

EMPIRICAL ANALYSIS OF DATA BREACH LITIGATION

Research-in-Progress

Sasha Romanosky
Heinz College,
Carnegie Mellon University
sromanos@cmu.edu

David Hoffman
Beasley School of Law,
Temple University
david.hoffman@temple.edu

Alessandro Acquisti
Heinz College,
Carnegie Mellon University
acquisti@andrew.cmu.edu

Abstract

Legal privacy scholarship has typically emphasized the various ways in which plaintiffs fail when bringing legal actions against entities when their personal information is lost or stolen. However, this scholarship is based on a limited set of published judicial opinions about large-scale data breaches. Little is actually known about the characteristics and disposition of a representative set of data breach lawsuits. Using a unique sample of manually-collected data from PACER, we analyze the court dockets of over 200 federal data breach lawsuits from 1998 to 2011. We use discrete outcome regressions to better understand which breaches are being litigated and which lawsuits are being settled. Our results generally (though not strictly) support theoretical notions of litigation. We find that while breach characteristics (size, cause and types of information compromised) are positively correlated with probability of filing a lawsuit, they contribute less when predicting the outcome of the suit. Instead, the probability of settlement is mostly driven by typical legal procedural matters such as actual harm and class certification.

Keywords: data breaches, identity theft, privacy litigation, docket analysis