Data Management Plan

Expected Data

The expected data to be generated during the course of this project are:

1. Technical papers describing our research.
2. Code for the proposed phishing page detection browser plug-in.
3. A dataset of logos and image maniplations.
4. Data from user studies.
5. Data from the RIT phishing training.
6. Security incident reports and logs of compromised user account statistics from the RIT IT department.
7. Course materials, such as lectures and slides, homework assignments, and projects.

Data Retention and Storage

All data generated by this project will be retained for a minimum of three years after the conclusion of the award, including data that are not specifically disseminated as described in the following section. We will use file storage services provided by RIT and department IT departments to ensure secure preservation of data during the retention period.

Data collected in user studies, RIT phishing training, and security incident reports and logs will be encrypted to prevent exposures. In our user studies, no identifiers will be kept and only basic demographic information will be associated with the data.

Data Dissemination

Technical papers generated during the course of the project will be published in academic journals and conference proceedings. Papers published in venues without archival proceedings, as well as technical reports not otherwise published, will be disseminated via arXiv.

The phishing page detection browser plug-in will be released under an Open Source License (as defined by the Open Source Initiative) and will be published in an open source project repository, such as SourceForge, GitHub, or Google code project hosting.

Data produced from the proposed research activities will be made available either publicly on PI Wright's website or, for sensitive data, available upon request by other researchers. Exceptions to this are the RIT phishing training data, which cannot be shared due to RIT policies on students and employees, and security incident reports and logs, which are sensitive for RIT's security posture. For these data sets, we will work with RIT's Global Risk Management team to find ways for other researchers to access the data, perhaps in conjunction with the PIs.

Shared data will be stored in standard formats, such as text files and XML, whenever it is possible to do so. If binary formats are necessary (e.g., for saving space), utilities for converting this data into human-readable formats will be made available.

Curricular materials such as lecture notes and slides will be made available via PI Wright's website.