A Bureau of Cyber-statistics

David Clark, MIT and kc claffy, CAIDA, UCSD

The recent Cyber Solarium Commission report[1] sets out a strategic plan to improve the security of cyberspace. Among its many recommendations is that the government establish a Bureau of Cyber Statistics, to provide the government with the information that it needs for informed planning and action. A recent report from the Aspen Institute echoed this call.[2] Legal academics and lobbyists have already started to consider its structure.[3] The Internet measurement community needs to join this conversation.

The particular focus of these reports is cybersecurity, but there are many other issues related to the character of cyberspace that the government needs to understand. If established, a Bureau of Cyber Statistics would need a much broader remit than security. Or to be blunt, we cannot secure what we do not understand.

The Solarium report proposes some specific characteristics: they recommend the bureau be in the Department of Commerce, and funded and authorized to gather necessary data. The report also says that "*the center should be funded and equipped to host academics as well as private sector and independent security researchers as a part of extended exchanges*". We appreciate that the report acknowledges the value of academic researchers, but this objective requires careful thought to achieve. The report specifically mentions "purchasing private or proprietary data repositories". Is "extended exchanges" the only pattern of access, where an academic would work under an NDA and not be able to publish results that relied on proprietary data? Would this allow graduate students to participate? The proposal does not indicate any understanding of how academic research actually works.

As an illustrative example, the authors were hired by AT&T as "independent measurement experts" to propose and oversee methods for AT&T to satisfy FCC reporting requirements imposed as a merger condition.[4] All the data we received was covered by an NDA, and we were not able to publish any details about what we learned. This sort of work is not academic research, it is consulting.

The bureau must be organized in such a way that academics are able and incentivized to utilize the resources of the bureau for research on questions that motivate the creation of the bureau in the first place. But this requires that when the U.S. government establishes the bureau, it makes apparent the value of academic participation and the modes of operation that will allow it. Thus, the critical question we pose is how can the research community demonstrate its value as an independent voice helping to inform the future of the Internet, to justify the argument that a bureau of cyber-statistics should be organized to allow third-party access to the data it holds, in a way that makes academic research practical.

This demonstration will not be effective if it is hypothetical. The research community must demonstrate its value through real projects that produce important results. But real projects are tricky, because the

---

[1] https://www.solarium.gov/report

[2] https://www.aspeninstitute.org/longform/a-national-cybersecurity-agenda-for-resilient-digital-infrastructure/

[3] https://www.lawfareblog.com/considerations-structure-bureau-cyber-statistics

[4] https://catalog.caida.org/details/paper/2016_att_ime_first_amended_report
https://www.caida.org/publications/papers/2016/att_ime_justification/att_ime_justification.pdf

data does not necessarily exist yet, and if it does, may be proprietary. So the barrier we must overcome is the “chicken and egg” problem of how to demonstrate the value of an independent research community before the bureau exists.

The starting point must be to work with public data, and translate research results into forms that are meaningful to a constituency broader than the research community. But this path reveals more specific barriers: Who would fund such research? What are the incentives of the academic research community to undertake it? But if we do not recognize this challenge, the independent research community may essentially be written out of the story, as more and more data is proprietary and hidden away.

The call for a bureau of cyber statistics is thus an opportunity and a threat. It is a threat if the data is not actually available to the research community on practical terms. It is an opportunity in that this recommendation, if it continues to get traction, is a focal point for discussion about collection and use of data, where we have an opportunity and responsibility to make abstract calls for access to data more specific and concrete.