Flow aggregation traffic monitor
To easily understand network issues, the NECOMA Project includes R&D of traffic monitoring tools.

Flow aggregation
Traffic flow is multidimensional information represented by five-tuple (sender address, destination address, sender port, destination port, and protocol). Generally, the distribution of flow characteristics in traffic is biased, and clustering can identify characteristic groups. In other words, aggregating flow in a multidimensional space allows attacks and scams from multiple individual flows to be consolidated, revealing general trends (Fig. 1). Moreover, changing the flow granularity exposes necessary details and can be applied to time (Fig. 2). However, flow aggregation in a multidimensional space is intense, and a practical tool did not exist.

Agurim: Two-step aggregation algorithm
Agurim, which is a flexible and efficient multidimensional flow aggregation tool, uses a two-step aggregation algorithm to make performance and flexibility compatible. Re-aggregating previously aggregated flow information dramatically reduces the search space for flow, allowing for efficient execution.

In primary aggregation, reusable aggregate flow information is generated from raw data such as pcap, NetFlow, and sFlow, and a performance-focused similarity algorithm can efficiently cluster the large amount of flow. In secondary aggregation, the aggregate granularity of flow space and time are adjusted according to the user needs to flexibly execute re-aggregation. By preparing re-aggregated flow information in 5 minute, 1 hour, and 1 day time intervals, and by further re-aggregating this information, interactive operations are possible through a web user interface.

The Agurim prototype, which is available as open source software, is composed of a flow parser, primary and secondary aggregation engines, and the web user interface (Fig. 3).
Agurim
Traffic Monitoring by Flow Aggregation

Agurim Web user interface
Aggregate flow information is plotted on the web browser (Fig. 4). The home screen shows aggregate flow information for the last 24 hours aggregated by byte and packet number. The legend for each aggregated flow shows details, including the overall ratio of aggregated sender and destination addresses in the first row and the protocol details for the aggregate flow in an aggregated form in the second row. Asterisks (*) denote wild card characters.

Users can zoom and pan using the buttons at the bottom of the screen and change the datasets or protocol displays using the drop-down menu. In the protocol display, the flow is aggregated and the details are displayed as aggregate addresses. Additionally, data can be obtained via APIs, allowing cooperation with other tools.

Open dataset
Agurim has been operational since February 2013 as a part of the WIDE Project and is publically available. By allowing network operators and researchers access, we hope to help promote sharing and research on traffic information. Please see the URL below for datasets and source code information:

http://mawi.wide.ad.jp/~agurim/about.html
https://github.com/necoma/agurim

Please see http://www.necoma-project.jp for details on research activities and results of NECOMA.