ISYE 3232 Final Report

Abba Otieno Ndomo

December 18, 2025

Data Collection Methodology

Objectives and Scope

We collected vehicle and pedestrian data at a T-intersection with three vehicle approaches (Ferst Dr K, Hemphill H, Ferst Dr C) and associated crosswalks. Our goal was to record arrival, queue progression and exit times so that we could construct inter-arrival, service and waiting-time distributions for later queueing analysis.

Team Roles

For each vehicle approach, a road observation team included:

- **Arrival Recorder**: noted when a vehicle joined the back of the queue.
- **Exit Recorder**: noted when a vehicle cleared the intersection.
- **Data Logger**: entered Vehicle ID, Vehicle Type, Arrival, Stop Sign Arrival (if used), Exit Time and Direction From/To.

A dedicated **Pedestrian Observer** recorded pedestrian arrival at the crosswalk, crossing start and exit times.

Data Schema and Tools

Vehicle sheets contained time fields and automatically computed processing times and counts of cars still in queue (when Exit Time was blank). Pedestrian sheets similarly computed waiting and crossing durations. A summary sheet aggregated cars in queue per road and total pedestrians waiting or crossing. Observers used timestamp shortcuts (Ctrl+Shift+;) and consistent direction and vehicle-type codes.

Procedures and Quality Checks

Times were logged immediately at each event to avoid recall errors. Periodic cross-checks between Arrival and Exit recorders and the Data Logger helped reconcile missing or inconsistent rows. The final dataset consisted of per-approach vehicle records, pedestrian crossing records and summary totals for sanity checks before downstream analysis.

Queueing Analysis of Intersection with Pedestrians

We analysed vehicle data from a signalised intersection with three approach streams and observable pedestrian crossings. For each car, we parsed arrival, stop and exit times. **We defined inter-arrival time as the difference between consecutive arrivals, service time as the time-in-the-intersection (exit minus stop), and waiting time as stop minus arrival.** The objective was to model the system as a single queue with a single server using QPLEX, then compare predicted waiting times to observed waits in order to assess the impact of pedestrians on vehicle delay.

```python
# Helper functions
def _to_time(series: pd.Series) -> pd.Series:
    return pd.to_datetime(series.astype(str).str.strip(), format="%H:%M:%S", errors="coerce")

def _to_seconds(series: pd.Series) -> pd.Series:
    return pd.to_timedelta(series, errors="coerce").dt.total_seconds()

def _replace_nonpositive_with_min_positive(series: pd.Series) -> pd.Series:
    series = series.replace([np.inf, -np.inf], np.nan).dropna()
    positives = series[series > 0]
    if positives.empty:
        return series
    floor = positives.min()
    fixed = series.copy()
    fixed.loc[fixed <= 0] = floor
    return fixed

def load_vehicle_sheet(xl: Path, sheet_name: str) -> Dict[str, pd.Series]:
    df = pd.read_excel(xl, sheet_name=sheet_name)

    arrival = _to_time(df["Arrival Time"])
    stop = _to_time(df["Stop Sign Arrival"])
    exit_ = _to_time(df["Exit Time"])

    proc_col = _to_seconds(df["Processing time"])
    computed_proc = (exit_ - stop).dt.total_seconds()
    service = pd.Series(np.where(proc_col > 0, proc_col, computed_proc))
```

Figure 1: Snapshot of helper methods for cleaning the dataset

We cleaned the data by replacing any non-positive inter-arrival, entry, service or waiting times with the minimum positive value observed in that series and removing missing values. We then formed empirical PMFs by binning times into **5-second intervals, capping support at 30 bins (0–150s)** to keep the QPLEX state space tractable. These PMFs were supplied to the `StandardMultiserver` model with `number_of_servers = 1`. After a 200-step burn-in, we sampled over 400 steps, truncating the state space at 400 states.

Snapshot of the dataset (vehicles, combined across streams)

- Inter-arrival (n=297): mean ~ 5.24s, median ~ 3s, std ~ 7.32s.
- Service (n=298): mean ~ 14.9s, median ~ 13s, std ~ 10.1s.
- Waiting (n=298): mean ~ 96.3s, median ~ 81s, std ~ 87.1s.
- Per-sheet stats are in the earlier cells (`per_sheet`).

Figure 2: Snapshot of the cleaned dataset

For the combined vehicle stream, the mean inter-arrival time was approximately $E[A] \approx 7.3\,\mathrm{s}$, and the mean service time $E[S] \approx 15.0\,\mathrm{s}$, giving an arrival rate $\lambda \approx 0.137$ cars/s and traffic intensity $\rho =$

$\lambda E[S] \approx 2.05 > 1$. QPLEX returned a truncated steady-state distribution with $E[N] \approx 4066$ and $E[Q] \approx 4065$, and Little's law implied $W_q \approx 29{,}700\,\mathrm{s}$ ($\sim 8.25\,\mathrm{h}$). **These values are not physically realistic, since observed waits are on the order of tens of seconds.** Instead, they indicate that, **under the assignment's one-queue/one-server framing, the system is theoretically unstable:** pedestrian-inflated service times combined with observed arrivals exceed single-server capacity. Thus, QPLEX qualitatively signals **that pedestrians significantly reduce effective capacity,** but realistic quantitative predictions would require multiple servers **(e.g. lanes or phases), reduced service times, or lower arrival rates.**