

# Анализа данных трафика научного учреждения с использованием вычислительных систем

РЫГОВСКИЙ ИВАН АНАТОЛЬЕВИЧ  
e-mail: [ivan.rigovsky@csc.sibsutis.ru](mailto:ivan.rigovsky@csc.sibsutis.ru)

Представлены методы для ретроспективного анализа данных IP трафика научно-исследовательского академического института, результаты проделанной работы. Полученные результаты позволяют установить закономерности и некоторые особенности в поведении пользователей.

На различных этапах анализа были использованы, как стандартные статистические методы, так и методы интеллектуального анализа данных (Data Mining). При этом были разработанные собственные параллельные алгоритмы, с использованием стандартов OpenMP и MPI, и исследована их эффективность на различных наборах данных и конфигурациях вычислительных систем.

Предложенные методы позволяют выявить новые, ранее неизвестные кластеры пользователей, провести анализ известных классов, а также делать выводы о развитии сети на различных отрезках времени.

<http://dl.dropbox.com/u/14919248/img.png>

Рисунок 1. Распределение пользователей на плоскости, относительно количества потребляемого трафика.

К примеру, на рисунке 1 можно увидеть полученное при помощи метода многомерного шкалирования распределение пользователей на плоскости, относительно их поведения в сети, с градацией по количеству потребляемого трафика (цвет) и по различию в поведении (расстояние между точками на плоскости). Видно, что пользователи с различным объёмом трафика имеют схожее поведение, что позволяет отнести их к одному классу.

Предложенные методы позволяют оценить поведение пользователей с различных углов зрения, учитывая особенности анализируемой сети и конкретных её пользователей, применяя индивидуальный подход к анализу их поведения.