

MAWI since 1999

open traffic data efforts by WIDE

MAWI WG chair: Kenjiro Cho (IIJ)
February 2026



about MAWI

- MAWI: Measurement and Analysis of the WIDE Internet
 - the oldest active working-group in WIDE
- <https://mawi.wide.ad.jp/mawi/>
- known for open packet traces
 - open traces taken almost every day since Feb 1999
 - daily traces for 15 minutes during 14:00-14:15 JST
 - occasional longer traces for 24-72 hours
- hundreds of papers use MAWI data
 - google scholar: 2,550 hits for "mawi traffic" (2026-02)

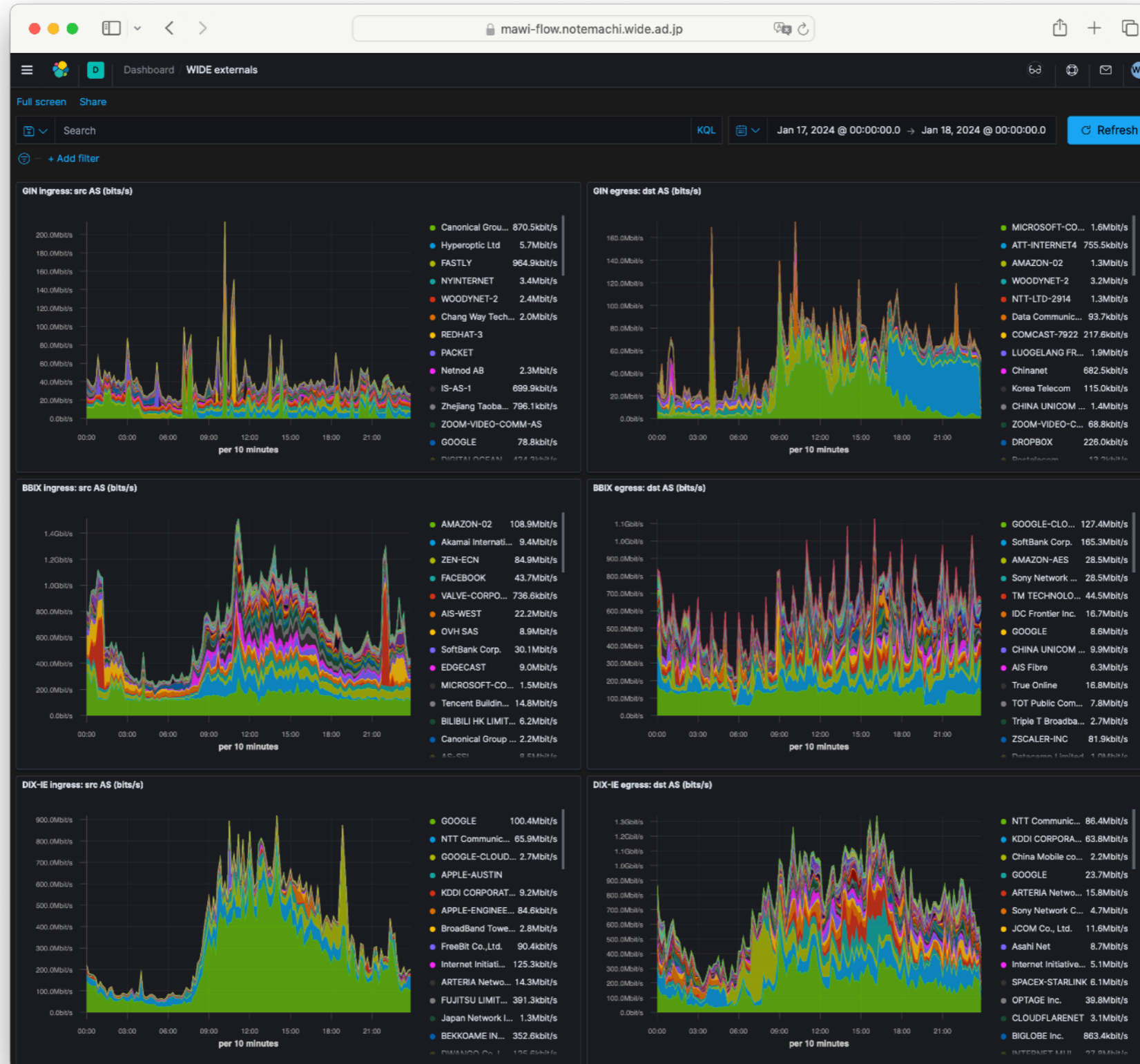
purpose of MAWI open data

- purpose: to promote data analysis and tool development
- background:
 - hard to obtain backbone data
 - NSFNET terminated in 1995
 - universities moved to commercial services
 - WIDE motivated for promoting IPv6
- unofficial purpose: for our own interests, for our own use
 - access privilege: only the operators in charge

current datasets

- Samplepoint-F
 - the transit link of WIDE (to AS2914 NTT GIN): GbE, 2006-
 - an IX link to BBIX: 10GbE, 2023-
- Samplepoint-G (suspended by a machine failure)
 - the WIDE-DIXIE link: 10G, 2018- (every Wednesday)
- other datasets
 - agurim (aggregated flow summaries)
 - iBGP from AS2500 (suspended)
 - netflow (by elastiflow since 2023)

traffic on major external links



other activities

- anomaly labeling (mawilab): terminated in 2024
 - for user convenience
 - comparing different classifiers
 - providing a reference dataset
- aggregated flow summaries
 - summarizing traffic by spatial and temporal aggregation
 - multi-dimensional aggregation technique: agurim

why WIDE does it? discussions in 1999

- for whom, for what purposes?
 - networking researchers and engineers
 - traffic research, tool development
- consistent with the mission of WIDE?
 - if not, it doesn't work or doesn't last
 - Internet culture/WIDE spirit
 - contributing to communities by sharing tools and datasets
- can it be done by others?
 - difficult for US people (due to lawsuit risks)

Lessons Learned

- make data (that we want to use for our own research) open to the public
 - after 25+ years, many tools are already outdated
 - responding to requests from others, exploring how to cut corners
- do not be too ambitious
 - adequate traffic mix and size
 - granularity: hard to track individual flows but enough to identify anomalies
- we do not pursue quality data (e.g., accuracy, perfection)
 - too expensive for long-term effort
- Rome was not built in a day
 - mawi data: recognized as unique datasets
 - for 25+ years, almost everyday
 - missing data due to equipment failures, operational mistakes
 - automation, low-cost, group effort

summary

- open traffic data efforts by WIDE
 - make data (we want to use) open to the public
 - provide what others request
 - adequate size and quality for long-term effort
 - now used by many researchers
- significance and risks of open data
 - give to the one who asks you (Matthew 5:42)