EPOC Network Data Privacy Policy
February 19, 2019

I. Introduction

The Engagement and Performance Operations Center (EPOC) uses the NetSage measurement and monitoring tool (http://netsage.global) to gather network data for use in understanding file transfer performance and debugging user data transfer issues. The NetSage tool is an open privacy-aware network measurement, analysis, and visualization service designed to better understand the behaviors and needs of today's research and education (R&E) networks.

Although currently in use by other projects, the NetSage Project was originally funded by the US National Science Foundation (NSF) International Research Network Connection (IRNC) program to better understand the use of the IRNC-funded backbone networks and exchange points. In much the same way as other large-scale NSF facilities track their end users, the NetSage tool was created to understand the use of the IRNC networks. For example, the XSEDE (https://www.xsede.org/) high performance computing platform tracks end users by institution, science domain, and project, as does the Open Science Grid computing consortium (https://www.opensciencegrid.org/). NCAR/UCAR (https://library.ucar.edu/) tracks the use of their data resources in similar ways.

The EPOC project uses the NetSage Tool to understand and visualize large data flows associated with research, education, and science projects for its associated partners, and also uses this data to debug performance issues. The data is de-identified and used primarily to understand the network behaviors of large flows and to better understand the general use and functionality of the monitored networks and exchange points.

EPOC works with their *Regional Partners* to collect data from networks, exchange points, and archives associated with those partners and with their partners' permission. The EPOC data privacy policy for data collected by the NetSage tool strives to balance the privacy interests of users whose data transits the networks that the NetSage tool monitors, the operational needs of the EPOC project and the Regional Partners, and the need to demonstrate the broader benefit of the NSF-funded resources. We are committed to protecting privacy and informing interested parties about our policies and practices.

II. Scope of this Policy

This policy identifies:

- The information the NetSage tool collects about data transferred by its infrastructure;
- The ways in which this information may be used and disclosed to third parties; and
- The security measures adopted to prevent unauthorized access to this information.

III. What Information Is Collected?

The NetSage tool captures and collects *active* networking data (for example, latency and throughput from a tool such as perfSONAR) and *passive* network metadata (for example, SNMP and flow data). This data may consist of packet headers in addition to performance data, but will never contain payload data from flows. Data sets are de-identified at the source before being stored. The NetSage tool uses current accepted practices in the R&E community, such as anonymizing the IP addresses in a prefix-preserving manner, to de-identify the data released to researchers to ensure user privacy wherever possible. The data is then highly aggregated and does not contain information about traffic flows specific to individual users. EPOC partners may choose to release additional data to the EPOC team, however that data is not part of this policy.

IV. Disclosure of Data

EPOC is the steward of all the network data it collects. EPOC, at the direction of the EPOC PI, may share network data under the following circumstances:

1. EPOC plans to make summaries of de-identified Regional Partner network traffic data public on the EPOC website (http://epoc.global).
2. Upon request, Regional Partners can have access to the full, de-identified data sets for their site.
3. In rare cases where there is an ongoing performance issue for a specified flow, access to raw data may be needed to debug performance issues. If both endpoints of a flow agree, the EPOC collaboration may, for a limited time, use raw data at a collection point to help identify ongoing performance problems between two sites. Internal procedures exist to ensure this is done securely. The EPOC team takes the following actions when collecting raw data:
    a. Permission is obtained from the authorized representative at Regional Partner.
    b. Affected organizations are notified.
    c. Internal logs are maintained documenting what data is collected, the time period covered, who collected the data, and why the data was collected.
    d. When the issue has been resolved, the Regional Partner will also be informed and the raw data files will be subsequently destroyed using industry best practice data sanitization techniques.

V. How Data Is Collected, Retained, and Protected

All network data is managed under the control of EPOC project members authorized by the EPOC Principal Investigator at Indiana University.

EPOC takes appropriate steps to protect collected network data from unauthorized access or disclosure. Additionally, EPOC employs industry standard security measures to protect against the disclosure, loss, misuse, and alteration of the information under our control.

VI. Notice for Updates and Changes to Policy

This document is derived from the original NetSage project data privacy document (http://www.netsage.global/home/netsage-privacy-policy), which was derived from ESnet's privacy policy (available at https://www.es.net/about/governance/data-privacy-policy/), which itself is derived from the Internet2's policy on privacy of network flow data (available at http://www.internet2.edu/policies/network-flow-data-privacy-policy/). EPOC reserves the right to update this privacy policy at any time to reflect changes in the manner in which it deals with traffic, whether to comply with applicable regulations and self-regulatory standards, or otherwise. Then Privacy Policy posted here will always be current. We encourage you to review this statement regularly.

VII. Who to Contact if You Have Questions

If you have any questions about this privacy policy, please contact Dr. Jennifer M. Schopf, the PI of the EPOC project, at jmschopf@iu.edu.

VIII. Glossary

Networking data: Active — Data collected by tools that perform active tests at the user level, such as perfSONAR

Networking data: Passive — Network traffic data, such as netflow or sFlow data, or data collected using passive monitoring tools such as Tstat or from routers directly

Regional Partners — At the time of this writing, EPOC's regional partners included:

- The Indiana State Network (I-Light)
- Ohio State R&E Network (OARnet)
- Keystone Initiative for Network Based Education and Research (KINBER)
- Great Plains Network (GPN)
- The Texas State R&E Network (LEARN)
- Front Range Gigapop (FRGP)