Universität Basel

Publication

Anonymization of System Logs for Preserving Privacy and Reducing Storage

System logs constitute valuable information for analysis and diagnosis of systems behavior. The analysis is highly time-consuming for large log volumes. For many parallel computing centers, outsourcing the analysis of system logs (syslogs) to third parties is the only option. Therefore, a general analysis and diagnosis solution is needed. Such a solution is possible only through the syslog analysis from multiple computing systems. The data within syslogs can be sensitive, thus obstructing the sharing of syslogs across institutions, third-party entities, or in the public domain. This work proposes a new method for the anonymization of syslogs that employs de-identification and encoding to provide fully shareable system logs. In addition to eliminating the sensitive data within the test logs, the proposed anonymization method provides 25% performance improvement in post-processing of the anonymized syslogs, and more than 80% reduction in their required storage space.