

# Columbia University Morningside or Lamont Doherty Consent Form

**Attached to Protocol: IRB-AAAO0259**

**Principal Investigator: Steven Bellovin**  
(smb2132)

**IRB Protocol Title: Cross-device Internet User Tracking**

**Consent Number: CF-AAAT2535**

**Participation Duration: 1 month**

**Anticipated Number of Subjects: 100**

## Contact

<u>Contact</u>	<u>Title</u>	<u>Contact Type</u>	<u>Numbers</u>
Sebastian Zimmeck		Emergency	Telephone: 347-694-4712 Cell: 347-694-4712
Sebastian Zimmeck		Administrative	Telephone: 347-694-4712 Cell: 347-694-4712
Sebastian Zimmeck		Co-Investigator	Telephone: 347-694-4712 Cell: 347-694-4712

## Research Purpose

The purpose of this study is to examine the extent to which it is possible to correlate different computing devices (smartphones, laptops, etc.) to one particular Internet user based on computer usage data (e.g., the type of browser or the visited websites). We also want to find out what characteristics of a person can be learned (e.g., gender or age) from the usage data.

## Information on Research

The purpose of this form is to give you information to help you decide if you want to take part in a research study. This consent form includes information about:

- why the study is being done;
- the things that you will be asked to do if you participate in the study;
- any known risks involved; and
- any potential benefits.

A co-investigator for this project, Sebastian Zimmeck (sebastian@cs.columbia.edu), is available to discuss the study with you. If at any time you have questions about the study, please contact the co-investigator. Take all the time you need to decide whether you want to take part in this research study and let us know of any questions that you might have.

## Study purpose

We are doing this research study to find out to which extent it is possible to identify the different computing devices (e.g., laptop, smartphone, tablet) of one particular Internet user among the devices of other Internet users and what can be learned about that user (e.g., age and gender). The identification and analysis is based on computer usage data (e.g., web history and app usage data). For, example from observing that one laptop visits a chess website and an Android phone accesses a chess app, it could be concluded that both devices belong to the same user. Finding out to which extent this correlation is possible helps us in our understanding of Internet privacy. We also want to find out what can be learned about Internet users, in the example, that a user is interested in chess.

We are seeking study participants between 18 and 65 years old from the Columbia University community (i.e., students, staff, professors). About 100 to 200 participants are expected to be enrolled in this study in total.

We are not aware of any major risks associated with this study. Please note, though, while we apply state-of-the-art security practices, the loss and abuse of data can never be fully excluded. To avoid such data loss or abuse we will strictly adhere to organizational and technological measures to secure the data (e.g., data access is only available to project members on a need-to-know basis and data is stored in encrypted format). It should also be noted that we do not collect any sensitive data. In particular, we do not collect health, financial, or children data. Please also consider that much of the data that we receive is routinely available to your Internet service provider, work/school network administrator, and the web services (e.g., Google or Facebook) that you use.

## Procedures

Study participants will be enrolled online at this website. When you start the enrollment process you will need to download and install a software program for each browser and mobile device that will enable us to collect your browsing history (and app usage in case of mobile devices). After you installed the software we will analyze certain technical information about your browser (e.g., whether you are using Google Chrome or Apple's Safari browser). You will also need to fill out a short questionnaire with personal information and a few questions on your interests (e.g. whether you are interested in sports). After you submitted the information your browser is set up for the study. You will repeat the enrollment process for every browser on all devices that you want to enroll as well as for all mobile devices.

After you have completed the enrollment process we will receive your data. For desktop/laptop devices the data transmission will be an automated process that does not require you to do anything. For mobile data, however, we require your assistance to send us some data through a manual upload process at the end of the data collection process. We will send you an e-mail at that time with

instructions how to send us the data.

Your study participation is purely online and does not require any personal visits (except if you wish to speak to us personally, in which case a co-investigator will be available to meet with you). We estimate that the time you need to spend for downloading and installing the software as well as other study requirements will not take more than one hour. The study has a duration of up to three months. In other words, we will collect your data for a period of up to three months. After we obtained your data at the end of that period you will receive your compensation for taking part in the study in form of an Amazon gift card.

## **Compensation**

---

### **WILL I GET COMPENSATED?**

#### **Payment**

You will be compensated for your study participation. Study participants for the full month of data collection will be compensated \$50 in form of two Amazon gift cards (\$10 and \$40). We will send the gift cards by e-mail to you after one month of data collection is complete. However, please note that we reserve the right to exclude any study participant for the second amount if there is not sufficient data submitted. So, please make normal use of your computers and phones during the study to guarantee that you will receive the full compensation for three months.

The Amazon gift cards do not have an expiration date. We will send them to the e-mail address that you specify when signing up for the study on this website.

Future testing and research on the collected data may lead to the development and use of information and products that have commercial value. You will not receive any financial compensation that may result from this research.

## **Confidentiality**

---

### **WHAT ABOUT CONFIDENTIALITY?**

#### **Confidentiality Protection**

We collect certain information from you (e.g., your name, web history, app usage history, and interests). If you do not want us to collect information at some point during your participation in this study, please pause the software for the time that you would not want to submit information (In our app click "Stop Browsing History Collection". In your browser, disable our browser extension). You can also decide to not enroll certain devices or browsers. In that case we would not be able to collect any data from the devices and browsers that you are not enrolling. However, please keep in mind that you need to enroll at least one mobile device (smartphone, tablet, etc.) and one desktop device (desktop, laptop, etc.). Please also note that we need a certain amount of information for our study to work. So, visiting one website per day, for example, would be insufficient use and leaving you

unqualified for the full compensation.

Any information collected from you during this study will be kept confidential. We will apply state-of-the-art technologies to keep your data secure and private. In particular, our data transmission and storage makes use of provably secure encryption. We also use best practices for data management. In particular, we collect as little information as possible and only members of our study group have data access on a need-to-know basis. Despite our efforts, however, we cannot promise you absolute confidentiality protection. Unanticipated problems, such as a stolen computer and secure keys may occur, although it is highly unlikely.

Please note that we will not share any of your information with anyone outside of our research group except as explained in the following:

The following individuals and/or agencies will be able to look at and copy your research records:

The investigator, study staff and other professionals who may be evaluating the study

Authorities from Columbia University, including the Institutional Review Board ('IRB')

The United States Office of Human Research Protections ('OHRP')

After the conclusion of our study we plan to make our dataset in anonymized form publicly available to researchers and other interested parties. This release will not include any of your personal information (such as your name or e-mail address). We plan to publicly release your app/browsing histories, device fingerprints, interests, and personas. As the browsing histories might contain identifiers (e.g., if you searched for your name on Google), we will anonymize the data before the release. More specifically, we will apply a secure cryptographic hash function ([https://en.wikipedia.org/wiki/Cryptographic\\_hash\\_function](https://en.wikipedia.org/wiki/Cryptographic_hash_function)) with a salt ([https://en.wikipedia.org/wiki/Salt\\_\(cryptography\)](https://en.wikipedia.org/wiki/Salt_(cryptography))) to every item of your browsing history. Such hash is provably very hard to break. To maintain utility we might hash the different parts of each item separately, such that, for example, <http://www.facebook.com> and <http://www.facebook.com/privacypolicy/> can be identified as coming from the same domain, i.e. the two urls would be <http://www.cbe648909034c0624c205fe219d3fbd10052c715.com> and <http://www.cbe648909034c0624c205fe219d3fbd10052c715.com/3a08d23a0703e42e42f8a0c867fb6399735b4f4c>, respectively.

## Risks

### WHAT ARE THE RISKS OF THE STUDY?

#### Loss of confidentiality

We will protect the confidentiality of your data using state-of-the-art technology and privacy procedures (which are described in the 'confidentiality' section of this consent form). However, you should know that the possibility of a loss of confidentiality can never be excluded with certainty. Loss of confidentiality includes having your information disclosed to someone who is not on the research team and was not supposed to see or know about your information. In case of such loss the research

team will notify you of the confidentiality breach and undertake efforts to protect you from any possible harm.

## **Benefits**

---

### **ARE THERE BENEFITS TO TAKING PART IN THE STUDY?**

#### **Direct benefit**

A benefit of taking part in this study is that we can tell you to which extent your browsing and app usage habits make you identifiable on the Internet and what others can learn about you. If you wish to receive such information, please contact us.

#### **Indirect benefit**

In addition, the information collected from this research may help us to better understand privacy on the Internet in general.

## **Alternative Procedures**

---

### **WHAT OTHER OPTIONS ARE THERE?**

#### **Alternatives**

You may choose not to take part in this research study.

## **Voluntary Participation**

---

### **DO I HAVE TO BE IN THE STUDY?**

#### **Voluntary Participation**

Taking part in this study is your choice. You may refuse to take part in the study or withdraw from the study at any time without jeopardizing your employment, student status, or any other rights. However, please note that you might not receive your (full) compensation (see the Compensation section). The study team reserves the right to withdraw participants in case of misuse or deceit.

## **Additional Information**

---

### **WHOM DO I CALL IF I HAVE QUESTIONS OR PROBLEMS?**

#### **Questions**

If you have any questions or concerns on this research study, you should contact the co-investigator

Sebastian Zimmeck  
sebastian@cs.columbia.edu  
347-694-4712

or the principal investigator

Steven M. Bellovin  
smb@cs.columbia.edu  
212-939-7149.

If you have any questions about your rights as a research study participant in this study, you should contact the Institutional Review Board (IRB) at Columbia University by phone at 212-851-7040 or by email at askirb@columbia.edu.

More information about taking part in a research study can be found on the IRB website at <http://www.columbia.edu/cu/irb>

## STATEMENT OF CONSENT

### My Consent

I have read the above purpose of the study, and understand my role in taking part in the research. I volunteer to take part in this research. I have had a chance to ask questions. If I have questions later, about the research, I can ask the co-investigator or principal investigator listed above. I understand that I may refuse to participate or withdraw from participation at any time without jeopardizing my employment, student status or other rights to which I am entitled. If I have questions about my rights as a research participant, I can contact the IRB as described above. I certify that I am 18 years of age or older, but not older than 65, and freely give my consent to participate in this study. I will receive an electronic copy of this document for my records.

By clicking on the "I agree" button at the enrollment website, I consent to the above terms.