Abstract

An individual’s healthcare data may be the most private information a person possesses. Current regulations, such as the Health Insurance Portability and Accountability Act of 1996 (HIPAA), safeguard patient data by assigning a sensitivity level to data items. However, this approach is limited when domain knowledge is used to infer additional patient data. In our research, we investigate privacy violations occurring when non-confidential patient data is combined with medical domain ontologies to disclose a patient’s protected health information (PHI).

We developed a framework that detects privacy violations and eliminates undesired inferences. Our inference channel removal process is based on controlling the release of the data items that lead to undesired inferences. These data items are either blocked from release or generalized to eliminate the disclosure of the PHI. We first developed an exhaustive framework to disrupt the undesired inferences, then improved on the methods using a heuristic-based approach. Our privacy model includes traditional security assessments (i.e., HIPAA) as well as considering safety and patient privacy preferences. We developed a graphic user interface that allows patients to control the release of their data. We also visualize the inferred data using the healthcare domain knowledge.