Name: _____ **USC ID:** _____

INF 529 Final Exam

Spring 2017

Instructions:

Show all work. No electronic devices are allowed. This exam is open book, open notes, but no use of electronics are permitted during the exam. You have **120 minutes** to complete the exam.

Please prepare your answers on separate sheets of paper. You may write your answers on the sheet of paper with the question. Use the space following the question, until the next question, and if instructed, use the back of the same sheet to answer. Do **NOT** extend your answer onto other sheets. I am looking for relatively short answers, and answers that extend onto additional sheets will not be considered. I am not looking for you to copy as much material from the notes as possible in hope that something you say might be correct. Part of a quality answer is for you to select the correct answer, rather than list as much as possible.

I will provide scratch paper so that you may take notes and plan your answer before committing it on the exam. If you need to change an answer and cannot fit the change into the allotted space, I will consider a replacement answer on a fresh sheet of paper, using no more than the same space allotted on the original exam sheet.

Be sure to include your **name** and **USC ID** number **on each page**.

There are **100 points** in all and **3 questions.**

	Q1	**Q2**	**Q3**		**Total Score**
Score					

Name: _____ **USC ID:** _____

1) The Impact of "Big Data" on Privacy

Big data, or data science, is enabled by several differences in the handling of data in today's systems. These differences have a profound impact on the actual privacy that exists and to the balance of power between businesses and individuals. Two of these differences are the capability to store much more data than in the past, and in the accessibility of this data.

a) Explain how the sheer volume of data that can be practically stored has changed our expectations of privacy. What kinds of records can be stored now that could not be stored in the past? What else about those records has changed because we are no longer limited by the size of our storage repositories? How does this storage capability affect out privacy? Give examples from real world systems (there were examples provided in class) where our privacy is affected because companies are storing more data. Include in your examples, systems that have changed because of the kinds of data, as well as those that have changed because of the amount of data that can be stored. (confine your answer to the front of this page - 15 points)

b) Explain how the ability to search data has changed our expectations of privacy. When searching data, what characteristics in the data being searched most profoundly affect the impact to our privacy? Give examples from real world systems (there were example provided in class) where our privacy is affected by the ability to search collected data, and thus impact our privacy. For those examples, explain the characteristics of the data that enable effective search, and any technological advances affecting those characteristics. Be sure to consider the ability to search multiple kinds of data. (answer only on the back of this page - 15 points)

Name: _____ **USC ID:** _____

c) Explain how machine learning, data mining, and statistical inference methods can use "big data" about us in ways that are against our personal well-being. How can these techniques uncover (or discover) possibly incorrect information, and how do they create or reinforce profiles and stereotypes that society has long sought to abolish. (answer only on the front of this sheet - 10 points)

Name: _____ **USC ID:** _____

2) Legislation, regulation, and legal limits on "Big Data"

a) Explain how certain regulations or business practices, both within the US, and in Europe, place limits in certain specific cases on the "sheer volume" of data that is retained about our activities. Some of these rules might limit the collection or retention of certain data, or they provide procedures to remove data. In answering, consider several rules or practices that have been discussed in class. (answer only on the front of this page - 15 points)

b) Drawing upon discussions in class, discuss several rules or regulations that impact our ability to search data and share data. In answering this questions, consider limits on the characteristics of the data as discussed in 1(b), whether resident in our own data bases, or in the datasets we share. (answer only on the back of this sheet - 15 points)

Name: _____ **USC ID:** _____

3) Technologies that improve privacy and security

a) "A release of data is said to have the **k-anonymity** property if the information for each person contained in the release cannot be distinguished from at least **k**-1 individuals whose information also appear in the release."

Give an example of a problem with k-anonymity? Specifically, constructs an example of a data set for which personally identifiable information might be available even if the k-anonymity property holds. (select 2 or 3 for k so that your example is short).

(10 points - answer only on the front of this page)

b) Proxies and onion-routing

Why does the use of an anonymizing proxy, or better yet, onion routing, improve our privacy? Specifically, what is the characteristic of data collected about our activities (as discussed in 1(b)) that is removed from the data potentially collected and stored, and why might removal of that characteristic improve out privacy? (10 points - answer only on the back of this sheet)

c) end-to-end encryption

How does the use of end-to-end encryption in a system like PGP, or Whats Ap improve our security? Why are we more secure with end-to-end encryption that if we relied on SSL or TLS to protect our communication with a server.

(10 points - answer only on this front of this page)