Practical 4 – Reputation Systems

Reputation systems are regularly used to develop trust among the users of ecommerce and other online services. Given their growing importance, the numerous security threats that can affect them raise serious concerns for both researchers [1,2] and policy makers [3,4].

[1] Kevin Hoffman, David Zage and Cristina Nita-Rotaru. A Survey of Attack and Defense Techniques for Reputation Systems, *ACM Computing Surveys*, vol. 42. no. 1, December 2009. http://dl.acm.org/citation.cfm?id=1592452.
[2] Qinyuan Feng, Ling Liu, and Yafei Dai. 2012. Vulnerabilities and countermeasures in context-aware social rating services. ACM Trans. Internet Technol. 11, 3, Article 11 (February 2012). http://dl.acm.org/citation.cfm?id=2078319.
[3] UK Competitions & Markets Authority. Online reviews and endorsements report. 19 June 2015. https://www.gov.uk/government/consultations/online-reviews-and-endorsements.
[4] European Commission. Study on Online Consumer Reviews in the Hotel Section. Final Report, 2014. http://bookshop.europa.eu/en/study-on-online-consumer-reviews-in-the-hotel-sector-pbND0414464/.

Exercise 1

This exercise reformulates parts of a question from the 2012 PSEC assessment, to allow you to explore security aspects of reputation-based systems within the space of a practical session.

In everyday life the concept of "reputation" is of considerable importance. Working in teams, consider the following questions:

- What is a reputation?
- How is a reputation made?
- By what means do we protect our reputations from attack?
- How may reputations be exploited, abused or subverted?
- What do we do when faced with someone of unknown reputation?
- How may reputations change?
- Are reputations necessary?

Now address the above questions for computer-based systems.

Make notes summarising your team's findings for a class discussion.

Exercise 2

Major providers of many online services, including Amazon and eBay, collect customer ratings of products and/or sellers, and use *reputation metrics* to derive measures of reputation from these data.

This exercise (adapted from the 2015 PSEC assessment) is concerned with evaluating the effectiveness of different *statistical measures of central location*[1] as reputation metrics for discrete-valued rating systems. These are systems in which customers rate products or sellers on an integer scale between 1 and *MAX_RATE*, where 1 represents the worst rating and *MAX_RATE>1* represents the best rating. An example is the Amazon product rating system, for which *MAX_RATE* = 5 (stars), as illustrated in Figure 1.

[1] *NIST/SEMATECH e-Handbook of Statistical Methods*, Section 1.3.5.1 – Measures of Location, 2012. http://www.itl.nist.gov/div898/handbook/eda/section3/eda351.htm.

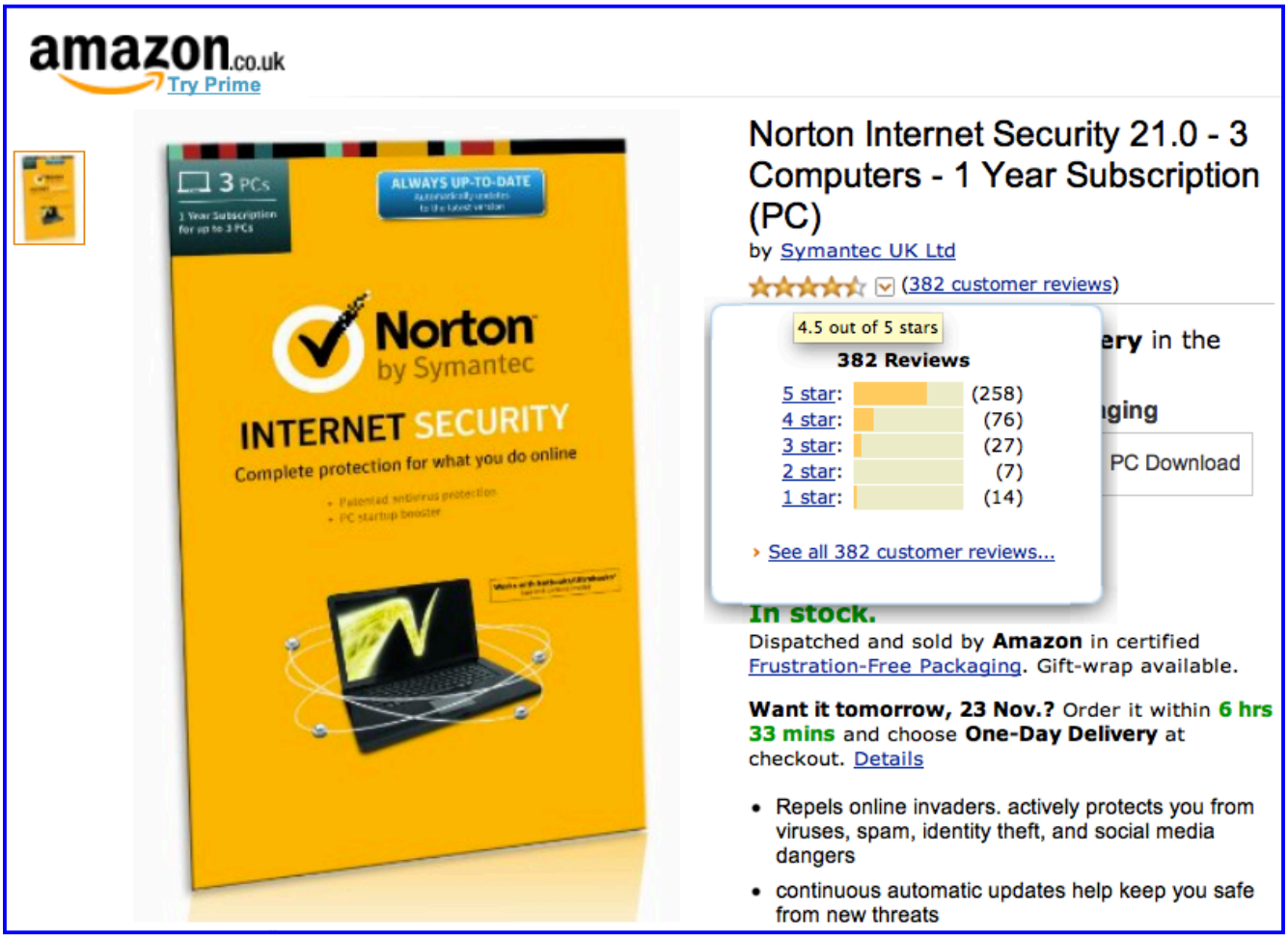

Figure 1: Amazon's use of the mean rating as a product reputation metric (NB: Recently, this calculation was replaced with a proprietary machine learning solution)

Your task is to **complete the implementation of `ReputationMetricModeller.java`**, which is a simple software tool for analysing and comparing the resistance of several reputation metrics to *self-promoting* and *slandering* attacks. In a self-promoting attack, malicious users give only ratings of *MAX_RATE* to the rated product or service. In a slandering attack, malicious users give only ratings of 1.

Notes:

1. Load the incomplete implementation of `ReputationMetricModeller.java` (provided on the module VLE) into an IDE such as Eclipse, and add the code for the methods marked with the comment 'Implement this method'.
2. If you use the lab PC, boot it into Windows (you will find Eclipse under Programs/Software development), as you will also need to use Excel to visualise the results produced by the tool.

The tool takes as input the name of a two-line file containing:

a) the customer rating data for a product in the format

$MAX\_RATE,n_1,n_2,\ldots,n_{MAX\_RATE}$

on line 1, where n_1, n_2, …, n_{MAX_RATE} represent the numbers of customers that gave the product ratings of 1, 2, …, *MAX_RATE*, respectively;

b) a self-promoting or slandering attack descriptor in the format

+min,*max*

for a self-promoting attack and

–min,*max*

for a slandering attack on line 2, where *min* and *max* represent the minimum and maximum number of expected fake ratings for the attack, respectively.

As an example, the input file containing the lines

```
5,14,7,27,76,258
+10,25
```

corresponds to the customer rating data for the Amazon product from Figure 1, and to a self-promoting attack comprising between 10 and 25 fake '5 star' additional ratings, respectively. Conversely, a '`-10,25`' attack descriptor would specify a slandering attack comprising between 10 and 25 fake '1 star' additional ratings.

Your complete implementation of the tool will analyse the impact of all attack sizes between *min* and *max* on the overall rating of the product, when each of the following six statistical measures of central location is used as a reputation metric:

a) mean;
b) median;
c) 5% and 10% *trimmed mean* (see footnote 1 on page 1);
d) 5% and 10% *Winsorized mean* (see footnote 1 on page 1).

The tool will report the result of this analysis by generating one line of output with the format

$i,\delta_1,\delta_{2,\ldots,}\ \delta_6$

for each attack size *i* between *min* and *max*, where δ_1 to δ_6 represent the changes (accurate to two decimal places) induced by the attack in the overall product ratings for the six reputation metrics.

Once you completed the implementation of the missing methods, carry out the tasks below:

(i) Use your tool to analyse the impact of self-promoting and slandering attacks comprising between 5-200 fake ratings on the products with the genuine ratings given by

```
5,4,30,200,706,60
5,49,686,215,43,7
```

An input file (`input1.txt`) encoding the first attack you are required to analyse is provided on the module VLE.

(ii) Plots these data generated by the tool using Microsoft Excel (or similar software).

(iii) Use the results of your analyses to suggest the reputation metric or metrics that should be used by an online retailer that expects product rating patterns similar to those in (ii) for the majority of its products, in each of the following scenarios:
- The retailer's reputation system is affected by self-promoting and slandering attacks of small magnitude (0.5-1% fake customer ratings)
- The retailer's reputation system is affected by self-promoting and slandering attacks of medium magnitude (2-4% fake ratings)
- The retailer's reputation system is affected by self-promoting and slandering attacks of large magnitude (10-20% fake ratings).

(iv) (optional) Which criterion or criteria other than resistance to attacks should be considered when choosing the reputation metric for the type of reputation systems from this question? To what extent are these additional criteria satisfied by the reputation metrics from part (ii) of the question?