ML Driven Bandwidth Forecasting

A Data Driven Approach To Bandwidth Calculation

NANOG 86

Venkata Naga Chaitanya Munukutla*

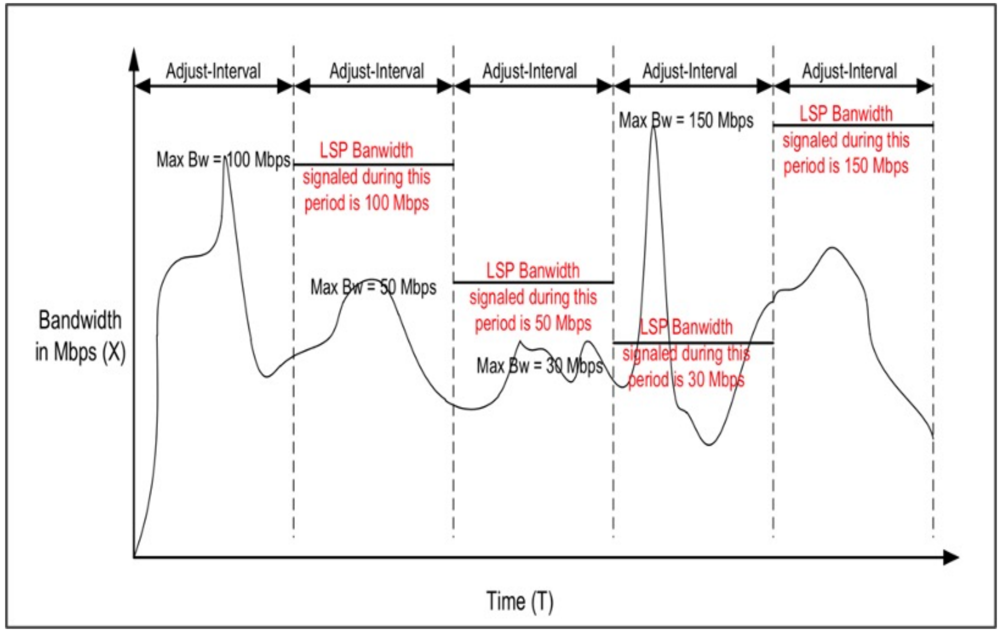
Colby Barth, Krishna Karthik Gadiraju, Vishnu Pavan Beeram

AUTO BANDWIDTH (AUTOBW)

- BW management is crucial for network management.
  - Routing process collects per LSP traffic statistics at certain frequency.
  - AutoBW automatically adjusts BW allocation based on the volume of traffic flowing through the tunnel in real time.
  - Adjustments are made every adjust-interval.
- In each adjust-interval:
  - Max-Average is computed as:

    max-average = max (sampled byte-rates from preceding adjust-interval)

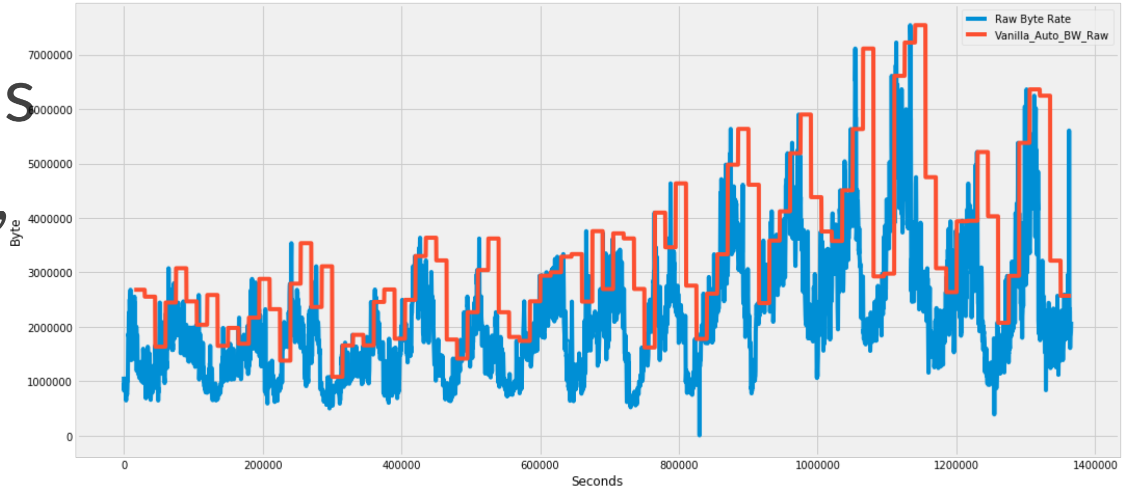
  - BW is adjusted if max-average > adjust-threshold
  - New paths are computed if BW is adjusted



MOTIVATION

- BW management remains a problem for many operators: config recommendations/struggles, numerous requests to “go faster”, more accurate, …
- Current auto-bandwidth algorithm is always “catching-up”
- Many, interdependent configurations to tune the behavior

Current MPLS BW Algorithm



GOALS

- Use patterns in historical BW values to forecast future behavior (“proactive” in nature)
  - Treat BW calculation as a Time Series Forecasting problem
  - Faster BW allocation for traffic rate changes by ‘forecasting’ future BW values based on historical patterns
- Data-driven & Low footprint approach
  - Leverage real-time data (e.g., telemetry, logs)
  - Real-time statistics now available at higher frequency, enabling ML-based solutions even for short historical windows
  - Require minimal changes to network design
  - Reduce operational overhead

TIME SERIES FORECASTING

What is a Time Series?

- A series of observations/measurement taken at specified times at equal intervals.
- Can be used to predict future values based on past observed values.
- **Univariate:** only one variable is observed at each time
- **Multi-variate:** two or more variables are observed at each time

What are the properties we should observe?

- Trends (increasing, decreasing etc.)
- Seasonality (hourly, daily, weekly, monthly)

Which methods can be used?

- Time series forecasting methods such as ARIMA [1] etc. can exploit patterns in the data to predict (forecast) future values.
- Regression-based methods such as Support Vector Regression [2] can predict future values.
- Deep learning methods [3]

[1] Hyndman, Rob J., and George Athanasopoulos. "8.9 Seasonal ARIMA models." *Forecasting: principles and practice. oTexts. Retrieved* 19 (2015).

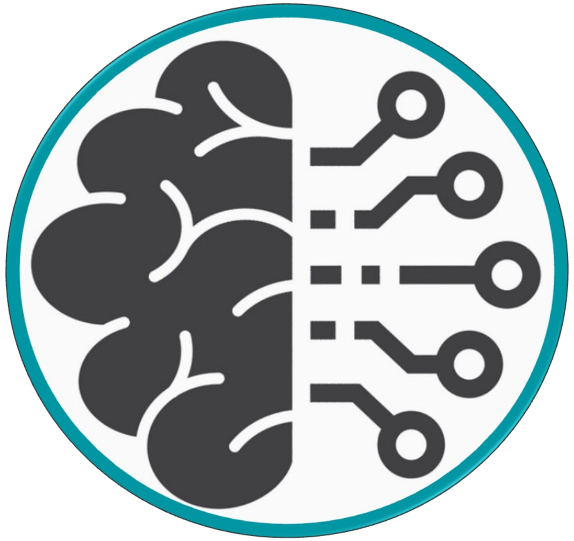
[2] Smola, Alex J., and Bernhard Schölkopf. "A tutorial on support vector regression." *Statistics and computing* 14, no. 3 (2004): 199-222.

[3] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

ML DRIVEN TE APPLICATION FRAMEWORK

Evolution of Bandwidth Forecasting

ML Driven Bandwidth Forecasting



ALGORITHM AGNOSTIC

Max-Avg

ML/DL/Statistical Methods

. . .

TECHNOLOGY AGNOSTIC

RSVP-TE

Native IPv4/6

Segment-routing

WAN/DC/Enterprise

EXTENSIBLE TO OTHER APPLICATIONS

HEALTH MONITORING

SIMPLIFIED OPERATIONS THROUGH AUTOMATION

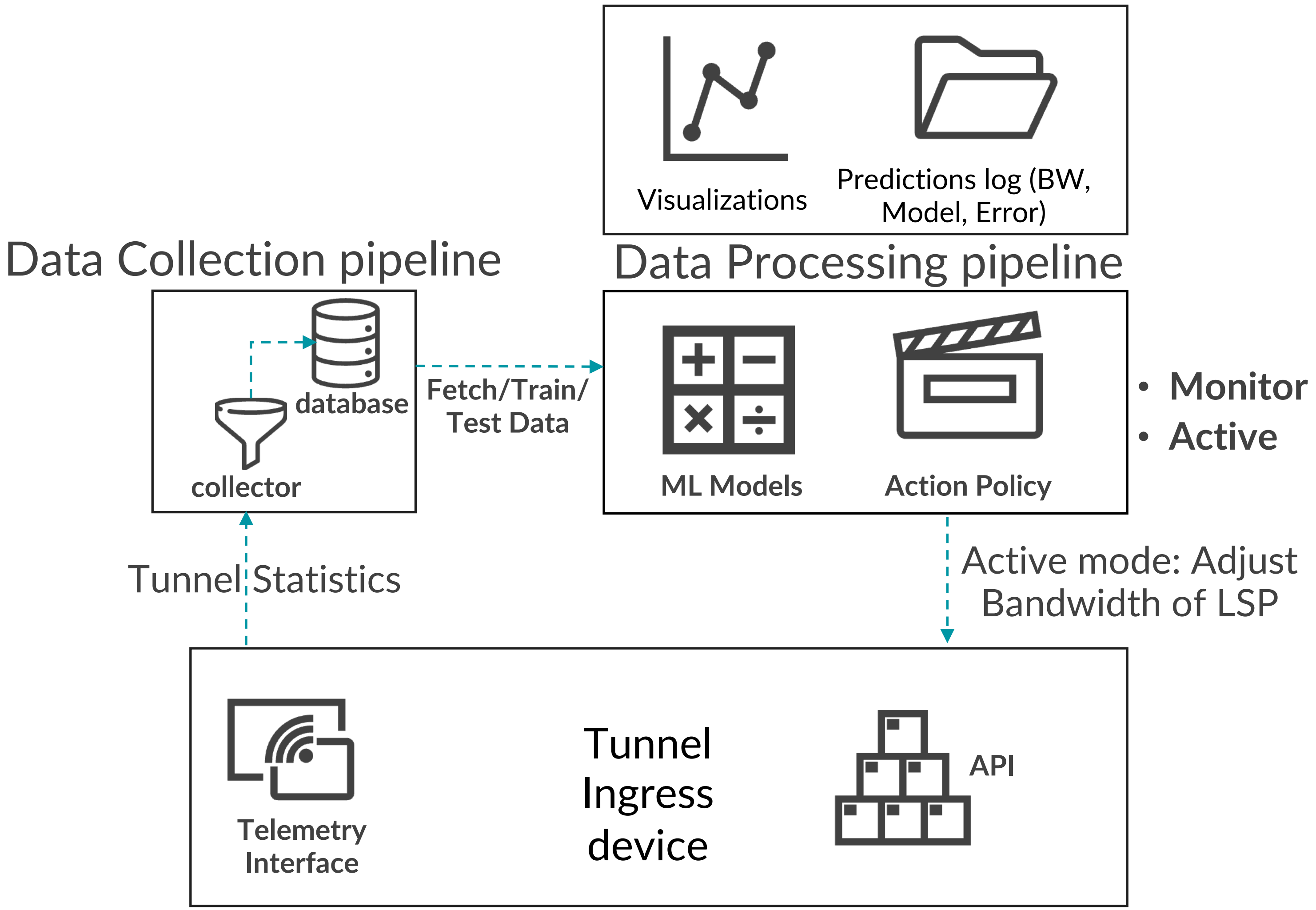
Minimal configuration changes

Light-weight

Deploy anywhere

Micro services-based architecture

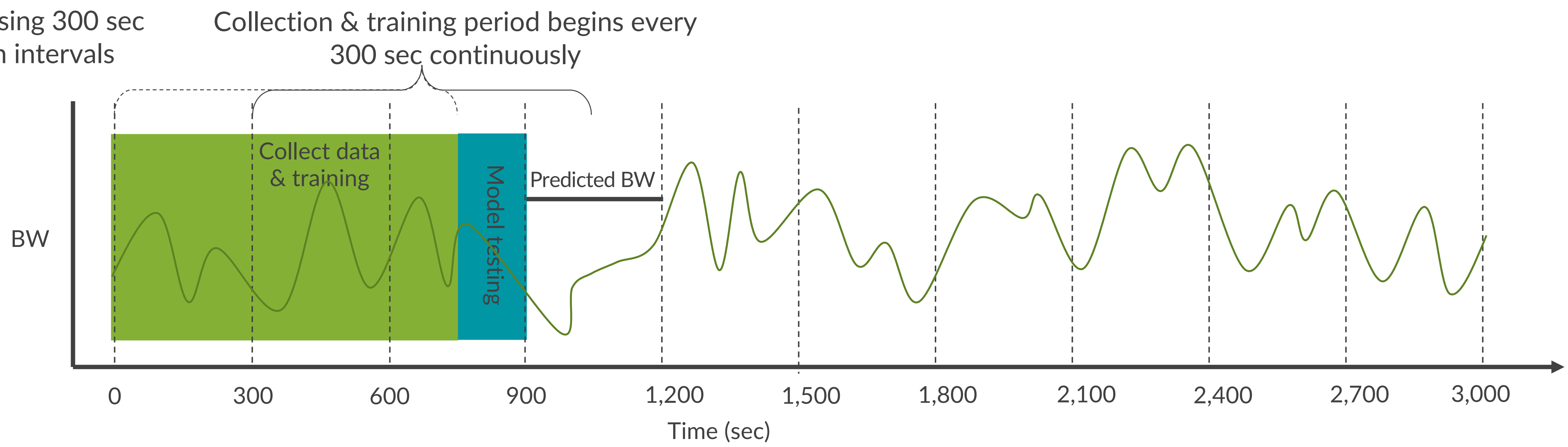
FRAMEWORK FOR FORECASTING



ML BASED FORECASTING LOGIC

- Data collection & model(s) training
  - Model(s) testing
    - Model selection & prediction

* Example using 300 sec prediction intervals

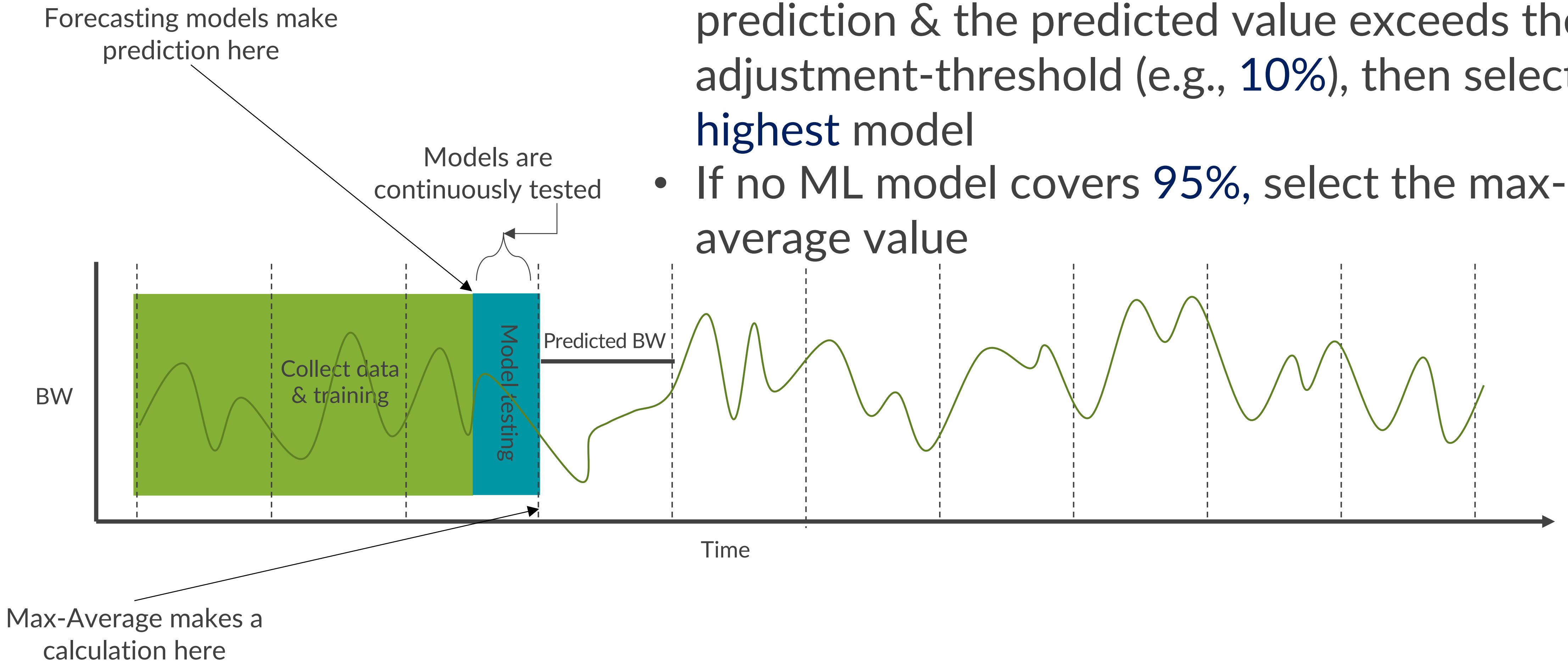


REAL-TIME MODEL VALIDATION

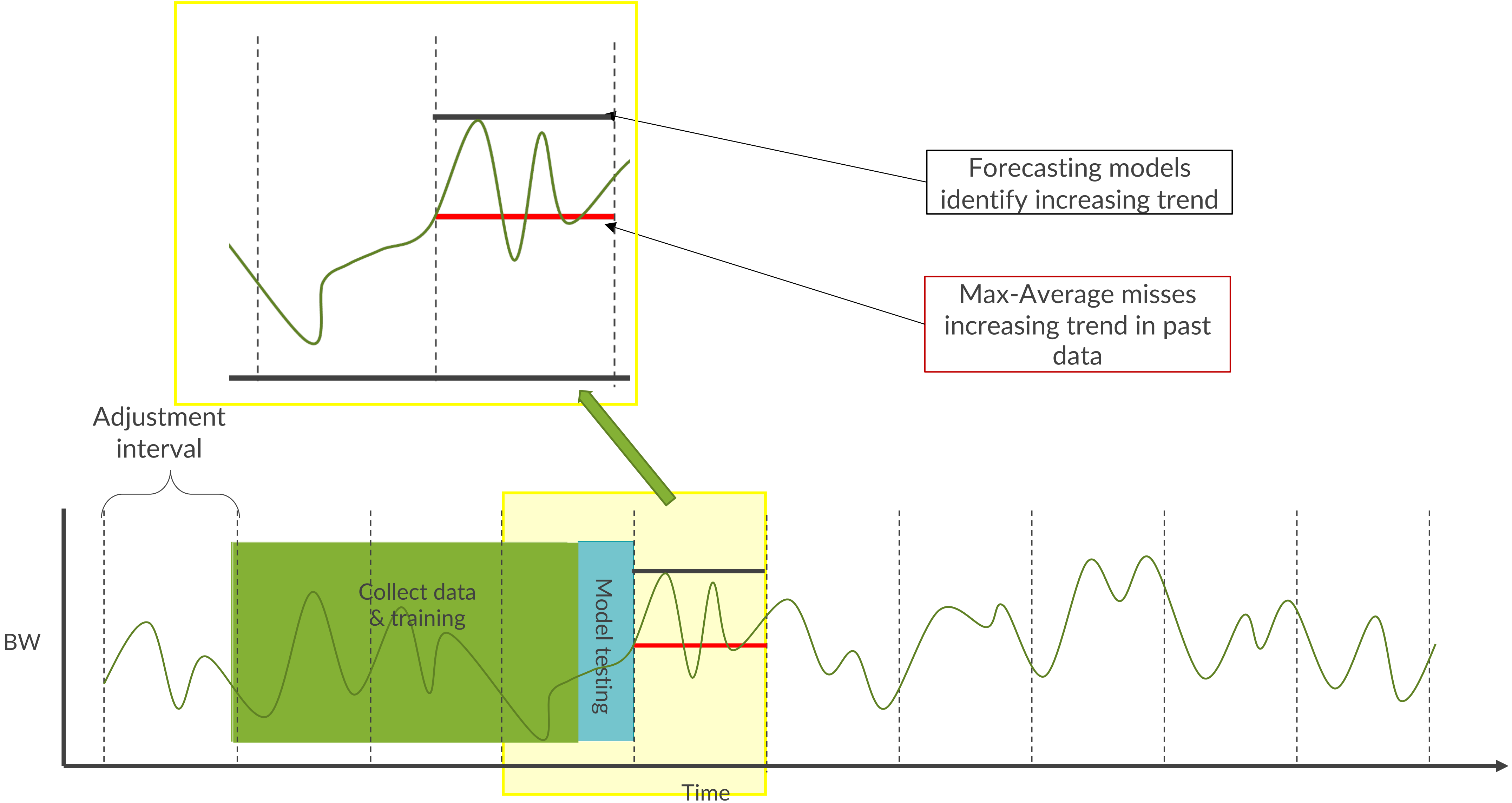
Configurable parameters

Model selection & prediction

- If 95% of traffic is covered by a model's prediction & the predicted value exceeds the adjustment-threshold (e.g., 10%), then select the highest model
- If no ML model covers 95%, select the max-average value



TIME SERIES PATTERN IDENTIFICATION



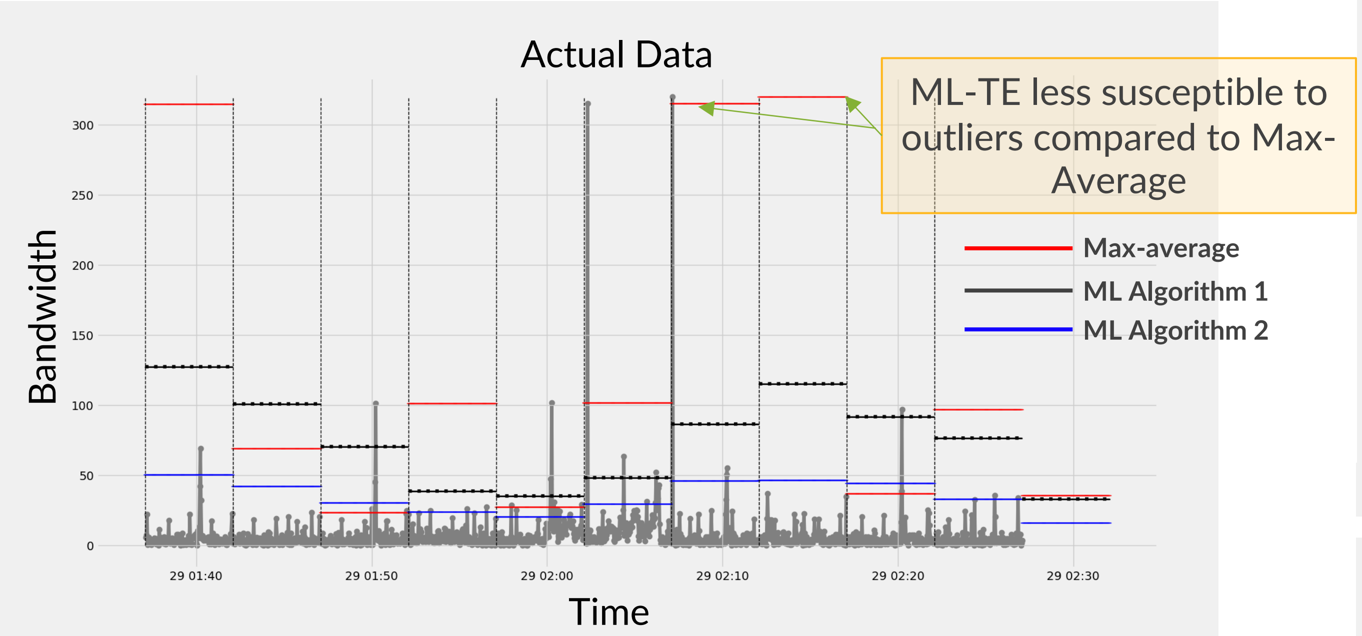
ML BASED BW FORECASTING IN ACTION

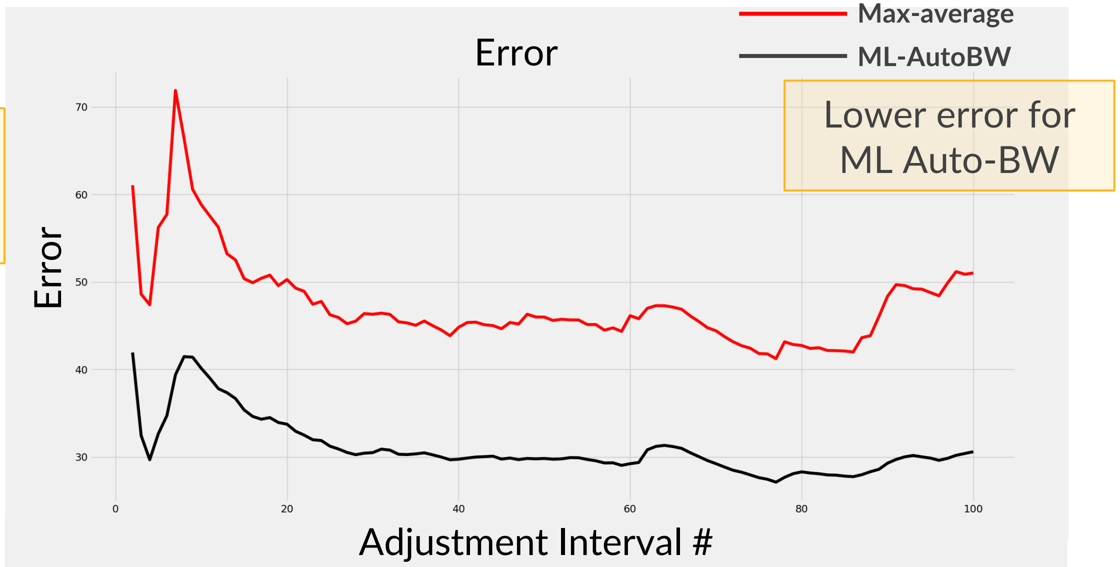
- Informational tracking of collection, training and test periods per Tunnel
- Model specific prediction reporting (predicted BW, % covered reported BW, ..)
- Model selection & error reporting
- Summary for lifetime of predictions

```
**************************** DEVICE: RESERVATION STATUS **************************
Current Mode: monitor. BW Update will not be triggered
Adjust Interval: 300s
Current Signaled BW:                   2.3283 Mbps
Forecasting Model Algo1 Predicted BW:   10.4219 Mbps (% change = 347.6207%)
Forecasting Model Algo2 Predicted BW:   16.2964 Mbps (% change = 599.932%)
Max-Average BW:                         15.7633 Mbps (% change = 577.0358%)
Best technique based on primary analysis: Algo2

                        Adjustment-intervals:
                                # completed: 2354
                                # exceeding adjustment threshold: 2292
                                # not exceeding adjustment threshold: 62
                        Bandwidth updated by:
                                # Algo1: 166
                                # Algo2: 1948
                                # max-average = 178
*************************************************************************************
```

IS ML BASED FORECASTING BETTER?

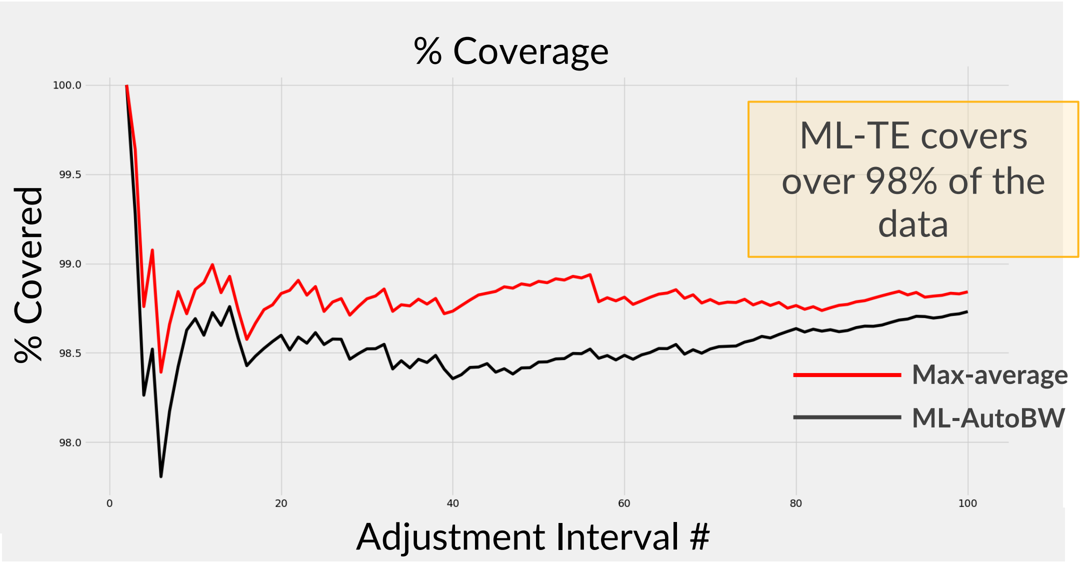




Metrics calculation:

$$MAPE(actual, predicted) = \frac{1}{n}\sum_{i=0}^{n-1}\frac{|actual - predicted|}{max(\epsilon, |actual|)}$$

$$\%covered = \frac{\#\, points\ captured\ by\ method}{total\ points}$$



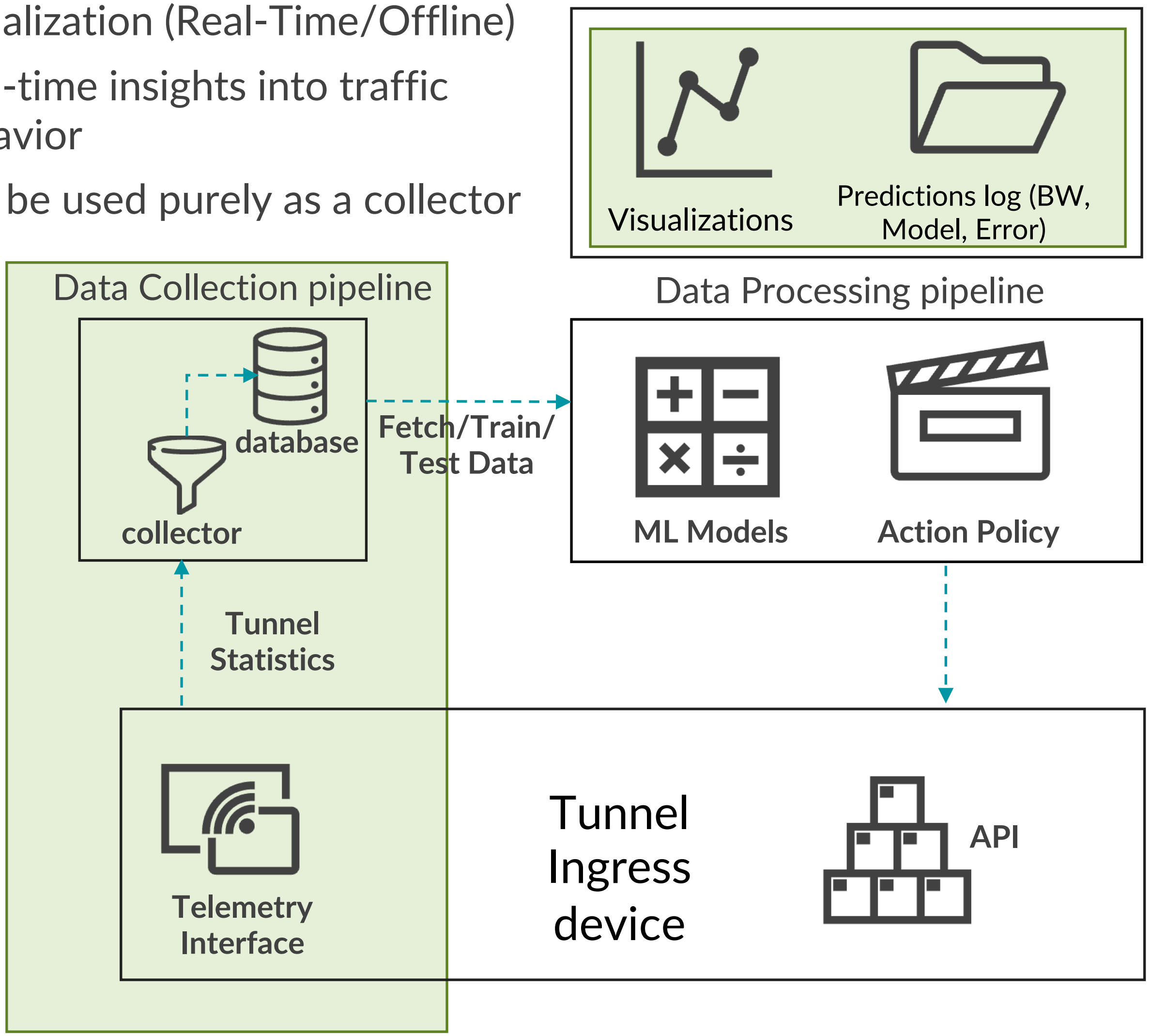
SUMMARY

ML driven TE application framework

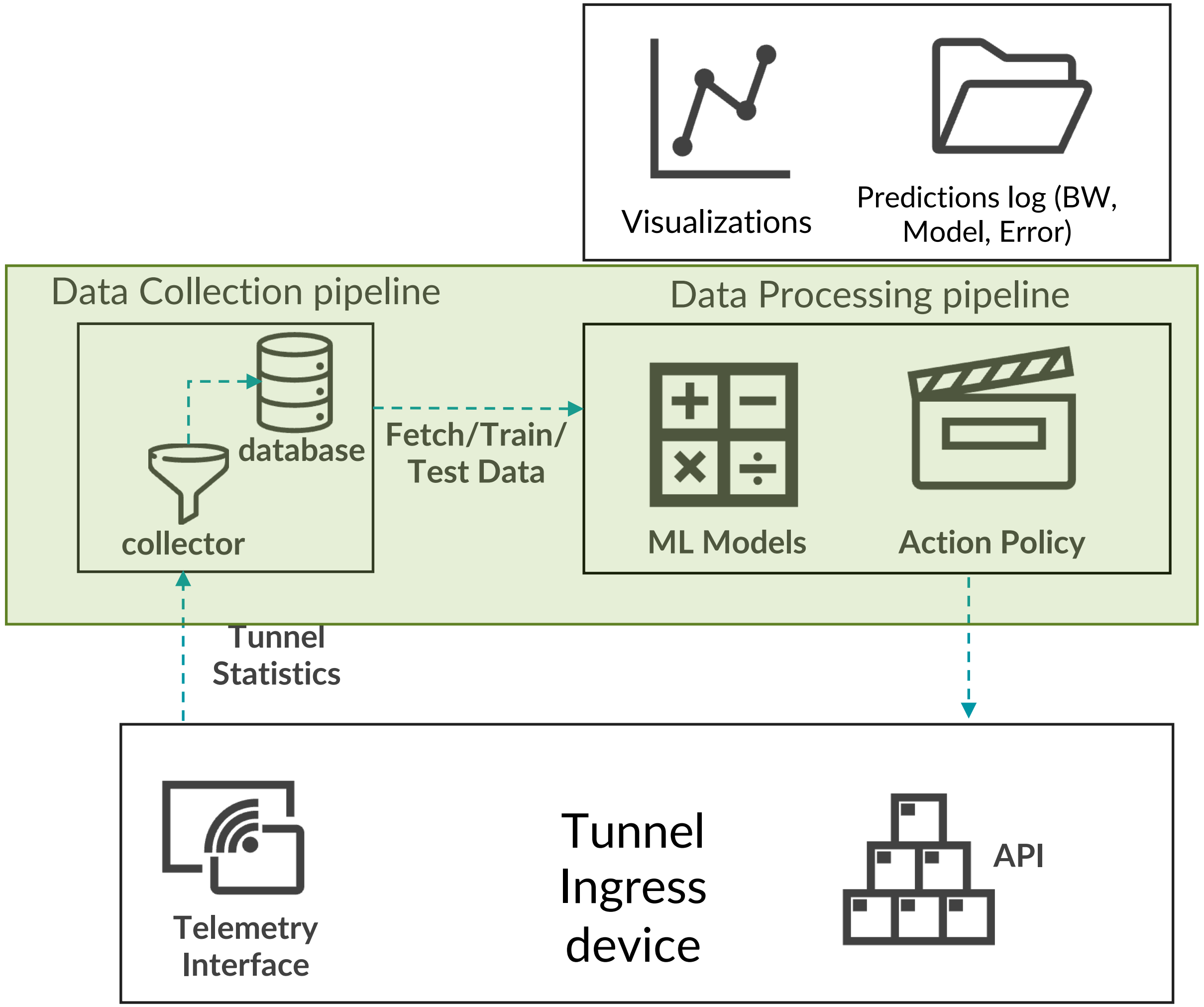
- leverages real-time data.
- allows us to exploit patterns in the historical data (trends, seasonality etc.)
- enables proactive bandwidth forecasting.
- allows for faster adjustment.
- offloads bandwidth calculation.

APPLICATIONS – REAL-TIME INSIGHTS

- Visualization (Real-Time/Offline)
- Real-time insights into traffic behavior
- Can be used purely as a collector



APPLICATIONS – CONFIGURATION EVALUATION



Towards AutoConfig ML Applications:

- Evaluating various device configs using historical data


Thank you