Data and Network Security project outline

1. Overview
   1. Identify the problem
   2. What are we trying to solve?
   3. Purpose of the paper
2. Previous work
   1. What other solutions have been proposed for the problem of anonymous release of data sources?
   2. What are potential issues that could arise from these approaches?
   3. What issues did the writers identify?
3. Our solution
   1. Cover both requirements:
      1. Policy
      2. Data
   2. Details and specifics of the model
   3. Test case implementation with data
      1. One solution I had in mind was taking the identifiers of records and clustering based on those identifiers. We Would then publish data with no identifier except for the cluster and run ANOVA and Tukey HSD to tell whether the groupings had statistically significant differences from the data grouped by actual identifier. We would simply replace the identifier of the dataset with a cluster number and do no aggregation to the data. We would then provide the characteristics of the clusters to the data users. For instance, cluster one is female with an average age of 65 and std dev of 5 years…The hope would be that we fail to reject the null for these cases.
   4. Potential attacks
      1. Outline potential pitfalls of the approach