Quasi-identifier values can impact the severity of a data leakage. For example, a leakage involving diseases can have different impacts on individuals at certain stages of their lives. A 96 year old might be disclosed to have HIV through a data breach. This individual may not be as severely impacted as a child under 10. The 96 may already know that they have the disease and have been living with it for a long period of time. The child under 10 may be aware, but would not like this information to be disclosed in their school. Using the regular L-Severity model, the model only takes into account severity of sensitive attributes. Non-sensitive attributes can also be a target of a data breach. The severity of quasi-identifiers and sensitive attributes must be maintained over time. If an employee responsible of maintaining this data does not update the attribute classification, an attacker can target non-sensitive attributes to remain under the radar. Tracking data coming in and out of a system can be beneficial, many times a breach is not realized until it is reported or made public. A DLP may be tracking when sensitive data is released or based on some measurement of severity. Often severity can involve many dimensions, but there are few publications that focus strictly on the data and the definitions around that. Vavilis et al. attached values to different sensitive attribute values and created a model to measure this severity called L-Severity. L-Severity considers a model that is catered only towards sensitive attributes. We propose a new model that places more emphasis on quasi-identifier attributes and non-sensitive attributes. Non-sensitive attributes are important because data may be unstructured. For example, JSON can have many properties associated with an object. These properties are dynamic and data can be added on without the end system realizing the addition or removal. The varying properties will give different records different levels of severity based on the number of attributes associated with that object. This research does not go into the validation of data. The analysis of the impact of DF was done comparing the algorithm that M-Score uses and a k-anonymity. It was found that there is not be a significant impact between M-Score and k-anonymity because the DF impact on severity is measured on the number of similar quasi identifiers. For example, if a table is conforming to k-anonymity the DF will remain constant. Using M-Score’s algorithm, this will remain the same, a table that is nonconforming, less than k, will have higher severity in both scenarios. Although DF is a good metric to use to measure privacy, it is too general. Similar to sensitive attributes, quasi-identifiers and non-sensitive attributes should be scored. The number of tuples in each unique quasi-identifier group will have a positive correlation with a defined score. Having a score can give a more defined measurement of the impact of the data on a person’s life. However, this value can be incorrect if not setup up by a domain expert in privacy. Having a score also allows more granular analysis. For example, a row may have higher sensitivity than others. This sensitivity will be better reflected with an accurate measurement in privacy and taking into consideration non-sensitive attributes. Setting parameters to each score classification can allow an analyst place more emphasis on one over the other. For example, a database with information about public figures can monitor data being retrieved and placing a higher emphasis on privacy. Distinguishing Factor is important because one record may stand out from the others.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Job | City | Sex | Age | Disease | Medication |
| Student | LA | Male | 100 | HIV | Vitamins |
| Student | LA | Male | 20 | Heart Attack | Aspirin |
| Student | LA | Male | 30 | Migraine | Paracetamol |
| Student | LA | Male | 40 | Hypertension | Aspirin |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Job | City | Sex | Age | Disease | Medication |
| Student | LA | Male | 10 | HIV | Vitamins |
| Student | LA | Male | 20 | Heart Attack | Aspirin |
| Student | LA | Male | 30 | Migraine | Paracetamol |
| Student | LA | Male | 40 | Hypertension | Aspirin |