The topic of the article reviewed was developing algorithms which made use of collaborative filtering data to infer what items a user reviewed. The algorithm looks at the items that are being recommended to a user and reverses the process to determine what items lead to that recommendation. This presents a privacy concern because information such as medical condition or political affiliation could be revealed by knowing what items were purchased. If this tactic was used by an insurance company, individuals could have their rates raised based on a condition they were not required to report.

The data used in the algorithms that were developed is publicly available from sites such as Amazon, Netflix, and Last.fm. The type of data provided by these systems may be item similarity lists, item-to-item covariances or relative popularity of items. The model also assumes that for some users, auxiliary information is available to the attacker about the user’s preferences. This information could come from 3rd party review sites or social media.

The attack model used by the researchers focuses on a passive attack which an attacker makes use of this publicly available data and does not create fake accounts. These fake accounts would be used to manipulate the recommendations and try to find a pattern. The attacker then monitors this data over a period of time. If the attacker notices an increase in the relation between the auxiliary and public information for an item, it can be inferred that item has been added by the user.

The main metrics that indicate the algorithm’s effectiveness are the yield and accuracy of the inferences made. The yield is the number of inferences per user per observation period. The accuracy is the percentage of those inferences that were correct. The researchers note that high yield and high accuracy are not needed to achieve damage with an attack of this type. A single item could reveal a very personal fact about an individual.

The researchers tested their algorithms against the real system mentioned above. The results when tested against Amazon showed a greater than 80% accuracy rate for the first 200 inferences made. The accuracy decrease as the number of inferences increases. This trend is the same for the other systems tested.
The researchers provided several items that could be used to mitigate this attack. These mitigations could be used by myself or others in future software projects that require some sort of collaborative filtering. The top two are limiting the length of related items lists and limit the speed or rate of data access. The lists that are publicly available from Last.fm may have more than 100 items for a user. This is a lot of information for an attacker to make use of. When it comes to limiting data access, one of the APIs used during the research allowed up to 5,000 queries per day. The algorithm was able to be used with only 500 to 2,500 queries per day per user.

This attack could potentially be very damaging to an individual’s reputation. However it is very time consuming and requires a lot of data to perform. An attacker must be very motivated in order to mount an attack such as this. An attack of this nature is probably not going to be used against the average individual unless large organizations such as insurance companies begin to use it.

**Bibliography**