



Minimizing your long-term security and privacy risks online . . .

Assistant Professors Elena Zheleva and Chris Kanich receive an \$800K NSF grant "SaTC: CORE: Medium: Collaborative: Enabling Long-Term Security and Privacy through Retrospective Data Management"

\$800K grant aims to reduce long-term online data risks

Assistant Professors Elena Zheleva and Chris Kanich were awarded an \$800K NSF grant entitled "SaTC: CORE: Medium: Collaborative: Enabling Long-Term Security and Privacy through Retrospective Data Management" that runs from August 1, 2018 until July 31, 2022.

Both Zheleva and Kanich share an interest in online data and privacy as their main areas of research.

This is part of an overall \$1.22 million collaborative project with Professor Blase Ur in the University of Chicago Computer Science Department.

"SaTC: CORE: Medium: Collaborative: Enabling Long-Term Security and Privacy through Retrospective Data Management"

Chris Kanich, PhD & Elena Zheleva, PhD

ABSTRACT

Online data storage, everything from past conversations to tax returns to playdate invitations, may be retained at full fidelity for years or decades. Although the data being saved in online archives does not change, the personal and social contexts surrounding them do. Those life changes may necessitate changing or deleting stored data but, unfortunately, the vast quantity of data in users' online archives makes manual management infeasible. The goal of this project is to develop methods and tools that enable users to manage the data they have accumulated over many years, leveraging user-centered design and machine learning to partially automate the process. These tools will enable a better understanding of retrospective privacy in the context of modern long-lived online archives. They will also empower users to more effectively manage the risks embedded in these archives. The findings shared with the research community will advance discovery beyond this project.

The understanding of user conceptualizations of security and privacy over time, as contexts change, has been sty-

mied by a lack of broad, carefully collected datasets within this domain. This project will collect anonymized datasets, with users' permission, that enable further research in this area. The team is conducting one of the first longitudinal studies of how desired security and privacy decisions change over time. The project will also gather qualitative insights about users' perceptions of risk and utility for long-term data, as well as the acceptability of retrospective management mechanisms. Furthermore, it is not currently understood how temporality impacts the use of machine learning techniques for privacy, nor how to capture concept drift to ensure that latent threats can be identified within immense archives. These tasks require new machine learning approaches and predictive models that can both account for the temporal dimension and minimize user burden when automating archive management. Finally, the project will design and implement novel user-centered interfaces that address the currently unmet need of helping users efficiently minimize security and privacy risks in their large, long-term online archives.