Public Review for

The Devil and Packet Trace Anonymization

Ruoming Pang, Mark Allman, Vern Paxson, & Jason Lee

Reproducible research hinges on the ability to use common datasets, and so public datasets containing measurements of real, operational networks are important for the Internet measurement community, and subsequently for all those fields of networking that use their research. There are several such datasets available now, but given the ongoing evolution of the Internet, more are always needed. However, such data is often considered proprietary, or its release raises privacy or security concerns, and so access to data is often limited to the organization that collected it. Anonymization of data has sometimes been used to remove the offending components of a dataset so that the remainder can be made public.

As with all such work, the Devil is in the detail. Different organizations have their own requirements for anonymization, while on the other hand, researchers have different interests in a dataset. Particular aspects of an anonymization may remove the component of the data that is of interest in a study. Anonymization policies that balance security and privacy with research value are needed. It is a challenging problem --- as one reviewer put it “the difference between a researcher and an attacker cannot be expressed in pure data analysis terms. It is motivation or funding source that makes the difference.”

In studying the issues of their own organization, the authors of this paper sought not to add a new tool that would perform anonymization according to their own policies, but rather they aimed to create a new (freely available) tool ‘tcpmkpub,’ which would allow one to implement complicated, multifaceted anonymization policies, considerably beyond the capabilities of existing tools. The paper also provides considerable discussion of the issues surrounding anonymization policies, and finally provides access to a large dataset (11 GB set of packet traces anonymized using their tool).

The reviewers were all largely positive about the paper, for instance saying “it raises the level of thoroughness at which packet traces are anonymized,” and “The authors present a tool able to fulfil this task [anonymization] in an elegant way and whose flexibility promises to make it easy to adapt in different environments.” However, two reviewers noted the fundamental problem of making a trace totally attacker proof is not solved here, though this problem is highly challenging, and perhaps not solvable --- so there is plenty of remaining research to be performed in this area.

Public review written by

Matt Roughan
University of Adelaide,
Australia

acm sigcomm