Deliverables: Software prototype and design documents.

TASK 3.2: **Objective**: Extend AdObserver, DiscriminationObserver to the collaborative use case.

General Description: Port the tools to the collaborative version of Hubble and re-run measurements in a simulated collaborative scenario for evaluation.

Responsible Organization and Location: Columbia University (NYC)

Exit Criteria: Transparency tools that can be run collaboratively by the end users without the need to trust a third-party.

Deliverables: Software prototype and design documents.

TASK 3.3: **Objective**: Develop and evaluate statistical testing methodology for stratification structure.

General Description: Develop methods for discovering latent population stratification (clustering), together with hypothesis tests that leverage this stratification structure to increase the statistical power to detect targeting.

Responsible Organization and Location: Columbia University (NYC)

Exit Criteria: Software tool for computation of statistical tests.

Deliverables: Software prototype and design documents.

TASK 3.4: **Objective**: Extend sparse linear regression techniques to use adaptive multi-stage experimental designs, and incorporate statistical testing methods to generate higher-order targeting hypotheses.

General Description: Develop multi-stage methodology for exploiting groups of related targeting inputs and outputs. The group structures are inferred in a first experimental stage, and the subsequently exploited in a second stage using group-sparse linear regression methods to discover group-level targeting hypotheses.

Responsible Organization and Location: Columbia University (NYC)

Exit Criteria: Scalable and empirically-validated implementation of multi-stage group-sparse linear regression approach, incorporated into Hubble pipeline.

Deliverables: Software prototype and design documents.

TASK 3.5: **Objective**: Finalize privacy-preserving, collaborative transparency building blocks and integrate into Hubble.

General Description: XXX.

Responsible Organization and Location: Columbia University (NYC)

Exit Criteria: XXX.

Deliverables: XXX.

TASK 3.6: **Objective**: Finalize LocationObserver tool and run studies of impact of transparency on user actions.

General Description: XXX.

Responsible Organization and Location: Columbia University (NYC)

Exit Criteria: XXX.

Deliverables: XXX.

TASK 3.7: **Objective**: User study of third-party content visualization tool.

General Description: We will conduct a usability study of the previously developed third-party content visualization tool, to understand whether and how the tool is effective with real users: does it effectively convey information to users? Do users take useful actions in response to this information? etc.

Responsible Organization and Location: University of Washington (Seattle, WA)

Exit Criteria: Conduct user study with a sufficient number of participants to generate statistically significant results and inform interactive improvements to the tool.

Deliverables: User study results and iterative improvements to the software prototype.

TASK 3.8: **Objective**: Larger-scope web privacy user awareness tool.

General Description: We will developer a more full-fledge web tracking user awareness tool, integrating functionality from the previously developed third-party content visualization tool and from CollectionObservatory. This tool will be informed by the findings of user studies and measurements and will build on other infrastructure developed in the project.

Responsible Organization and Location: University of Washington (Seattle, WA)

Exit Criteria: Develop a more full-fledged web tracking user awareness tool informed by and building on other aspects of the project.

Deliverables: Software prototype and design documents.

TASK 3.9: **Objective**: Demonstrate our final TA2 technology on TA3 Research and Existing Systems.

General Description: Integrate final implementation of Hubble and transparency tools with the Research and Existing Systems implemented by TA3 researchers. Finalize integration of our tools with some TA1 and TA2 protection-oriented technologies to enable auditing of their effectiveness.

Responsible Organization and Location: Columbia University (New York), University of Washington (Seattle)

Exit Criteria: Successful detection of data use in TA3 Research and Existing Systems. Successful auditing of effectiveness of other TA1, TA2 technologies with which we integrate, as applied to the same TA3 systems.

Deliverables: Software prototypes, design documents, and results from evaluation.

8 Schedule and Milestones

The Gantt chart below provides a graphic representation of the project schedule at the level of sub-tasks, all of which fall with the one overall task of Hubble, aimed at developing a complete and demonstrable Hubble prototype and tools. The performing organization is indicated via color: blue tasks correspond to Columbia University, green tasks correspond to University of Washington.

Task	**Period**	**PI(s)**
Phase 1		
Task 1.1	Months 1-18	Geambasu
Task 1.2	Months 1-18	Geambasu
Task 1.3	Months 1-18	Hsu
Task 1.4	Months 1-18	Hsu
Task 1.5	Months 1-18	Chaintreau
Task 1.6	Months 1-18	Chaintreau
Task 1.7	Months 1-18	Roesner
Task 1.8	Months 1-18	Roesner
Task 1.9	Months 15-18	All
Phase 2		
Task 2.1	Months 19-36	Geambasu
Task 2.2	Months 19-36	Geambasu
Task 2.3	Months 19-36	Hsu
Task 2.4	Months 19-36	Hsu
Task 2.5	Months 19-36	Chaintreau
Task 2.6	Months 19-36	Chaintreau
Task 2.7	Months 19-36	Roesner
Task 2.8	Months 19-36	Roesner
Task 2.9	Months 33-36	All
Phase 3		
Task 3.1	Months 37-54	Geambasu
Task 3.2	Months 37-54	Geambasu
Task 3.3	Months 37-54	Hsu
Task 3.4	Months 37-54	Hsu
Task 3.5	Months 37-54	Chaintreau
Task 3.6	Months 37-54	Chaintreau
Task 3.7	Months 37-54	Roesner
Task 3.8	Months 37-54	Roesner
Task 3.9	Months 51-54	All

Table 4: **Timeline.** [XXX Guys: if you have good drawing tools, please help me transform this table into a pretty gantt chart (see instructions in text).]

Program milestones are indicated via bullets, and the duration of each sub-task is provided in the final column of the graphic.

[Someone, please can you generate this gantt chart? I don't know how to make it nice, I only use OpenOffice and it's very primitive. Look at the MEERKATS proposal I sent for guidance. Table 4 contains the timeline data for us.] XXX

9 Cost Summary

Entire Performance Period (Total: $3,960,419)

	Columbia U (prime)	**U Washington (sub)**	**Category Total**
Direct Labor	1,270,820	402,258	1,673,078
Materials ODC	783,786	324,082	1,107,868
Indirect Costs	910,322	269,151	1,179,473
Member Totals	2,964,928	995,491	3,960,419

GFY 15 (Total: $104,438)

	Columbia U (prime)	**U Washington (sub)**	**Category Total**
Direct Labor	23,921	7,844	31,765
Materials	1,517	1,196	2,713
ODC	18,594	12,725	31,319
Indirect Costs	32,962	5,679	38,641
Member Totals	76,994	27,444	104,438

GFY 16 (Total: $885,835)

	Columbia U (prime)	**U Washington (sub)**	**Category Total**
Direct Labor	297,808	94,277	382,085
Materials	39,200	12,855	52,055
ODC	142,269	42,071	184,340
Indirect Costs	205,205	62,148	267,353
Member Totals	674,481	211,351	885,832

GFY 17 (Total: $923,294)

	Columbia U (prime)	**U Washington (sub)**	**Category Total**
Direct Labor	296,940	96,167	393,107
Materials	39,200	12,930	52,130
ODC	146,087	58,067	204,154
Indirect Costs	210,684	63,219	273,903
Member Totals	692,911	230,383	923,294

GFY 18 (Total: $949,555)

	Columbia U (prime)	**U Washington (sub)**	**Category Total**
Direct Labor	306,367	97,493	403,860
Materials	39,200	12,984	52,184
ODC	150,019	63,182	213,201
Indirect Costs	216,340	63,970	280,310
Member Totals	711,926	237,629	949,555

GFY 19 (Total: $898,226)

	Columbia U (prime)	**U Washington (sub)**	**Category Total**
Direct Labor	294,556	86,783	381,339
Materials	31,683	12,209	43,892
ODC	141,732	68,808	210,540
Indirect Costs	204,745	57,711	262,455
Member Totals	672,716	225,510	898,226

GFY 20 (Total: $199,074)

	Columbia U (prime)	**U Washington (sub)**	**Category Total**
Direct Labor	61,228	19,694	80,922
Materials	0	4,926	4,926
ODC	34,285	22,130	56,415
Indirect Costs	40,387	16,424	56,811
Member Totals	135,900	63,174	199,074

References

[1] Facebook react. http://facebook.github.io/react/docs/reconciliation.html.

[2] Time - lendup.com. http://business.time.com/2012/11/16/can-a-payday-lending-start-up-use-facebook-to-create-a-modern-community-bank/.

[3] G. Acar, C. Eubank, S. Englehardt, M. Juarez, A. Narayanan, and C. Diaz. The web never forgets: Persistent tracking mechanisms in the wild. In *21st ACM Conference on Computer and Communications Security*, 2014.

[4] G. Acar, M. Juarez, N. Nikiforakis, C. Diaz, S. Gürses, F. Piessens, and B. Preneel. FPDetective: Dusting the web for fingerprinters. In *20th ACM Conference on Computer and Communications Security*. ACM, 2013.

[5] Adblock plus. https://adblockplus.org.

[6] A. Agarwal, A. Beygelzimer, D. Hsu, J. Langford, and M. Telgarsky. Scalable nonlinear learning with adaptive polynomial expansions. In *Advances in Neural Information Processing Systems 27*, 2014.

[7] R. Amadeo. Adware vendors buy chrome extensions to send ad- and malware-filled updates. Ars Technica. http://arstechnica.com/security/2014/01/malware-vendors-buy-chrome-extensions-to-send-adware-filled-updates/.

[8] A. Anandkumar, K. Chaudhuri, D. Hsu, S. M. K. akade, L. Song, and T. Zhang. Spectral methods for learning multivariate latent tree structure. In *Advances in Neural Information Processing Systems 24*, 2011.

[9] A. Anandkumar, D. P. Foster, D. Hsu, S. M. Kakade, and Y.-K. Liu. A spectral algorithm for latent Dirichlet allocation. *Algorithmica*, 72(1):193–214, 2015.

[10] A. Anandkumar, R. Ge, D. Hsu, and S. M. Kakade. A tensor approach to learning mixed membership community models. *Journal of Machine Learning Research*, 15(Jun):2239–2312, 2014.

[11] A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade, and M. Telgarsky. Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, 15(Aug):2773–2831, 2014.

[12] P. Barford, I. Canadi, D. Krushevskaja, Q. Ma, and S. Muthukrishnan. Adscape: Harvesting and Analyzing Online Display Ads. *WWW '14: Proceedings of the 23nd international conference on World Wide Web*, Apr. 2014.

[13] A. Beygelzimer, D. Hsu, J. Langford, and T. Zhang. Agnostic active learning without constraints. In *Advances in Neural Information Processing Systems 23*, 2010.

[14] P. J. Bickel, Y. Ritov, and A. B. Tsybakov. Simultaneous analysis of lasso and dantzig selector. *Ann. Statist.*, 37(4):1705–1732, 08 2009.

[15] E. J. Candès, J. K. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[16] K. Chaudhuri and D. Hsu. Sample complexity bounds for differentially private learning. In *Twenty-Fourth Annual Conference on Learning Theory*, 2011.

[17] K. Chaudhuri and D. Hsu. Convergence rates for differentially private statistical estimation. In *Twenty-Ninth International Conference on Machine Learning*, 2012.

[18] K. Chaudhuri, D. Hsu, and S. Song. The large margin mechanism for differentially private maximization. In *Advances in Neural Information Processing Systems 27*, 2014.

[19] W. Cheng, Q. Zhao, B. Yu, and S. Hiroshige. Tainttrace: Efficient flow tracing with dynamic binary rewriting. In *Proceedings of the 11th IEEE Symposium on Computers and Communications*, 2006.

[20] S. Dasgupta and D. Hsu. Hierarchical sampling for active learning. In *Twenty-Fifth International Conference on Machine Learning*, 2008.

[21] S. Dasgupta, D. Hsu, and C. Monteleoni. A general agnostic active learning algorithm. In *Advances in Neural Information Processing Systems 20*, 2007.

[22] A. Datta, M. C. Tschantz, and A. Datta. Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination. *arXiv.org*, Aug. 2014.

[23] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[24] E. Dwoskin. Why You Can't Trust You're Getting the Best Deal Online. *online.wsj.com*, Oct. 2014.

[25] P. Eckersley. How unique is your web browser? In *Proceedings of the International Conference on Privacy Enhancing Technologies*, 2010.

[26] Electronic Frontier Foundation. Privacy Badger, July 2014. `https://www.eff.org/privacybadger`.

[27] W. Enck, P. Gilbert, B. gon Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth. TaintDroid: An information-flow tracking system for realtime privacy monitoring on smartphones. In *Proc. of the USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2010.

[28] S. Englehardt, C. Eubank, P. Zimmerman, D. Reisman, and A. Narayanan. Web Privacy Measurement: Scientific principles, engineering platform, and new results. *Princeton University*, June 2014.

[29] J. P. Finis, M. Raiber, N. Augsten, R. Brunel, A. Kemper, and F. Färber. Rws-diff: Flexible and efficient change detection in hierarchical data. In *Proceedings of the 22Nd ACM International Conference on Conference on Information &#38; Knowledge Management*, CIKM '13, pages 339–348, New York, NY, USA, 2013. ACM.

[30] R. Geambasu, A. Levy, T. Kohno, A. Krishnamurthy, and H. M. Levy. Comet: An active distributed key/value store. In *Proc. of the USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2010.

[31] Ghostery Enterprise. Ghostery. `https://www.ghostery.com/`.

[32] D. B. Giffin, A. Levy, D. Stefan, D. Terei, D. Mazières, J. C. Mitchell, and A. Russo. Hails: protecting data privacy in untrusted web applications. In *OSDI'12: Proceedings of the 10th USENIX conference on Operating Systems Design and Implementation*. USENIX Association, Oct. 2012.

[33] A. Hannak, P. Sapiezynski, A. M. Kakhki, B. Krishnamurthy, D. Lazer, A. Mislove, and C. Wilson. Measuring personalization of web search. In *WWW '13: Proceedings of the 22nd international conference on World Wide Web*. International World Wide Web Conferences Steering Committee, May 2013.

[34] A. Hannak, G. Soeller, D. Lazer, A. Mislove, and C. Wilson. Measuring Price Discrimination and Steering on E-commerce Web Sites. *IMC '14: Proceedings of the 14th ACM SIGCOMM conference on Internet measurement*, 2014.

[35] D. Hsu and S. M. Kakade. Learning mixtures of spherical Gaussians: moment methods and spectral decompositions. In *Fourth Innovations in Theoretical Computer Science*, 2013.

[36] D. Hsu, S. M. Kakade, J. Langford, and T. Zhang. Multi-label prediction via compressed sensing. In *Advances in Neural Information Processing Systems 22*, 2009.

[37] D. Hsu, S. M. Kakade, and P. Liang. Identifiability and unmixing of latent parse trees. In *Advances in Neural Information Processing Systems 25*, 2012.

[38] D. Hsu, S. M. Kakade, and T. Zhang. A spectral algorithm for learning hidden Markov models. *Journal of Computer and System Sciences*, 78(5):1460–1480, 2012.

[39] J. Huang and T. Zhang. The benefit of group sparsity. *Annals of Statistics*, 38:1978–2004, 2010.

[40] S. Kamkar. Evercookie—virtually irrevocable persistent cookies. `http://samy.pl/evercookie/`.

[41] M. Lecuyer, G. Ducoffe, F. Lan, A. Papancea, T. Petsios, R. Spahn, A. Chaintreau, and R. Geambasu. XRay: Enhancing the Web's Transparency with Differential Correlation . In *23rd USENIX Security Symposium (USENIX Security 14)*, San Diego, CA, 2014. USENIX Association.

[42] P. G. Leon, B. Ur, Y. Wang, M. Sleeper, R. Balebako, R. Shay, L. Bauer, M. Christodorescu, and L. F. Cranor. What Matters to Users? Factors that Affect Users' Willingness to Share Information with Online Advertisers. In *Symposium on Usable Privacy and Security*, 2013.

[43] B. Liu, A. Sheth, U. Weinsberg, J. Chandrashekar, and R. Govindan. AdReveal: improving transparency into online targeted advertising. In *HotNets-XII: Proceedings of the Twelfth ACM Workshop on Hot Topics in Networks*. ACM Request Permissions, Nov. 2013.

[44] A. M. McDonald and L. F. Cranor. Americans' Attitudes about Internet Behavioral Advertising Practices. In *Proceedings of the Workshop on Privacy in the Electronic Society*, 2010.

[45] J. Mikians, L. Gyarmati, V. Erramilli, and N. Laoutaris. Detecting price and search discrimination on the internet. In *HotNets-XI: Proceedings of the 11th ACM Workshop on Hot Topics in Networks*. ACM Request Permissions, Oct. 2012.

[46] Mozilla. Lightbeam. `https://www.mozilla.org/en-US/lightbeam/about/`.

[47] N. Nikiforakis, A. Kapravelos, W. Joosen, C. Kruegel, F. Piessens, and G. Vigna. Cookieless Monster: Exploring the Ecosystem of Web-based Device Fingerprinting. In *Proceedings of the IEEE Symposium on Security and Privacy*, 2013.

[48] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2009.

[49] E. Rader. Awareness of behavioral tracking and information privacy concern in facebook and google. In *Symposium on Usable Privacy and Security*, 2014.

[50] F. Roesner, J. Fogarty, and T. Kohno. User Interface Toolkit Mechanisms for Securing Interface Elements. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, 2012.

[51] F. Roesner and T. Kohno. Securing Embedded User Interfaces: Android and Beyond. In *Proceedings of the USENIX Security Symposium*, 2013.

[52] F. Roesner, T. Kohno, and D. Molnar. Security and Privacy for Augmented Reality Systems. *Communications of the ACM*, 57:88–96, 2014.

[53] F. Roesner, T. Kohno, A. Moshchuk, B. Parno, H. J. Wang, and C. Cowan. User-Driven Access Control: Rethinking Permission Granting in Modern Operating Systems. In *Proceedings of the IEEE Symposium on Security and Privacy*, 2012.

[54] F. Roesner, T. Kohno, and D. Wetherall. Detecting and defending against third-party tracking on the web. In *Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, 2012.

[55] F. Roesner, D. Molnar, A. Moshchuk, T. Kohno, and H. J. Wang. World-Driven Access Control for Continuous Sensing Applications. In *ACM Conference on Computer and Communications Security*, 2014.

[56] F. Roesner, C. Rovillos, T. Kohno, and D. Wetherall. ShareMeNot: Balancing Privacy and Functionality of Third-Party Social Widgets. *USENIX ;login:*, 37, 2012. `https://sharemenot.cs.washington.edu/`.

[57] F. Roesner, C. Rovillos, A. Saxena, and T. Kohno. Trackingobserver: A browser-based web tracking detection platform, 2013. `https://trackingobserver.cs.washington.edu/`.

[58] L. Sweeney. Discrimination in online ad delivery. *Communications of the ACM*, 56(5), May 2013.

[59] Y. Tang, P. Ames, S. Bhamidipati, A. Bijlani, R. Geambasu, and N. Sarda. CleanOS: Mobile OS abstractions for managing sensitive data. In *Proc. of the USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2012.

[60] The Wall Street Journal. What they know, 2010–2012. `http://www.wsj.com/public/page/what-they-know-digital-privacy.html`.

[61] R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288, 1994.

[62] B. Ur, P. G. Leon, L. F. Cranor, R. Shay, and Y. Wang. Smart, useful, scary, creepy: perceptions of online behavioral advertising. In *8th Symposium on Usable Privacy and Security*, 2012.

[63] X. Xing, W. Meng, D. Doozan, N. Feamster, and W. Lee. Exposing Inconsistent Web Search Results with Bobble. *cseweb.ucsd.edu*.

[64] X. Xing, W. Meng, D. Doozan, N. Feamster, W. Lee, and A. C. Snoeren. Exposing Inconsistent Web Search Results with Bobble. *Passive and Active Measurements Conference*, 2014.

[65] H. Xu, Q. Wu, H. Wang, G. Yang, and Y. Jia. Kf-diff+: Highly efficient change detection algorithm for xml documents. In R. Meersman and Z. Tari, editors, *On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE*, volume 2519 of *Lecture Notes in Computer Science*, pages 1273–1286. Springer Berlin Heidelberg, 2002.

[66] T.-F. Yen, Y. Xie, F. Yu, R. P. Yu, and M. Abadi. Host Fingerprinting and Tracking on the Web: Privacy and Security Implications. In *Proceedings of the Network and Distributed System Security Symposium*, 2012.

[67] K. Zhang and T. Jiang. Some {MAX} snp-hard results concerning unordered labeled trees. *Information Processing Letters*, 49(5):249 – 254, 1994.

[68] Y. Zhu, J. Jung, D. Song, T. Kohno, and D. Wetherall. Privacy scope: A precise information flow tracking system for finding application leaks. Technical Report UCB/EECS-2009-145, EECS Department, University of California, Berkeley, Oct 2009.

10 Appendix A

10.1 Team Member Identification

Individual Name	Role (Prime, Subcontractor or Consultant	Organization	Non-US?		FFRDC or Govt?
			Org.	Ind.	
Geambasu	Prime	Columbia University	N/A	N/A	N/A
Chaintreau	Prime	Columbia University	N/A	N/A	N/A
Hsu	Prime	Columbia University	N/A	N/A	N/A
Roesner	Subcontractor	University of Washington	N/A	N/A	N/A

10.2 Government or FFRDC Team Member Proof of Eligibility to Propose

NONE

10.3 Government or FFDRC Team Member Statement of Unique Capability

NONE

10.4 Organizational Conflict of Interest Affirmations and Disclosure

NONE

10.5 Intellectual Property (IP)

The Offeror and subcontractors reserve the right to independently or jointly seek intellectual protection for the results of the work under this program. These rights will not compromise the values of the proposed work to the Government because it will have access to and use of the research and results of this work.

10.6 Human Subjects Research (HSR)

The proposed work includes user studies that will involve human subject research. The proposed studies will be designed and conducted according to procedures approved by the organizations' Institutional Review Boards (IRBs). Ample time will be allotted to complete the approval process for each study.

10.7 Animal Use

NONE

10.8 Representations Regarding Unpaid Delinquent Tax Liability or a Felony Conviction under Any Federal Law

(a) The proposer represents that it is [ ] is not [**X**] a corporation that has any unpaid Federal tax liability that has been assessed, for which all judicial and administrative remedies have been exhausted or have lapsed, and that is not being paid in a timely manner pursuant to an agreement with the authority responsible for collecting the tax liability.
(b) The proposer represents that it is [ ] is not [**X**] a corporation that was convicted of a felony criminal violation under a Federal law within the preceding 24 months.

10.9 Cost Accounting Standards (CAS) Notices and Certification

NONE