# CPSC 457/557 – Fall 2011 – Answer Key for Exam 2

Question 1: Privacy and Online Advertising

(a) General-interest websites such as http://www.cnn.com would be severely affected. Without behavioral information about users, advertisers would not be able to target ads very precisely, because the mere fact of these users' visiting http://www.cnn.com does not say much about which products and services they would be interested in; precisely targeted ads command higher prices, and losing them would severely impact the general-interest websites' revenues. By contrast, special-interest websites such as http://www.hotels.com would be only modestly affected, because there is enough known about users of these websites to allow advertisers to tailor their messages without tracking and modeling individuals.

(b) Privacy advocates believe that the Kerry-McCain proposal is too weak to protect consumers, and they object to its not offering a "universal opt-out" mechanism analogous to the telephone "do-not-call list."

(c) No, AdSense and AdWords do not currently use them, because Gmail doesn't store them.

(d) DoubleClick cookies

(e) The NY State Bar Association regards Gmail as a "contractor," analogous to other commercial third parties, such as litigation services, to which attorneys outsource work and thus reveal privileged information. If such a contractor takes appropriate steps to secure the information, then an attorney's sharing the information with the contractor does not violate client confidentiality. The Association also regards as significant the fact that the "concepts" that Gmail uses to choose ads are extracted by a computer program, not a person.

(f) One advantage of this approach is that it may protect against ``linkage attacks'' that can reveal the user's identity as well as his location. One disadvantage is that it requires a trusted third party.

Question 2: Differential Privacy

(a) See Definition 1 in "A Firm Foundation for Private Data Analysis," by C. Dwork. This paper was covered in class, and a link to it is provided on the course website.

(b) Dalenius's desideratum is that anything that can be learned about a respondent from the statistical database should be learnable without access to the database. No, it is not attainable using differential privacy; in fact, it is not attainable at all.

(c) A subset of the records in the database is chosen uniformly at random and published; statistics can then be computed on the subset and, if it is sufficiently large, the results may be quite similar to the corresponding statistics for the entire database. To see why subsample-and-release cannot serve as the function $K$ in the definition of differential privacy, consider databases $D$ and $D^1$ such that record $X$ is in $D$ but not $D^1$. If $S$ is the set of all subsamples that contain $X$ (or, equivalently, the property that record $X$ is "released,"), then clearly the probability that $K(D^1)$ is in $S$ is zero. If this $K$ were differentially private, then the probability that $K(D)$ is in $S$ would have to zero as well, but that is nonsensical: If $X$ is in $D$,

and a uniformly random subsample of $D$ is released, then $X$ will be released with nonzero probability.

Question 3: Digital Currency

(a) The four properties are:
   i. Forgery of a bill should be infeasible.
   ii. Duplication of bills (by a user trying to spend them or a vendor trying to deposit them) should be preventable or detectable.
   iii. Users should remain anonymous.
   iv. Online interaction with the bank should not be necessary when a user pays a vendor.

(b) In a blind-signature scheme, party *A* can obtain party *B's* signature on a document *M* without revealing *M* to *B*. Blind signatures are used in the digital-cash scheme to provide user anonymity.

(c) If the bank receives two triples *(M, s, RIS), (M, s, RIS')*, where *RIS* and *RIS'* are different, from two different vendors, then it is overwhelmingly likely that a user double-spent *M*. For some $j$ such that the $j^{th}$ bit of *RIS* differs from the $j^{th}$ bit of *RIS'*, the bank should obtain $x_j$ and $x_j^1$ from the vendors, XOR them to recover the Username, and report the user to the authorities. If the bank receives the same triple twice from the same vendor, then he knows that that vendor is trying to deposit the same bill twice and should report this vendor to the authorities. For the bank to receive the same triple from two different vendors, the two vendors would have had to choose the same *RIS*; because these strings are chosen independently and uniformly at random, the probability that they are identical is $2^{-K}$.

(d) BitCoin maintains a *public* record of all transactions, whereas in basic HashCash the client and server conduct a transaction bilaterally and need not share the record of this transaction with anyone else. The difficulty of a proof of work in the BitCoin protocol changes over time in order to maintain the rate of one new block per 10 minutes; there is no analogous rate-limiting mechanism in basic HashCash.

Question 4: Online identity and anonymity

(a) The main argument in favor is the same as the argument in favor of protection of anonymous communication in any medium: Throughout US history, anonymous and pseudonymous communication have played important political and social roles, and courts have granted them First-Amendment protection; this protection must be extended to the Internet now that it is a crucial communication medium. One argument against is that, unlike newspapers and magazines, websites often contain user-generated content that is not edited; thus, it is not feasible on the Web, as it is in older print media, to "catch" wildly inaccurate and potentially harmful utterances before they reach millions of people.

(b) The website operator could store the IP addresses or account information of its users but not reveal them online; he could reveal them offline to law-enforcement personnel if he receives a subpoena. Users who are not covered by the subpoena could continue to use the website anonymously or pseudonymously.

(c) An "identity credential" contains identity data about the bearer of the credential, *e.g*., name, age, address, or nationality. An "attribute credential" contains a property of the bearer, *e.g.,* that he or she is a Yale student or a NYS employee, and it need not contain any information that identifies the bearer more precisely. By providing a credential, the issuer asserts that he has verified the bearer's identity or the fact that the bearer has the required property.
(d) Phishing