Privacy in GLAV Information Integration

Alan Nash[1]* and Alin Deutsch[2]**

[1] IBM Almaden Research Lab
`anash@ibm.us`
[2] University of California San Diego
`deutsch@cs.ucsd.edu`

Abstract. We define and study formal privacy guarantees for information integration systems, where sources are related to a public schema by mappings given by source-to-target dependencies which express inclusion of unions of conjunctive queries with equality. This generalizes previous privacy work in the global-as-view publishing scenario and covers local-as-view as well as combinations of the two.

We concentrate on *logical* security, where malicious users have the same level of access as legitimate users: they can issue queries against the global schema which are answered under "certain answers" semantics and then use unlimited computational power and external knowledge on the results of the queries to guess the result of a secret query ("the secret") on one or more of the sources, which are not directly accessible. We do not address issues of *physical* security, which include how to prevent users from gaining unauthorized access to the data.

We define both absolute guarantees: how safe is the secret? and relative guarantees: how much of the secret is additionally disclosed when the mapping is extended, for example to allow new data sources or new relationships between an existing data source and the global schema? We provide algorithms for checking whether these guarantees hold and undecidability results for related, stronger guarantees.

A copy of this technical report can be obtained by sending a request to `deutsch@cs.ucsd.edu`.

---

* Work performed while the author was at the University of California, San Diego.

** Funded by an Alfred P. Sloan fellowship and by NSF CAREER award IIS-0347968.