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	
University of Washington Technical POC Administrative POC	 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 Andrei Stabrovski Department of Computer Science & Engineering, Box 352350 University of Washington, Paul Allen Center, 185 Stevens Way Seattle, WA 98195-2350 206-543-7165 (v) 206-543-2969 (f) andreis@uw.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 them and what not, or what protection tools they should use to prevent misuse. Our goal is to develop *transparency 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 transparency 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 building blocks and programming abstractions to facilitate the building of many such transparency tools. Using Hubble, we will develop and thoroughly evaluate several transparency tools, and will study how transparency can help shape user actions and enable them to better manage their online privacy. Our effort targets *Technical Area #2*.

Key Technical Challenges: Constructing transparency 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 [**?**, **?**, 8, 13]. 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 transparency tools that do not themselves create new privacy challenges is a difficult challenge. Intuitively, to reveal how data is being used, a transparency 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 transparency tools on the end-users is an open question. For transparency tools to be effective, they must not only help educate users (and watchdog organizations like the EFF or the FTC) 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 auditing tools to track the use of personal data at fine granularity (e.g., individual emails, photos) 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 grain is extremely limited. Our own recent system, XRay [16], 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 [30] 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 [35], AdFisher [**?**], and OpenWPM [9], 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 [**?**, 35].

Expected Impact: The greatest impact of our work will be to increase user awareness about the implications of their online actions. 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 auditing 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 transparency 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.

Cost, Duration and Team: Our proposed effort will last 4.5 years (starting on 09/01/2015), with a total cost of $4,000,000. The team members are from Columbia University and University of Washington.

2 Goals and Impact

Many of the most 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 [16, 27]. Thus, many of today's users increasingly have some notion that this data collection is happening (e.g., through extensive media reporting on the topic [32]) and that they are exchanging some amount of private information for the use of free services (email, search, social media). an intuition embodied by the popular saying "if you're not paying, you're the product." 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 (e.g., [33]). **Thus, an 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 action*, and *verify the results of their actions*.

To this end, we propose the design, development, and evaluation of **user awareness tools** to 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 (e.g., through legal or policy means), these higher-level solutions do not help individual end users in the present moment.
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 (for example, two web domains may be owned by the same company, as the advertising network Doubleclick is owned by Google). 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.

We know of no awareness tools that meet all of these goals. For example, a number of tools exist that visualize third-party web trackers (e.g., Ghostery [12], Lightbeam [20]). 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 show users how their data is being actually used by the services that collect it.

An enormous number 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 and using the data for its marketing purposes, or whether a service shares their data with third parties? We believe that being able to discern and reveal such information to the users could result in inceeased, and potentially changing their XXX. Unfortunately, constructing such tools is tremendously difficult today,

Indeed, some questions have been posed in related research and analyzed using what we would call small-scale and/or purpose-specific experiments, where the tools needed to reveal a specific aspect about personal data collection and use are developed for the purpose of that specific purpose. For example, Wallstreet Journal's investigations of online price discrimination were done in the specific context of Orbitz and a few other websites [5, 34]. To study a new travel site, one would have to build that infrastructure from scratch. This approach, taken by many in the scientific community [4, 14, 15, 19, 35], results in redundancy between investigations, and typically in small-scale, one-off experiments.

Given that the Web is a large, ever-changing system with many different services, we believe that XXX. In its place, we envision a new architecture for what we call *scalable Web transparency*. Fig.1 illustrates this architecture. It consists of three components: (1) *Web services*, which must be audited, (2) *transparency tools*, which answer specific questions about data use in individual services, groups of services, or even large portions of the Web, and (3) *building blocks*, which support the building of those tools. Our hypothesis is that *there is a great reduction in the scale of each component*. Specifically, we believe that a few core building blocks can facilitate the development of tools to answer many interesting questions about even more services in the data-driven Web.

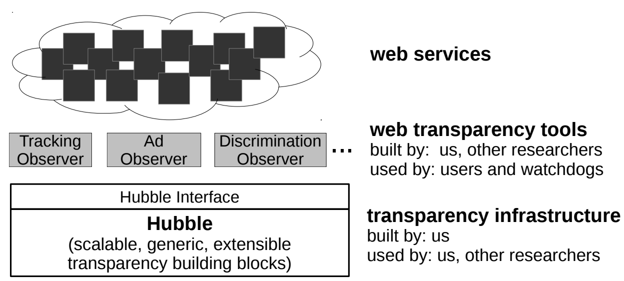

Fig. 1: **The Hubble Vision.** With a few building blocks, we hope to make it easy to build transparency tools to answer at scale many important questions about the data-driven Web.

Our specific plan is to drive the development of the building blocks by developing multiple tools and leveraging them in real measurements of data use on the Web. Described in detail in §4.2, the proposed tools we propose to build as part of this project include:

[**Give sufficient insight/motivation for each tool to make them compelling.**] xxx

1. *TrackingObservatory:* Modern web pages include large amounts of third-party content which invisible collects information about users' browsing behaviors, typically for the purposes of website analytics, targeted advertising, and other forms of personalization [27]. Though users have been made aware of such data collection through numerous prominent media reports (e.g. [32]), and effectively trade it for free access to many web services, it remains difficult for users to reason about this invisible data collection as they browse, and even more difficult for users to take action to protect their data from these trackers. We propose an extension to TrackingObserver, our previous (but limited) web tracking detection and measurement platform [30], to (1) detect a larger set of tracking behaviors (particularly fingerprint-based trackers, which are even more invisible to users that those trackers that use browser cookies)

and (2) visualize third-party content and data collection for users in a way that is effective and actionable, helping users take control of the collection and use of their web browsing behaviors. TrackingObserver is intended as a platform, allowing other researchers to adapt and build upon our tracking detection capabilities to build additional web privacy user awareness tools. [**I'd like to rename TrackingObserver if possible to distinguish from prior work. Unfortunately I'm terrible at names...** —Franzi] xxx

2. *AdObservatory:* Browser plugin that reveals which specific data in a user's profile – such as emails, website browsing history, or Facebook information – triggers which ads that the user is encountering as she surfs the web. The plugin will provide signals XXX.
3. *DiscriminationObservatory:* Browser plugin that reveals personalized content on arbitrary web pages. It is specifically geared toward XXX.
4. *LocationObservatory:* Mobile app that XXX.

[**Stress above that our goal is to enable others to build – hence our focus on reusable infrastructures.**] xxx

[**Impact.**] xxx

3 Collaborative Research Team Concept

4 Technical Plan

4.1 Thrust 1: Infrastructural Building Blocks for Web Transparency

4.1.1 The Hubble Infrastructure

[**Roxana**] xxx

4.1.2 Statistical Correlation and Causal Inference

[**Daniel**] xxx

4.1.3 Privacy-Preserving Transparency Protocols

[**Augustin**] xxx

4.2 Thrust 2: Awareness and Transparency Tools

[**Let's organize these tools into bigger tools and give them each a name. Let's have 4 tools in total. The description of each tool should be about one page in length.**] xxx

4.2.1 TrackingObserver: Revealing Third-Party Content and Tracking

[**I'm not really happy with this section but I've brain dumped it for now...** — Franzi] xxx

[**I wonder if I should name it something new, rather than TrackingObserver, to distinguish it more from prior work?** —Franzi] xxx

In prior work we developed TrackingObserver [27, 30], a browser-based web tracking detection and measurement platform. As part of the currently proposed work, we aim to extend TrackingObserver or a derivative tool to (1) detect additional web tracking behaviors and (2) develop useful user-facing visualizations or other information on top of TrackingObserver's detection and measurement capabilities.

Detecting fingerprint-based trackers. With respect to increasing the scope of TrackingObserver's detection, which currently handles primarily cookie-based trackers that explicitly store state in the user's browser, we will focus on detecting *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 [6]. While researchers have explored how fingerprinting works and conducted limited measurement studies of specific fingerprinting techniques or known fingerprinting libraries (e.g., [1,2,21,36]), there has been no extensive non-blacklist-based study of fingerprinting in the wild nor a user-facing tool to detect these behaviors. Extending TrackingObserver to support the automated detection of fingerprint-based tracking, 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) [27]. Indeed, some ads might not even have been intended by the web page developer, such as those injected by malicious browser extensions [3]. 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 TrackingObserver, 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 TrackingObserver 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 TrackingObserver 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 [7]), as we have done with ShareMeNot [29]) 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.

4.2.2 AdObserver: Revealing Targeting in Online Ads

[roxana] xxx

4.2.3 DiscriminationObserver: Revealing Discrimination in Arbitrary Online Content

4.2.4 LocationObserver: XXX Augustin's Location Tool

[augustin] xxx

4.3 Thrust 3: Measurement Studies

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., [17,18,22,33]); 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 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

Component	Sub-tasks	PI(s)
Hubble infrastructure	1.1, 1.10, 1.19	Geambasu
Statistical correlation and causation	1.3, 1.4, 1.12, 1.13, 1.21, 1.22	Hsu
Privacy-preserving transparency	1.5, 1.14, 1.23	Chaintreau
TrackingObserver	1.7, 1.16, 1.17, 1.25	Roesner
AdObserver	1.2, 1.11	Geambasu
DiscriminationObserver	1.11, 1.20	Geambasu
LocationObserver	1.6, 1.14, 1.22, 1.30	Chaintreau
User studies	1.8, 1.25, 1.26	Roesner, Chaintreau, Geambasu
Integration, Evaluation on TA3	1.9, 1.18, 1.27	All PIs

Table 1: **Team member responsibilities (research areas and subtasks).**

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

Table 2: **Team member commitments.**

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 [29]) as part of the Electronic Frontier Foundation's Privacy Badger tool [7]. 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. [**Roxana**] We will leverage AdObservatory and DiscriminationObservatory to run these studies to reveal websites that discriminate and warn users about that. xxx

5 Management Plan

The team members are facutly 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 subtasks, as identified in Table 3. 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 2.

[**Guys: in Table 3 please include in the tool tasks the measurement tasks, as well. Please feel free to combine people if you want to work together.**] xxx

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. 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 Geambasu 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, record keeping, and coordination. We will organize a website to make all of our findings publicly available. **All code resulting from this program will be released open-source.**

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. For detailed participant qualifications, biographies, and time commitments, please see Section 9. Please see Section 6 for a brief discussion of joint projects and other work highlighting the team expertise.

5.1 Integration and Evaluation

[**Let's see what goes here.**] xxx

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 transparency in computer systems for multiple years. As part of a DARPA MRC project (MEERKATS), she has built a series of new OS-level data management abstractions that leverage information flow systems to recover the structure of high-level application objects – such as emails, documents, or bank accounts – to provide a new level of transparency and control over users' data in clouds and on mobile devices [?, 31]. For example, using one of her recent systems, Pebbles [?], a user can tell exactly whether pieces an object in an arbitrary, unmodified, mobile app, are left after he deletes that object – a capability that is missing from today's OSes. Geambasu and Chaintreau have also recently developed a new tool, called *XRay*, which increases transparency of Web services with black-box testing and correlation [?]. The system is the very first to accurately reverse targeting within and across multiple services, including Gmail, Amazon, Youtube, Google News, as well as (most recently) Web-tracker-based targeting. Geambasu's work and interest in transparency dates back to her development of Keypad, an auditing file system for theft-prone devices that increases a user's visibility into what happens with their data after device theft [10]. Geambasu and Nieh have developed large-scale data-driven systems [?, 11], including a system called Synapse [?], which supports heterogeneous-DB replication to greatly facilitate the building of modern data-driven Web applications, which all tend to be very heterogeneous. That system is now deployed at a NYC startup, where it has been running in production with 650K users for about a year.

PI Chaintreau

PI Hsu

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 [27] 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 [?], a populer anti-tracking browser add-on, and ShareMeNot's code itself was incorporated into the Electronic Frontier Foundation's Privacy Badger [7] 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* [26] 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 [23, 24]. Her work has also focused on emerging security and privacy challenges in emerging augmented reality and continuous sensing platforms [25, 28].

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.
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.

7.1 Phase 1 (Months 1-18)

Our objective for Phase 1 is to develop a basic prototype of Hubble infrastructure and core building blocks, plus subset of basic transparency tools implemented and evaluated. In the last three months, we will work with TA3 performers to demonstrate our system on a TA3 Research System. Specific Phase 1 tasks follow.

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**: Add fingerprint tracking detection infrastructure to TrackingObserver.
General Description: We will extend TrackingObserver, our existing web tracking detection and measurement platform, to detect 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: Modified version of TrackingObserver that successfully detects a large fraction of fingerprint-based trackers, evaluated by a comparison with blacklist-based tracking detection tools.
Deliverables: Improved version of TrackingObserver 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: XXX
Responsible Organization and Location: Columbia University (New York, NY), University of Washington (Seattle, WA)
Exit Criteria: XXX.
Deliverables: XXX.

7.2 Phase 2 (Months 19-36)

Our objective for Phase 2 is to enhance the Hubble prototype, building blocks, and transparency tools to support more features and use cases. In the last three months, we will work with TA3 performers to demonstrate our enhanced technologies on a TA3 Research System. Specific Phase 2 tasks follow.

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 remain in the range of 100s-1000s inputs, but we expect its recall/precision to grow significantly 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: XXX
Responsible Organization and Location: Columbia University (NYC)
Exit Criteria: XXX
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 using sample-splitting to support testing of complex targeting hypotheses: in the first phase, we generate putative input combinations and groups based on correlations in an initial set of data; in the second phase, we test introduce new input combinations from the first phase and also exploit input group structure using group-sparse linear regression. 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 of group-sparse linear regression approach, 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 new additions to TrackingObserver.
General Description: With the improved version of TrackingObserver developed in Year 1, 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 TrackingObserver, 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: XXX
Responsible Organization and Location: Columbia University (New York, NY), University of Washington (Seattle, WA)
Exit Criteria: XXX.
Deliverables: XXX.

7.3 Phase 3 (Months 37-54)

Our objective for Phase 2 is to finalize the Hubble prototype, building blocks, and transparency tools, and to evaluate user reactions to transparency. In the last three months, we will work with TA3 performers to demonstrate our technologies on a TA3 Research System and a TA3 Existing System. Specific Phase 3 tasks follow.

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: XXX.
Responsible Organization and Location: Columbia University (NYC)
Exit Criteria: XXX.
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 TrackingObserver. 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 a TA3 Research System and on a TA3 Existing System.
General Description: XXX. Include Franzi's stuff from Y4.
Responsible Organization and Location: Columbia University (New York, NY), University of Washington (Seattle, WA)
Exit Criteria: XXX.
Deliverables: XXX.

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. 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 gannt 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. The only difference is that I think we need to reflect the program's milestones (see the BAA for this section).**] xxx

Here's the content of what will become a gannt chart:

9 Personnel, Qualifications, and Commitments

[**May need to move these, but please fill them in.**] xxx

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 ISAT group, having been appointed in 2014 to serve for a period of three years. She is co-organizing an ISAT workshop this summer on "Privacy in Today's Data-Driven World" (formerly known as "Where Are My Data?") For her work in privacy, Prof. Geambasu received a Microsoft Research Faculty Fellowship, a "Brillint 10" Popular Science nomination, an NSF CAREER award, 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, and the first Google Ph.D. Fellowship in Cloud Computing. Prof. Geambasu's research has been featured by high-profile media outlets, including The New York Times, The Economist, NPR, and others. Prof. Geambasu has been a member of the Information Science and Technology (ISAT) study group since Fall 2014.

Augustin Chaintreau

Daniel Hsu

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`. 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.

10 Cost Summary

References

[1] 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.

[2] 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.

[3] 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/`.

[4] N. Diakopoulos. Algorithmic accountability reporting: On the investigation of black boxes. Tow Center for Digital Journalism, Columbia University. February, 2014.

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

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

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

[8] 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.

[9] 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.

[10] R. Geambasu, J. P. John, S. D. Gribble, T. Kohno, and H. M. Levy. Keypad: An auditing file system for theft-prone devices. In *Proc. of the ACM European Conference on Computer Systems (EuroSys)*, 2011.

[11] 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.

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

[13] 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.

[14] S. Guha, B. Cheng, and P. Francis. Challenges in measuring online advertising systems. In *IMC '10: Proceedings of the 10th annual conference on Internet measurement*. ACM Request Permissions, Nov. 2010.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

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

[21] 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.

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

[23] 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.

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

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

[26] 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.

[27] 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.

[28] 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.

[29] 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/`.

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

[31] 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.

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

[33] 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.

[34] J. Valentino-Devries, J. Singer-Vine, and A. Soltani. Websites Vary Prices, Deals Based on Users' Information. *online.wsj.com*, pages 1–6, Dec. 2012.

[35] 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.

[36] 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.

11 Appendix A

11.1 Team Member Identification

The team member organizations are:

1. Columbia University in the City of New York (prime, Other Educational, US organization)
2. University of Washington (subcontractor, Other Educational, US organization)

11.2 Government or FFRDC Team Member Proof of Eligibility to Propose

NONE

11.3 Organizational Conflict of Interest Affirmations and Disclosure

NONE

11.4 Organizational Conflict of Interest Affirmations and Disclosure

NONE

11.5 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.

11.6 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.

11.7 Animal Use

NONE

11.8 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.

11.9 Cost Accounting Standards (CAS) Notices and Certification

NONE

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	Geambasu, Hsu, Chaintreau
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	Geambasu, Hsu, Chaintreau
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	Geambasu, Hsu, Chaintreau

Table 3: **Timeline. XXX This will become a gannt chart eventually – guys please help on this if you have good drawing tools.**

Participant	Project	Status	Level of Effort on Project			
			FY16	**FY17**	**FY18**	**FY19**
Roxana Geambasu	DARPA Brandeis	Proposed	8%	8%	8%	8%
	NSF SaTC CAREER	Current	8%	8%	8%	8%
	NSF SaTC Medium XRay	Pending	4%	4%	4%	4%
	DARPA MRC MEERKATS	Current	16%	0	0	0
Augustin Chaintreau	DARPA Brandeis	Proposed	8%	8%	8%	8%
	NSF XXX CAREER	Current	XX%	XX%	XX%	XX%
	NSF SaTC Medium XRay	Pending	4%	4%	4%	4%
Daniel Hsu	DARPA Brandeis	Proposed	8%	8%	8%	8%
	XXX	XXX	XX%	XX%	XX%	XX%
Franziska Roesner	DARPA Brandeis	Proposed	8%	8%	8%	8%
	XXX	XXX	XX%	XX%	XX%	XX%

Table 4: **Time Commitments (per project year, which coincides with fiscal).**