On the Evaluation of a Cluster-based Reputation Assessment Mechanism for Carpooling Applications

Emmanouil Mastorakis[1], Athanasios I. Salamanis[2], Dionysios D. Kehagias[2] and Dimitrios Tzovaras[2]
[1]*Aristotle University of Thessaloniki, School of Electrical & Computer Engineering, Thessaloniki, Greece*
[2]*Centre for Research & Technology Hellas, Information Technologies Institute, Thessaloniki, Greece*

Keywords: Carpooling, Reputation System, Clustering, Robustness, Attacks.

Abstract: Carpooling is a mobility concept that appears to be the answer when it comes to challenges in urban mobility derived by population growth. In carpooling, the same amount of people move with fewer vehicles leading to reduced traffic congestion and consequently to less CO_2 emissions, fuel consumption and drivers frustration. However, there has always been scepticism around carpooling due to the inherent mistrust between drivers and passengers. In recent years, some reputation systems have been proposed to reduce the impact of mistrust on carpooling applications. Among them, the work of Salamanis et al. (Salamanis, 2018), in which a reputation assessment mechanism based on clustering users travel preferences, was introduced. In this paper, we provide an extended version of the previous mechanism and we thoroughly evaluate its robustness in relation with different types of malicious attacks and clustering algorithms. In addition, we compare our mechanism with a benchmarking reputation system that utilizes the simple arithmetic mean to calculate reputation values based on users ratings. The evaluation results indicate that the extended reputation assessment mechanism exhibits more robust behavior compared to the benchmarking system in all types of attacks when using the hierarchical clustering algorithm.

1 INTRODUCTION

Public transportation services are ubiquitous in all the cities of the world. Using public transportation is environmentally friendly way of mobility, however it is not the perfect solution for coping with the increased traffic congestion. It might help reducing the number of vehicles on the road, but it does not necessarily reduce the number of miles driven, and it definitely does not reduce the amount of pollution generated by gas engines. "Smart cities" of tomorrow should make city living more efficient, more ecologically aware, and healthier. One of the main ways, in which that goal could be accomplished is through innovations designed to reduce traffic. Vehicular traffic should be considered as one of the greatest challenges facing cities all over the world and carpooling seems to be an alternative solution to this problem.

There is a clear socio-psychological barrier to the concept of carpooling, because people have to share their vehicles and travel with strangers. One way to mitigate this general sense of unease is the use of reputation systems. Reputation systems allow users to rate each other based on their activity in an online community in order to build trust through reputation. With the popularity of online communities for shopping, advice, and exchange of other important information, reputation systems are becoming vitally important to the online experience.

In recent years, several reputation systems have been emerged in the context of carpooling applications. Salamanis et al. (Salamanis, 2019) introduced a reputation assessment mechanism, which exploits the idea of grouping users based on their travel preferences. In particular, the user ratings are fine-tuned, through a weighted average scheme, based on the groups to which users belong. Thus, ratings between users that belong to the same or close groups receive more attention than others that belong to a distant group. The performance of such a system is affected by the choice of the appropriate clustering algorithm. Additionally, this system may behave differently (in terms of robustness) in different types of malicious attacks. In this paper, we introduce a variant of the reputation assessment mechanism proposed by Salamanis et al. (hereafter referred to as **C**luster-based **R**eputation **S**ystem (CRS)), which presents more robust behavior against different types of malicious attacks. In particular, we

Mastorakis, E., Salamanis, A., Kehagias, D. and Tzovaras, D.
On the Evaluation of a Cluster-based Reputation Assessment Mechanism for Carpooling Applications.
DOI: 10.5220/0007796602320241
In *Proceedings of the 5th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2019)*, pages 232-241
ISBN: 978-989-758-374-2

thoroughly evaluated the robustness of CRS with respect to different types of malicious attacks and clustering algorithms.

In particular, the types of malicious attacks evaluated in this work are:

1. Slandering Attack, in which a coalition of attackers provides false ratings to the system in order to lower the reputation of a victim user.
2. "Inverse" Slandering attack, which operates similarly to the slandering attack, but its objective, is to increase the reputation of a specific user.

Moreover, three different clustering algorithms were evaluated:

1. Hierarchical clustering
2. Mean shift
3. Affinity propagation

This evaluation process led to the conclusion that substituting the modified version of the k-means clustering algorithm of the original CRS with the hierarchical clustering algorithm results in a variant of the CRS (entitled **H**ierarchical **C**luster-based **R**eputation **S**ystem (HCRS)) which presents more robust behavior compared to the original CRS in both types of the evaluated attacks.

The rest of the paper is organized as follows. Section 2 reviews some of the current state-of-the-art reputation systems for several fields that are used in several online communities either in a business context or for research purposes. Section 3 introduces the idea of clustering users based on their travel preferences and presents the clustering algorithms used in this work to evaluate CRS. Section 4 introduces the reputation estimation process of the CRS model and benchmarking model. Section 5 describes in detail the evaluation approach adopted in this work as well as the findings of this process. Finally, Section 6 concludes the paper, reviewing the main contributions and suggesting future directions.

2 RELATED WORK

Reputation systems represent a significant trend in decision-making support for internet-mediated services. Common uses of these systems can be found in different types of online communities like e-commerce, social news, programming, wikis, Q&A, academia etc. This section reviews some of the state-of-the-art reputation assessment systems implemented either for commercial or research purposes.

E-commerce applications were among the first sectors for which reputation systems have been developed. For instance, Bizrate Insights (Bizrate Insights, 2019) is reputation system that provides services to both businesses and consumers. Consumers have access to ratings and reviews from verified buyers that help them to their purchase decisions. Additionally, TripAdvisor (Trip Advisor, 2019) utilizes a popularity-ranking algorithm, which is based on the quality and quantity of reviews that a business receives from users, along with the consistency of these reviews over time. Another application area in which reputation mechanisms are used is Q&A. Quora (Quora, 2019) is a Q&A website where questions are asked, answered, edited and organized by its community of users in the form of opinions. By combining social voting with sophisticated ranking algorithms, Quora enables users to better judge other users reputation and promote high quality content. Moreover, reputation systems are used for hospitality and social networking services, like CouchSurfing and Airbnb. In these services, members set up an online identity, rate other member by leaving comments for them, and develop a reputation that can help them be selected as an accommodation option.

In addition to the purely commercial online reputation assessment systems, several novel reputation systems have been developed in the context of research projects. For example, Bag et al. (Bag, 2018) proposed a privacy-aware decentralized reputation system for computing the personalized global reputation of a business entity by considering the trust scores from a set of trusted participants, without disclosing identities of participants in the trusted set and their trust-scores. Moreover, Gaoa and Zhouab (Gaoa, 2017) introduced an iterative group-based ranking method for evaluating user reputation in online rating systems. Reputation of users was calculated based on the weighted sizes of the user rating groups after grouping all users by their rating similarities. In this way, the high reputation users' ratings had larger weights in dominating the corresponding user rating groups.

In another approach, Lin et al. (Lin, 2018) proposed a reputation mechanism to enhance data trustworthiness (DTRM) for high-performance cloud computing. In this work, the sensitivity-level based data category, Metagraph-theory-based user group division, and reputation-transferring methods were integrated into the reputation query and evaluation process. Furthermore, Gusmini et al. (Gusmini, 2017) introduced a new comprehensive model and a multi-layer architecture for reputation evaluation

aimed to assess quality of Volunteered Geographic Information (VGI) content. Evaluation was based on reputation scores that summarized users' experiences with the specific content. Finally, Pera et al. (Pera, 2016) investigated whether self-storytelling is a powerful predictor of personal reputation. In this study, a qualitative–quantitative approach was adopted to investigate the meanings and stories contained in personal profile descriptions and their relation with reputation.

In recent years, the problem of mistrust in carpooling and ridesharing applications and the corresponding mitigation strategy through the use of reputation systems, has gained a lot of attention. For example, Montes et al. (Montes, 2018) introduced Teranga Go!, a carpooling platform integrating an intelligent decision support system. In this platform, participants of a carpooling ride act as experts that assess the driver aptitudes and determine, together with the history of the driver, a linguistic value for the driver's karma, which represents the collective opinion of people that have travelled with the driver. Moreover, Ferrer et al. (Ferrer, 2016) introduced a fully decentralized P2P ridesharing management network that avoids centralized ride-matching agencies along with a decentralized reputation management protocol that brings trust among peers.

3 TRAVEL PREFERENCES CLUSTERING

Travel preferences of the users include information that may be useful for the estimation of their reputation, and hence they are taken into account in the reputation assessment process. Table 1 presents the travel preferences that were taken into account by the CRS model. These travel preferences include several types of information like the preferred travel modes (e.g. bus, car, etc.) and the gender of the driver. Most of these travel preferences take values from a predefined set, whereas three of them, namely the maximum transfers, the maximum cost and the maximum walk distance, take arbitrary real values specified by the users.

3.1 Clustering Algorithms

The main objective of this work is to evaluate the robustness of the CRS in relation with different clustering algorithms. The original CRS model utilized a variant of the k-means algorithm to perform clustering. In this paper, we evaluated the following three additional clustering algorithms:

1. Hierarchical clustering
2. Mean shift
3. Affinity propagation

Table 1: User travel preferences.

Parameter	Values
Maximum Transfers	User specified
Maximum Cost (e)	User specified
Maximum Walk Distance	User specified
GPS Tracking	Yes, No
Travel Modes	Bus, Carpooling, Feet, Metro, Rail, Tram
Optimize Travel Solutions	By Price, Comfort, Speed, Safety, Distance
Car-pooler Gender	Male, Female
Car-pooler Age Range	[18; 30), [30; 40), [40; 50), [50; 60)
Impaired	Visual, Hearing, Elderly, Wheelchair
Smoking	Yes, No
Food	Yes, No
Music	Yes, No
Pets	Yes, No
Luggage	Yes, No

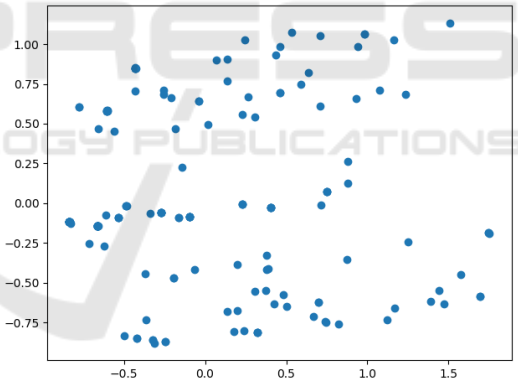

Figure 1: User travel preferences used by the CRS model.

As can be perceived by Table 1, the travel preferences of the users can be represented as multidimensional vectors of mixed-type data. For instance, the maximum cost is a continuous variable, as opposed to the car-pooler gender, which is a nominal variable, or food, which is a binary variable. For facilitating the clustering process, we encoded the categorical variables as numerical using one hot encoding. The resulting multidimensional vectors were depicted in the two-dimensional space, after performing principal component analysis (PCA) on them and keeping only the two principal components. This visualization is shown in Figure 1.

The clustering algorithms evaluated in this paper are presented in the following subsections.

3.1.1 Hierarchical Clustering

Hierarchical clustering is a general family of clustering algorithms that build nested clusters by merging or splitting them successively. This hierarchy of clusters is usually represented as a tree (or dendrogram). The root of the tree is the unique cluster that gathers all the samples, while the leaves are the clusters with only one sample. In Figure 2, the dendrogram of the travel preferences clustering is presented. We performed hierarchical clustering using a bottom up approach: each observation starts in its own cluster, and clusters are successively merged together (i.e. agglomerative clustering). The Euclidean distance was used as distance metric:

$$\|a - b\|_2 = \sqrt{\sum_i (a_i - b_i)^2} \quad (1)$$

The Ward's criterion was used for the merge strategy. Ward's criterion minimizes the sum of squared differences within all clusters. It is a variance-minimizing approach and in this sense is similar to the k-means objective function but tackled with an agglomerative hierarchical approach. Ward's minimum variance method can be implemented by the Lance-Williams formula:

$$d_{(ij)k} = a_i d_{ik} + a_i d_{jk} + \beta d_{ij} + \gamma \left| d_{ik} - d_{jk} \right| \quad (2)$$

where α_i, β, and γ are given by the following equations:

$$a_i = \frac{n_i + n_k}{n_i + n_k + n_j} \quad (3)$$

$$\beta = -\frac{n_k}{n_i + n_k + n_j} \quad (4)$$

$$\gamma = 0 \quad (5)$$

For example, for disjoint clusters C_i, C_j, and C_k with sizes n_i, n_j, and n_k respectively:

$$d(C_i \cup C_j, C_k) = \frac{n_i + n_k}{n_i + n_k + n_j} d(C_i, C_k) + \frac{n_j + n_k}{n_i + n_k + n_j} d(C_j, C_k) - \frac{n_k}{n_i + n_k + n_j} d(C_i, C_j) \quad (6)$$

We have chosen three (3) as the number of our clusters. In Figure 3, we can view the three clusters created by visualizing our multi-dimensional data on two-dimensional space using PCA. In addition, in Table 2 the distances of the cluster centroids are presented. It should be noted that the hierarchical clustering algorithm does not return the centroids of the clusters, and therefore these centroid were calculated by us, as the average of the vectors that belong to each cluster.

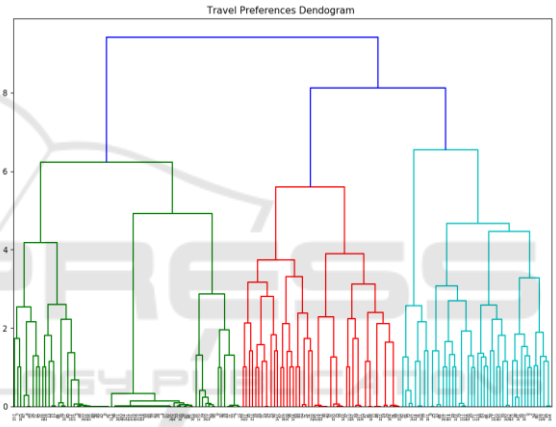

Figure 2: Travel preference dendrogram.

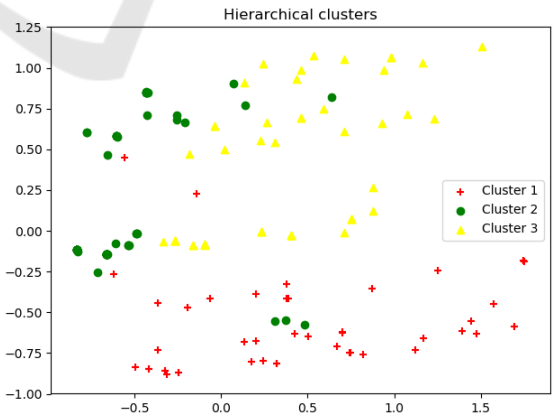

Figure 3: Hierarchical clustering results with three clusters.

Table 2: Distances between centroids in hierarchical clustering.

CLUSTER	1	2	3
1	0	1.286424	1.154078
2	1.286424	0	1.065389
3	1.154078	1.065389	0

3.1.2 Mean Shift

Mean shift is a non-parametric feature-space analysis technique for locating the maxima of a density function, which can be applied in the field of cluster analysis. Let a kernel function, $K(x_i - x)$, be given, which determines the weight of nearby points for re-estimation of the mean. The weighted mean of the density in the window determined by *K* is:

$$m(x) = \frac{\sum x_i \in N(x) K\,(x_i - x) x_i}{\sum x_i \in N(x) K\,(x_i - x)} \quad (7)$$

where *N(x)* is the neighbourhood of *x*, and $m(x) - x$ is called *mean shift*. A candidate centroid x_i in iteration *t*, is updated in the following iteration *t+1* based on the following equation:

$$x_i^{t+1} = m(x_i^t) \quad (8)$$

Equation (8) is compute for each centroid x_i. We applied the mean shift clustering on travel preferences data resulting in three clusters as shown in Figure 4. Table 3 presents the distances between the clusters centroids.

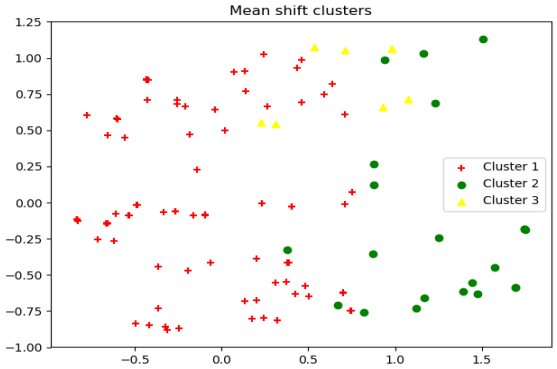

Figure 4: Mean shift clustering results with three clusters.

In contrast with the hierarchical clustering algorithm, the mean shift algorithm does return the coordinates of the clusters centroids, and therefore no post-processing for computing them is required.

Table 3: Distances between centroids in mean shift clustering.

CLUSTER	1	2	3
1	0	1.54581797	1.46446172
2	1.54581797	0	1.41064452
3	1.46446172	1.41064452	0

3.1.3 Affinity Propagation

Affinity propagation is a clustering algorithm that creates clusters based on the concept of "message passing" between data points. Similar to the *k*-medoids algorithm, affinity propagation identifies representative member of the input set called *exemplars*.

The messages sent between pairs of samples from the input set, represent the suitability for one sample to be the exemplar of the other, which is updated in response to the values from other pairs. This updating process takes place iteratively until convergence is reached, at which point the final exemplars are chosen and hence the final clustering is given. The messages sent between points belong to one of two categories. The first is the *responsibility r(i,k)*, which is the accumulated evidence that sample *k*, should be the exemplar for sample *i*. The second is the *availability a(i,k)* which is the accumulated evidence that sample *i* should choose sample *k* to be its exemplar, and considers the preference of all other samples *j* for *k* as an exemplar. In this way, exemplars are chosen by samples if they are:

1. Similar enough to many samples
2. Chosen by many samples to be representative of themselves

The responsibility of a sample *k* to be the exemplar of sample *i* is given by the following equation:

$$r(i,k) \leftarrow s(i,k) - \max_{k' \neq k}[a(i,k') + s(i,k')] \quad (9)$$

where *s(i, k)* is the similarity between samples *i* and *k*. The availability of sample *k* to be the exemplar of sample *i* is then given by the following equation:

$$a(i,k) \leftarrow \min[0, r(k,k) + \sum r(i',k)] \quad (10)$$

At the beginning, all values of *r* and *a* are set to zero. Then, their values are updated using the equations (9) and (10), until convergence is reached. In order to avoid numerical oscillations when

updating the messages, the damping factor λ is introduced:

$$r_{t+1}(i,k) = \lambda r_t(i,k) + (1-\lambda) r_{t+1}(i,k) \quad (11)$$

$$a_{t+1}(i,k) = \lambda a_t(i,k) + (1-\lambda) a_{t+1}(i,k) \quad (12)$$

where t indicates the iteration times.

As for the other clustering algorithms, we applied the affinity propagation algorithm to the travel preferences data, and the resulting three clusters are shown in Figure 5. In addition, Table 4 presents the distances between the clusters centroids.

Table 4: Distances between centroids in affinity propagation clustering.

CLUSTER	1	2	3
1	0	0.98990068	1.27398554
2	0.98990068	0	1.66263907
3	1.27398554	1.66263907	0

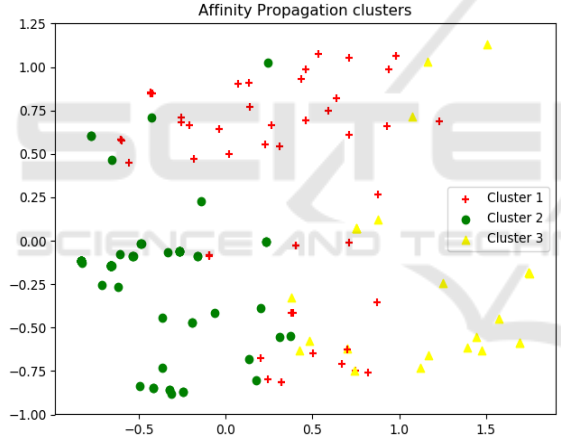

Figure 5: Affinity propagation clustering results with three clusters.

Similarly, to the mean shift algorithm, the affinity propagation algorithm computes the coordinates of the clusters centroids and therefore no post-processing for computing them is required.

4 REPUTATION ESTIMATION STRATEGIES

In this section, the evaluated reputation estimation strategies are briefly presented. In particular, we describe the reputation estimation process of the CRS model as well as the one of the simple-average reputation system.

The reputation estimation process for either model takes place at the end of each ride. Passengers rate the driver and the driver rates the passengers, no rating among passengers is allowed. These ratings are the main inputs of the reputation estimation models.

A user can rate another user based on N_f different rating variables, which describe different aspects of the journey experience. The value of each rating is an integer number between 1 and 5. For the purposes of this work, N_f was set to 4. A rating vector submitted to a reputation assessment system is $[f_1 \quad \cdots \quad f_{N_f}]^T$, where f_i is the value of feature i. From each rating vector, an average rating value R_{av}.

Finally, in our experiments, we are comparing the CRS model with the simple-average reputation system for two types of malicious attacks, namely the slandering and the inverse slandering.

4.1 CRS Reputation Estimation Strategy

In the CRS reputation estimation process, the total rating for a target user equals to the weighted average of the average ratings $R_{i,av}$ of the other co-passengers and it is given by the following equation:

$$R_{tot} = \sum_{i=1}^{N_r} w_i * R_{i,av} \quad (13)$$

where w_i is the weight of the user u_i and N_r is the total number of users. The weight w_i of a user u_i is defined by the cluster to which (s)he belongs. In particular, the following statements apply:

- Each user u_i belongs to a cluster c_i
- The distance between the cluster c_i of the rater user u_i and the target user u is d_i
- The weight w_i of the rater u_i is defined by the following equation:

$$w_i = 1 - \frac{d_i}{d_{max}} \quad (14)$$

where d_{max} is the distance between the cluster of the target user u and the most distant cluster
- The weights sum up to 1

The distances between the clusters are computed beforehand using the Euclidean distance metric and

are stored in a symmetric square matrix D_{kxk} (as shown in Table 2-4). In this matrix, the element d_{ij} is equal to the distance between clusters c_i and c_j. For each user u_i, the distance d_i of the cluster c_i from the cluster c of the target user can take values in the interval [0, d_{max}], therefore the weight w_i takes values in the interval [0, 1]. If a rater belongs to the most distant cluster from the evaluated user ($d_i = d_{max} \Rightarrow w_i = 0$), then his/her rating is not taken into account in the reputation estimation process.

After the total rating R_{tot} is calculated, it is transformed into positive or negative feedback (f) based on the following inequalities:

$$f \geq 0, if\ R_{tot} \geq 2.5 \quad (15)$$

$$f \leq 0, if\ R_{tot} \leq 2.5 \quad (16)$$

This feedback is used in order to update the reputation value of the target user. The CRS utilizes the beta reputation system (BRS) to model the reputation of a user. In particular, the reputation $R(u)$ of a user u in the CRS model is expressed as the expectation of the beta distribution as follows:

$$R(u) = \frac{r + 1}{r + s + 2} \quad (17)$$

where,

$$r_{new} = r_{old} + 1, f \geq 0 \quad (18)$$

$$s_{new} = s_{old} + 1, f < 0 \quad (19)$$

If no past feedbacks exist for a target user (i.e. $r = 0$ and $s = 0$), the value of $R(u)$ is 0.5.

4.2 Simple-average Reputation System

In the simple-average reputation system, the total rating for a target user equals to the average of the average ratings $R_{i,av}$ of the other co-passengers and it is given by the following equation:

$$R_{tot} = \frac{1}{N_r} \sum_{i=1}^{N_r} R_{i,av} \quad (20)$$

The reputation $R(u)$ of a target user u is then calculated by the following equation:

$$R(u) = \frac{n * R(u)_{prev} + R_{tot}}{n + 1} \quad (21)$$

where $R(u)_{prev}$ is the reputation of the user u before the last ride, and n is the number of previous rides.

5 EVALUATION

In this section, the framework and the results of the evaluation process of the CRS are presented. The main objective of this process was to evaluate the robustness the CRS in relation with different types of malicious attacks and clustering algorithms. In particular, two different types of attacks, namely the slandering attack and the "inverse" slandering attack (the one that tries to achieve the opposite result from the former) were evaluated. Also, the clustering algorithms presented in Section 3 were used by the CRS.

5.1 Evaluation Framework

This subsection describes the framework of the evaluation process of the CRS model. This framework includes the evaluation scenarios, the choice of malicious and non-malicious users, the rating strategies of the malicious and non-malicious users and the types of malicious attacks.

Due to lack of real, user-provided ratings, until the time that this manuscript was written, a simulation process with artificially generated user ratings was designed and implemented. This process is based on the scenario in which a coalition of malicious users is organized in order to perform a specific type of attack on the reputation system of a carpooling application. The malicious users are considered as legitimate users of the application (insider attackers).

As already mentioned, the user ratings are submitted to a reputation system at the end of each ride. Therefore, each simulation cycle represents a hypothetical ride in which the driver is always the target user, and among the passengers there are malicious and non-malicious users. At the end of this hypothetical ride, the passengers provide ratings to the reputation system regarding the driver. Initially, the probability of a passenger being a malicious user follows a uniform distribution. However, in each simulation cycle, the number of malicious users in the system increases and hence the probability of having a malicious user among the passengers of the hypothetical ride increases as well.

Attacks against reputation systems are classified based on the goals and methods of the attacker. The two types of attacks evaluated in this paper are the slandering attack and the "inverse" slandering

attack. In the former case, the attacker reports false data to *lower* the reputation of the target user. Therefore, such a malicious user submits rating vectors containing the smallest possible values, i.e. $[1 \quad ... \quad 1]^T$. In the latter case, the attacker tries to falsely *increase* the reputation of the target user by submitting rating vectors containing the largest possible values, i.e. $[5 \quad ... \quad 5]^T$. Finally, the non-malicious users are considered to rate in a completely random way, meaning that they submit feature vectors whose values are drawn from the uniform distribution in the interval [1, 5].

The simulation process was applied on both the CRS and the simple-average reputation system with the objective of evaluating their robustness. For the CRS model, when the slandering attack is concerned, the initial values of the previous positive (r) and negative (s) ratings were set to 3 and 1, respectively. This means that the initial reputation of the target is 0.8 (the result of equation (21) multiplied by 5). Also, all three clustering algorithms presented in Section 3, were used separately by the CRS model. On the other hand, when the inverse slandering attack is concerned, the respective values were 1 and 3, resulting in the initial reputation value of the target user to be 0.4. Additionally, for the simple-average reputation system, the number of past ratings (n) was set to 2, and the initial reputation value of the target user to 4 for the slandering attack and to 2 for the inverse slandering attack respectively. Finally, the percentage of malicious users in each reputation system (i.e. the penetration rate) starts from 0% and goes up to 100%, increasing in each cycle by 5%.

5.2 Results

Figures 6-8 present the reputation curves of the target user against the penetration rate of malicious users for the simple-average reputation system (blue line) and the CRS model (black line) for all three clustering algorithms.

In all three cases, the CRS model presents more robust behavior compared to the simple-average reputation system. In particular, the CRS model with hierarchical clustering, for up to 35% penetration rate, keeps the reputation value almost unchanged. When the penetration rate increases more than 35%, the reputation of the target user decreases in a steady pace. On the other hand, using the simple-average reputation system, the reputation value presents a downturn from the beginning of the simulation until the penetration rate reaches 50%, and then it is stabilized. For all penetration rate values, the reputation derived by the CRS model is always

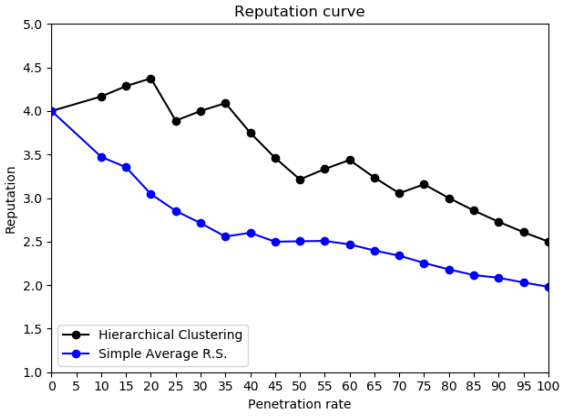

Figure 6: Reputation curves derived from the CRS model with hierarchical clustering and the simple-average reputation system in the slandering attack.

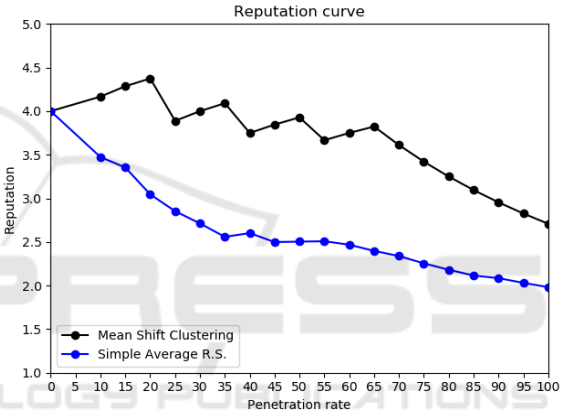

Figure 7: Reputation curves derived from the CRS model with mean shift clustering and the simple-average reputation system in the slandering attack.

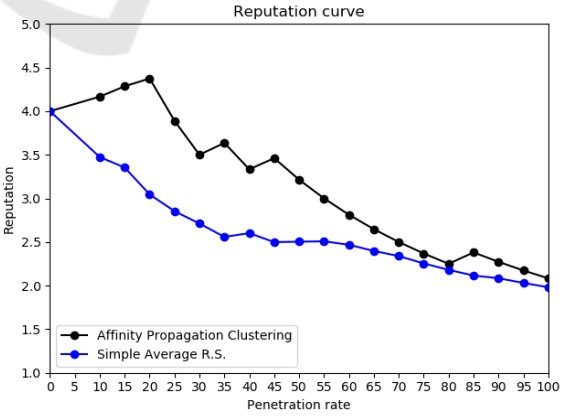

Figure 8: Reputation curves derived from the CRS model with affinity propagation clustering and the simple-average reputation system in the slandering attack.

higher than the reputation derived by the simple-average reputation system. Also, when the penetration rate is 100% (i.e. all users are malicious), the reputation value derived by the simple-average system has been decreased by approximately 50% compared to the initial reputation value, while the corresponding percentage for the CRS model is approximately 40%. These results are similar for both the mean shift and the affinity propagation clustering algorithms. Concerning which clustering algorithm leads to a more robust behavior for the CRS model in this type of attack, we can say that the hierarchical and the mean shift algorithms yield better results than the affinity propagation algorithm and therefore they can be considered as more appropriate for use with the CRS model.

Concerning the inverse slandering attack, the reputation curves are shown in Figures 9-11. In this case, there is a clear differentiation between the three clustering algorithms. In particular, in Figure 9, the reputation curve derived by the simple-average reputation system is constantly rising, while the reputation curve derived by the CRS model coupled with the hierarchical clustering algorithm presents fluctuations depending on the penetration rate. For a penetration rate from 0% to 35% the two reputation curves are very close, but when the penetration rate becomes greater than 35%, the values of the CRS model reputation curve become smaller than those of the simple-average reputation system. Moreover, when the penetration rate is 100%, the reputation value derived by the simple-average reputation system has been increased by approximately 100% compared to the initial reputation value, while the corresponding percentage for the CRS model is approximately 50%.

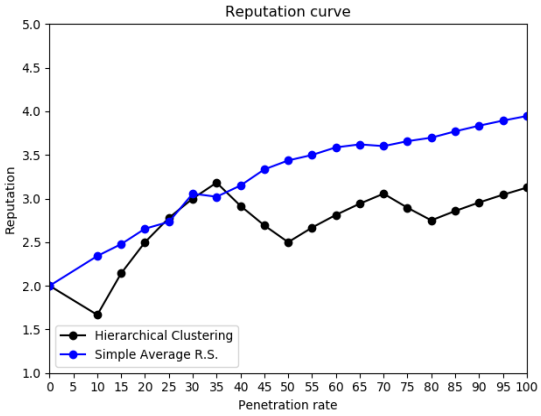

Figure 9: Reputation curves derived from the CRS model with hierarchical clustering and the simple-average reputation system in the inverse slandering attack.

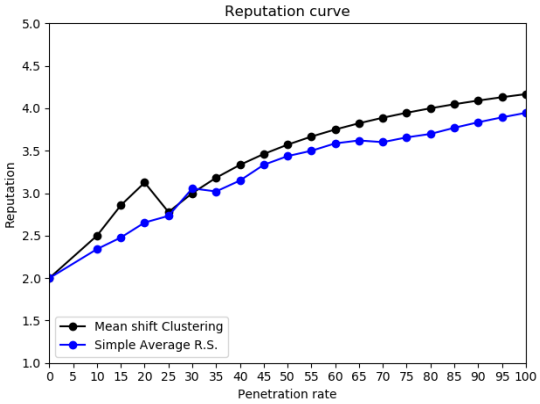

Figure 10: Reputation curves derived from the CRS model with mean shift clustering and the simple-average reputation system in the inverse slandering attack.

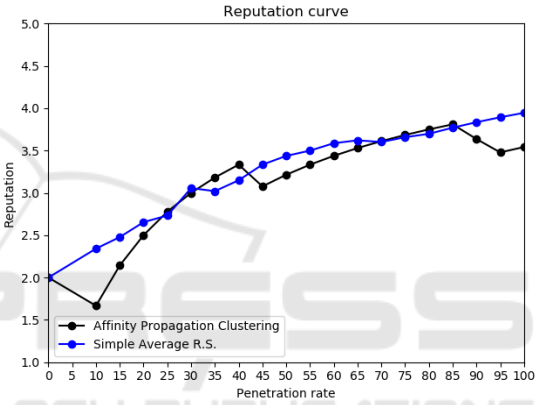

Figure 11: Reputation curves derived from the CRS model with affinity propagation clustering and the simple-average reputation system in the inverse slandering attack.

When the mean shift clustering algorithm is utilized by the CRS model, its robustness is worse than the robustness of the simple-average reputation system. In this case, both reputation curves are constantly increasing, and when the penetration rate becomes 100%, they both produce reputation values which are increased by approximately 100% compared to the initial reputation value.

Finally, when the affinity propagation clustering algorithm is used, the robustness of the CRS model is comparable with the robustness of the simple-average reputation system. For some penetration rate values the reputation values of the CRS model are higher than those of the simple-average reputation system, while for other values the opposite is true. When the penetration rate is 100%, the reputation value derived by the simple-average reputation system has been increased by approximately 100%, while the corresponding percentage for the CRS model is 87.5%. Based on these findings, we can say

that in the case of inverse slandering attack, the hierarchical clustering algorithm is more appropriate for the CRS model.

This result, in conjunction with the corresponding result for the case of the slandering attack, leads to the conclusion that substituting the modified version of the k-means clustering algorithm of the original CRS with the hierarchical clustering algorithm results in a variant of the CRS model (namely the HCRS model), which presents more robust behavior compared to the original CRS in both types of the evaluated attacks. It should be noted here that the above results should be re-verified on real, empirical data in order to prove their validity. Two main directions will be followed (as future work) towards this objective, namely to integrate the HCRS model into a carpooling application which will be handed to real users that will interact with the system by providing real ratings, and to use open ratings datasets (possibly from different domains).

6 CONCLUSIONS

In this paper, we thoroughly evaluated the robustness of CRS model in relation with different types of malicious attacks and clustering algorithms, and also we compared it with the simple-average reputation assessment system. In particular, two different types of malicious attacks, namely the slandering and the inverse slandering attacks, as well as three clustering algorithms, namely the hierarchical, the mean shift and the affinity propagation, were evaluated. Based on the findings of this work, we introduced HCRS, a variant of the original CRS model which utilizes the hierarchical clustering algorithm to create user groups (instead of the modified k-means algorithm used in the original CRS model). The evaluation results indicate that the HCRS model presents better performance, in terms of robustness, compared to the simple-average reputation system in both types of the evaluated attacks.

Future work involves experimenting with additional clustering algorithms and malicious attacks, and also setting up a formal process for the robustness evaluation of reputation assessment systems from different domains. Furthermore, another direction for future research is the investigation of the correlations between the quality of the reviews submitted to the system, and the robustness of the proposed reputation model.

ACKNOWLEDGEMENTS

The work presented in this paper is partially funded by the European Union's Horizon2020 Research and Innovation Programme through MyCorridor project under Grant Agreement No.723384.

REFERENCES

Bag, S., Azad, M. A., Hao, F., 2018. A privacy-aware decentralized and personalized reputation system, *Computers & Security Volume 77, August 2018, Pages 514-530*

Bizrate Insights, 2019. Bizrate Insights. https://bizrateinsights.com/, [Online; accessed 27-January-2019].

Coush Surfing, 2019. CouchSurfing https://www.couchsurfing.com/ [Online; accessed 27-January-2019].

Ferrer, J. D., Martínez, S., Sánchez, D., Comas, J. S., 2017. Co-Utility: Self-Enforcing protocols for the mutual benefit of participants *Engineering Applications of Artificial Intelligence, Volume 59, March 2017, Pages 148-158*

Gusmini, M., Jabeurb, N., Karam, R., Melchiori, M., Renso, C., 2017. Reputation evaluation of georeferenced data for crowd-sensed applications, *Procedia Computer Science Volume 109, 2017, Pages 656-663*

Jian Gaoa, Tao Zhouab 2017. Evaluating user reputation in online rating systems via an iterative group-based ranking method, *Physica A: Statistical Mechanics and its Applications Volume 473, 1 May 2017, Pages 546-560*

Lin, H., Hub, J., Xu, C., Mac, J., Yua, M., 2018. DTRM: A new reputation mechanism to enhance data trustworthiness for high-performance cloud computing, *Future Generation Computer Systems Volume 83, June 2018, Pages 293-302*

Montes, R., Sanchez, A. M., Villar, P., Herrera, F., 2018. Teranga Go!: Carpooling Collaborative Consumption Community with multi-criteria hesitant fuzzy linguistic term set opinions to build confidence and trust, *Applied Soft Computing Volume 67, June 2018, Pages 941-952*

Pera, R., Viglia, G., Furlan, R., 2016. Who Am I? How Compelling Self-storytelling Builds Digital Personal Reputation, *Journal of Interactive Marketing Volume 35, August 2016, Pages 44-55*

Quora, 2019. Quora https://www.quora.com/ [Online; accessed 27-January-2019].

Salamanis, A., Kehagias, D. D., Tsoukalas, D., Tzovaras, D., 2018. Reputation assessment mechanism for carpooling applications based on clustering user travel preferences, *International Journal of Transportation Science and Technology*

Trip Advisor, 2019. TripAdvisor. https://www.tripadvisor.com.gr//, [Online; accessed 27-January-2019].