THE UNIVERSITY OF AUCKLAND

SEMESTER TWO, 2015
Campus: City

COMPUTER SCIENCE

Software Security

(Time allowed: 0.25 hours)

NOTE: Your answers on this sample examination will not affect your final grade.

Please do NOT put your name on your answer sheet.

On the last day of lectures, we will discuss how we would mark some of the answers written by you or your classmates.

We advise you to spend about 5 minutes on a 5-mark question.

Very roughly: you should write two to four sentences when answering a 5-mark question.

We advise you to reserve at least 10 minutes at the end of the examination period, so that you can review your answers for accuracy prior to submitting your answer sheet.

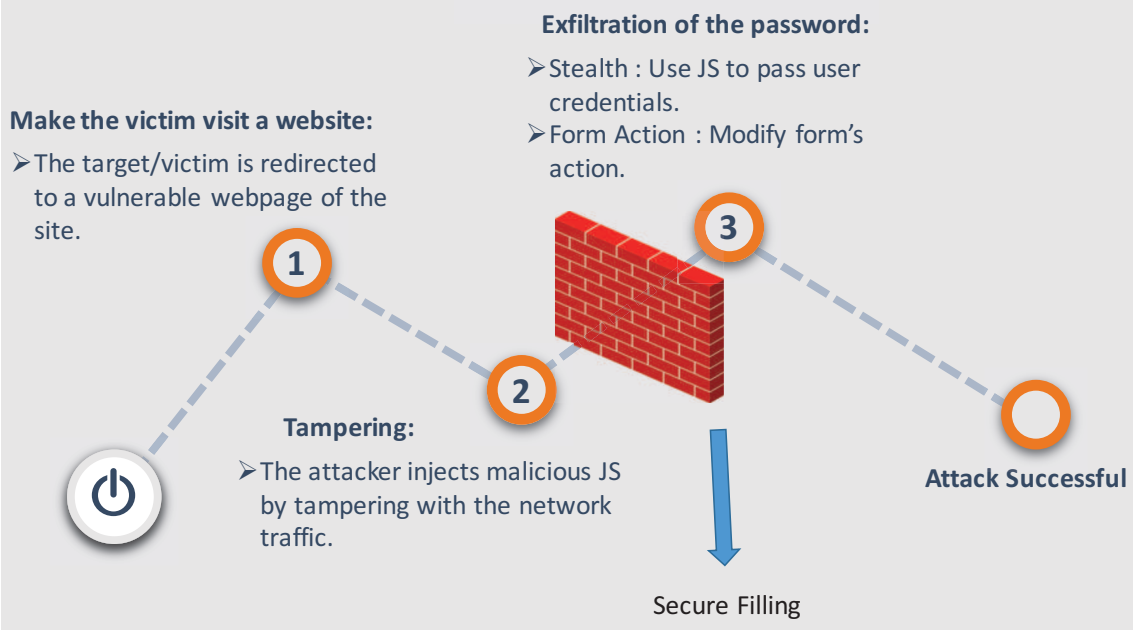
This is a **closed-book** examination.

Calculators are **not permitted** in this examination.

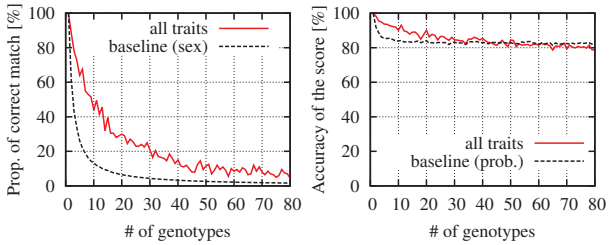
1. (20 marks, in total) Recall a threat scenario in [Humbert 2015]:

> ... if an adversary has access to phenotypic traits (e.g., visible traits) of an identified individual, he can use known correlations between phenotypic traits and genomic data to identify the genotype of this individual in a genomic database and to infer other sensitive information (such as predispositions to severe diseases) by using the de-anonymized genomic data... The adversary also has access to anonymized genotypes through a collaborative genome-sharing platform (such as the Personal Genome Project) and [wants] to de-anonymize them by relying upon phenotypic information gathered on online social networks (OSNs). The matching between OSNs and genomic profiles is (even) easier if, for example, the ZIP code is available with the genomic profiles, thus enabling the OSN profiles to be filtered before the matching attack.

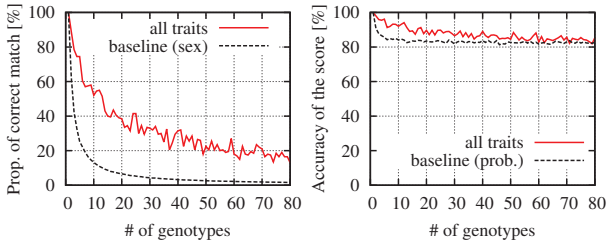
(a) (10 marks) The following figure appeared in a student's slideshow this semester. Draw a figure, in a similar style, which depicts threat scenario in the passage quoted above. To receive full marks, your figure must show a defensive "wall", and it must show at least two steps in an attack.



(b) (5 marks) What does the following figure tell you about the threat scenario of this question?



(a) unsupervised



(b) supervised

Figure 1. Performance of the identification attack for a single individual in a genomic database in the (a) unsupervised and (b) supervised cases, with respect to the proportion of correct matches (left) and the accuracy of the susceptibility score for Alzheimer's disease (right). The size of the database varies from 1 to 80.

(c) (5 marks) Do you agree with the following assertion in [Humbert 2015]? "Our results demonstrate the serious de-anonymization threat currently posed to individuals sharing their SNPs in genomic databases." In either case, justify your answer.