

VECTOR × HyperCity: final adversarial system audit

date	2026-10-25
branch	claude/vector-extremis-forensic-audit-f5vtxg (main untouched)
protocol frozen	commit 63bc12a , before any held-out seed ran (AUDIT_PROTOCOL.md)
held-out evaluation	commit 8c72db6 : 800 of 800 trials completed, 0 failed
post-hoc diagnosis	commit 8558862 and later, fresh seeds only (§5.7, DIAGNOSIS.md)
simulator	SUMO 1.28.0 through TraCI, 1 s step
audit completeness	COMPLETE. All six acceptance gates A–F pass (§6).
system performance verdict	FAIL against the strongest baseline. VECTOR is not better than SUMO's native actuated control in any of the 16 scenarios, and it is significantly worse in B after Holm correction. It is better than fixed-time control in 15 of 16 scenarios. Against a simple vehicle-actuated rule it is comparable to better, with no Holm rejection. It gives emergency vehicles large, consistent gains. No signal-safety constraint was violated in any trial, under the harness's external interlock.

These two verdicts are separate. A complete audit is not a passing system: the runner finishing tells you nothing about VECTOR.

This is simulation and only simulation. Nothing here establishes real-world safety, regulatory compliance, deployment readiness or operational effectiveness. Nothing here supports connecting VECTOR to live public traffic signals (§9).

1. Executive summary

The question. Can VECTOR, as implemented, control a connected 36-intersection simulated city? How does it compare with credible non-VECTOR control under adversarial conditions? The comparison uses paired seeds, with every failure kept.

What was run.

- The network: `city36`, a 6 × 6 SUMO grid. It has **36 connected, signal-controlled intersections**, validated from the network file and pinned by canonical hash.
- The arms, five in total:
 - VECTOR with its advisory shadow stack;
 - VECTOR without the shadow;
 - SUMO's native actuated programme, gap-out logic with matched timing (the strongest baseline available);
 - a simple vehicle-actuated rule controller;
 - 30 s fixed-time control.
- The scenarios: 16 (R0 and A–O).
- The seeds: 10 held-out seeds (95000–95009), run once after the protocol was frozen.
- The trials: 800 held-out trials, all completed, plus 105 smoke trials and 200 post-hoc diagnostic trials on fresh seeds.

What VECTOR is, in the execution path. A per-junction rule-based control law, `control_logic.decide_signals`. It has four rules:

1. emergency override;
2. no-demand rest;
3. starvation guard at 90 s red;
4. vehicle-count priority.

A rate-bounded queue observer filters its input, and a separate safety layer enforces its timing.

The forecasting and tiered-intelligence stack runs only **in shadow**, with no path to the decision. This is verified on every paired trial.

Many named modules are **not executed in the SUMO loop at all**:

- routing, coordination and perception;
- the Sentinel, Serpentine, Aegis, Atlas, Punisher, Phoenix, Hyperion, Terminus, Anarchy, ORACLE and PULSAR stack;
- the learned models;
- the `SignalSafetyGovernor`.

A "Decision Memory Engine" does not exist in the repository (§2).

Headline results, held-out, paired by seed. "Better" means lower P1, the mean time loss per requested ordinary vehicle.

comparison	result
VECTOR vs native actuated (primary, Holm over 16)	Never resolved better. Worse in B: +73.6 s/veh [26.6, 120.5], Holm p = 0.031. Unadjusted, also worse in C, D, H, J, K and O (+29 to +111 s/veh). Not resolved in the other 9.
VECTOR vs vehicle-actuated (secondary, Holm over 16)	No Holm rejection; the smallest is $p_{\text{holm}} = 0.0625$. Unadjusted, better in 7 scenarios (–20 to –127 s/veh) and worse in K (+35.6 [6.2, 64.9]).
VECTOR vs fixed-time 30 s (exploratory)	Better in 15 of 16 (–84 to –218 s/veh). J is not resolved.
emergency vehicles vs native	Ambulance –42 to –61 s, fire engine –65 to –138 s, police –39 s per trip, all with CIs excluding 0. The exceptions are in saturated K: the ambulance is not resolved, and the police car arrives in only 5 of 10 runs under VECTOR (7 of 10 under native).
safety	0 timing violations, 0 illegal signal states and 0 invalid requests in all 640 controlled trials. 0 collisions and 0 teleports under VECTOR. One baseline trial had a SUMO rear-end collision (§4.6).
runtime	VECTOR decide p95 2.0 ms, p99 2.8 ms, max 225 ms. The 36-junction cycle p95 is 24 ms, against a 1000 ms step. The shadow causes almost all of it: without the shadow, decide p95 is 0.06 ms.
prediction (shadow)	MAE is worse than persistence in 15 of 16 scenarios, and on every seed of each. Error doubles after the concept shift in H (1.25 → 2.60 vehicles). No model influences control.

Why VECTOR loses where it loses (§5). Across the 160 paired held-out cells, VECTOR's excess time loss over native tracks its excess *ineffective green* ($r = 0.91$) and *wasted green* ($r = 0.92$). Ineffective green is green on a queue that is not discharging.

The law counts halting vehicles as demand. A queue behind a blocked downstream link therefore looks like demand, and VECTOR keeps serving it until max green. Native gap-out logic sees no vehicles crossing its detectors and moves on. The law has no downstream or discharge term, and no junction knows anything about its neighbours.

A post-hoc diagnosis on fresh seeds (§5.7) tests whether the result repeats and which VECTOR mechanism contributes: on 10 fresh seeds, VECTOR is again worse than native in all five re-tested scenarios.

- R0: +21 s/veh [4, 38].
- J: +67 [17, 117].
- K: +116 [93, 140].
- C: +18 [–2, 38], borderline.
- B: +24 [–23, 70], not resolved. The held-out B effect (+74) did not replicate in size, so its Holm rejection stands as the pre-registered result, but its magnitude is probably overstated.

VECTOR's two mechanisms trade off:

- Green commitment costs about 15 s/veh at moderate demand (R0, C), but it is worth 65–103 s/veh under saturation (J, K).
- The queue observer is worth 14–37 s/veh.

No ablation closes the gap to native under congestion. That places the deficit in the control law itself, not in either mechanism.

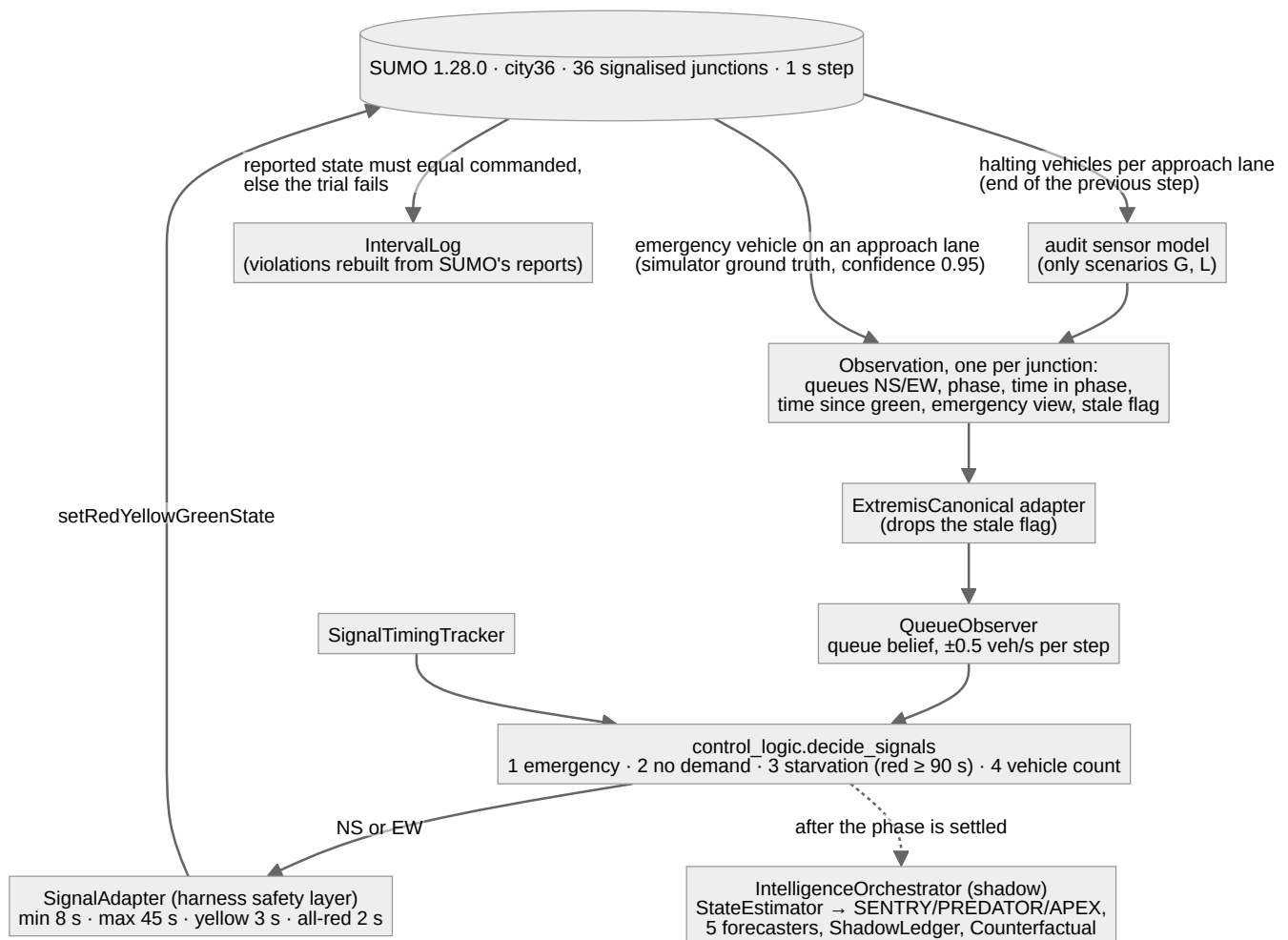
Pilot readiness. The evidence supports **neither connection to live signals nor an uncontrolled field trial**. The defensible next step is an offline, shadow-only research collaboration on a partner's historical data (§9).

2. System architecture: what the code actually runs

2.1 Verified execution path

Every statement below was checked against the code and against a real trial. The checks were:

- docs/architecture/DATAFLOW.md (verified map, 331 architecture tests);
- artifacts/final-adversarial-audit/integrity/module_inventory.json. This file comes from one real 36-junction VECTOR trial, run with call counters on the control path and a record of every module loaded.



There is **no edge from the shadow back to the decision**:

- in the code, `canonical_controller.decide` runs the shadow after `phase` is final;
- in the data, the signal sequence with and without the shadow is identical in every paired held-out trial (§4.7).

Not on the path at all. These modules are not even *imported* during a 36-junction VECTOR trial:

- Sentinel;
- Serpentine (protocol and routing);
- Aegis, Atlas, Punisher, Phoenix, Hyperion, Terminus;

- Anarchy (ORACLE, PULSAR, its Fusion Engine);
- the perception Fusion Engine and the camera/detector path;
- the `SignalSafetyGovernor`;
- the MORL/GNN/causal learners;
- temporal (sequence) inference.

A "**Decision Memory Engine**" **does not exist** in the repository. The protocol stack runs only under `run_extremis.py`, with synthetic perception and a simulated controller. It never runs against SUMO.

2.2 Module table

module	runs in the live loop?	output reaches a decision?	logged?	calls in the inventory trial (600 s, 36 junctions)	on invalid, late or missing output
<code>QueueObserver</code> (state estimate)	yes	yes: the law reads its belief, not the raw count	yes, <code>belief</code> in the trace	19 274 updates (2 per decision)	an impossible change is clamped to ± 0.5 veh/s per step; NaN and negatives are mapped to 0 upstream (<code>_lane_count</code>)
<code>control_logic.decide_signals</code> (decision)	yes	is the decision	yes, <code>reason</code>	9 637	INVALID_STATE → all-red proposal, mapped to the current approach by the adapter (defect D6)
<code>SignalTimingTracker</code>	yes	yes (red times for rule 3)	partly (snapshot)	every decision	—
<code>SignalAdapter</code> (harness)	yes	enforces timing on the decision	yes, <code>effect</code>	9 637 applies	an invalid request is replaced by the current approach and counted
<code>IntelligenceOrchestrator</code> + <code>StateEstimator</code>	yes (shadow, arm vector)	no	yes, <code>shadow</code>	9 637	an exception is caught, counted as <code>shadow_failures</code> , and control continues (scenario M); forecasters are skipped once a 50 ms wall-clock budget is spent, so its output depends on machine load
forecasters: persistence, physics, Kalman, spectral	yes (shadow)	no	the mean only	9 637 / 3 103 / 3 103 / 3 103	an exception is recorded in the registry and the model is skipped; the stack still reports itself healthy (§5.9)
autoregressive forecaster	never called (registered, not invoked)	no	—	0	—
<code>ShadowLedger</code> (prediction verification)	yes (shadow)	no	<code>ledger_verdicts_by_junction</code>	9 637 settle / 9 637 record	—

module	runs in the live loop?	output reaches a decision?	logged?	calls in the inventory trial (600 s, 36 junctions)	on invalid, late or missing output
TrainingGate	constructed	holds every model at UNTRAINED: nothing is authorised	—	—	—
IncidentView (INC1)	loaded, inert (production passes no incident source)	no	—	0	—
emergency detection	the harness reads SUMO ground truth	yes (rule 1)	ev	—	one emergency per junction: the input contract has a single slot
routing / rerouting	absent	—	route changes counted from SUMO's vehroute output	0	—
coordination between junctions	absent : 36 independent instances, no shared state, no offsets	—	—	—	—

A module that only writes a log entry is not counted as integrated. Only the `QueueObserver`, `decide_signals`, the timing tracker and the emergency input influence what SUMO executes. Everything in the shadow is integrated as a measured observer, not as a controller.

2.3 What a "control cycle" is here

A **traffic pass** is one SUMO step of 1 s. In each step, for each of the 36 junctions in fixed order:

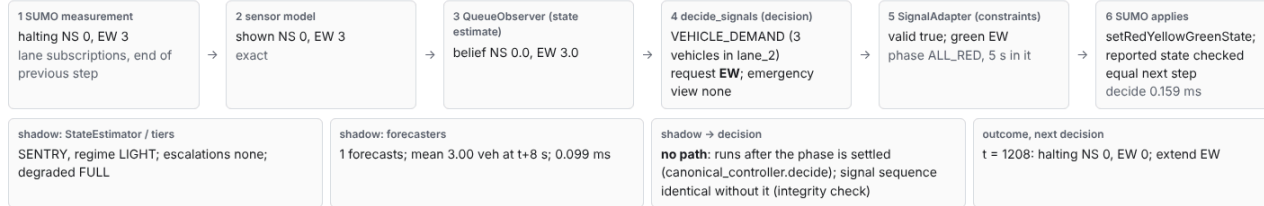
1. If the junction's `SignalAdapter.decision_due()` is true, the harness:

- builds that junction's `Observation`;
- calls its controller;
- applies the result through the adapter.

A decision is due when the minimum green has elapsed (then every second while green) or at the end of clearance. 2. The commanded state is set. 3. SUMO advances. 4. The reported states are checked and the measurements are taken.

Modules cooperate only within one junction's decision: observer → law → adapter (and, in shadow, the orchestrator). No module combines information across junctions. The *cycle latency* in §4.5 is the wall time of step 1 for all 36 junctions together. One recorded cycle, stage by stage, is evidence E18.

One control cycle, end to end: C21 at t = 1200 s (scenario A, seed 95000)



3. Experimental methodology

item	value
simulator	SUMO 1.28.0 (eclipse-sumo wheel), TraCI, step 1 s, teleporting disabled
network	city36: 6 × 6 grid, 36 signalised junctions, 4 two-lane arterials, 8 one-lane side streets, 180–260 m links, 24 fringe nodes; static and actuated variants pinned by canonical hash
demand	Poisson per second and OD route (552 routes); rates by OD class, calibrated with the baselines only; evaluation at 1.15 × base; demand 0–2100 s, drain to 2700 s
scenarios	16 (R0, A-O), scenario_matrix.json
arms	VECTOR (+ shadow), VECTOR without shadow, native actuated, vehicle-actuated, fixed-time 30 s
seeds	calibration 90000–90001 (baselines only); smoke 91000–91001; held-out 95000–95009 , run once after the freeze (63bc12a); diagnostic 96000–96009 (post-hoc)
trials	16 × 5 × 10 = 800 held-out trials (plus 105 smoke, 200 diagnostic)
primary endpoint	mean time loss over all requested ordinary vehicles; unfinished trips are charged their loss so far
statistics	paired by seed; 95% t and bootstrap intervals; Wilcoxon signed-rank; Holm over 16 scenarios in two named families
integrity checks	harness ≡ validated runner (VECTOR and native, same seed); shadow non-influence (every paired trial); health gate per trial (accounting, teleports, SUMO state = commanded state, timing violations)
reproduction	\$10 and AUDIT_PROTOCOL.md

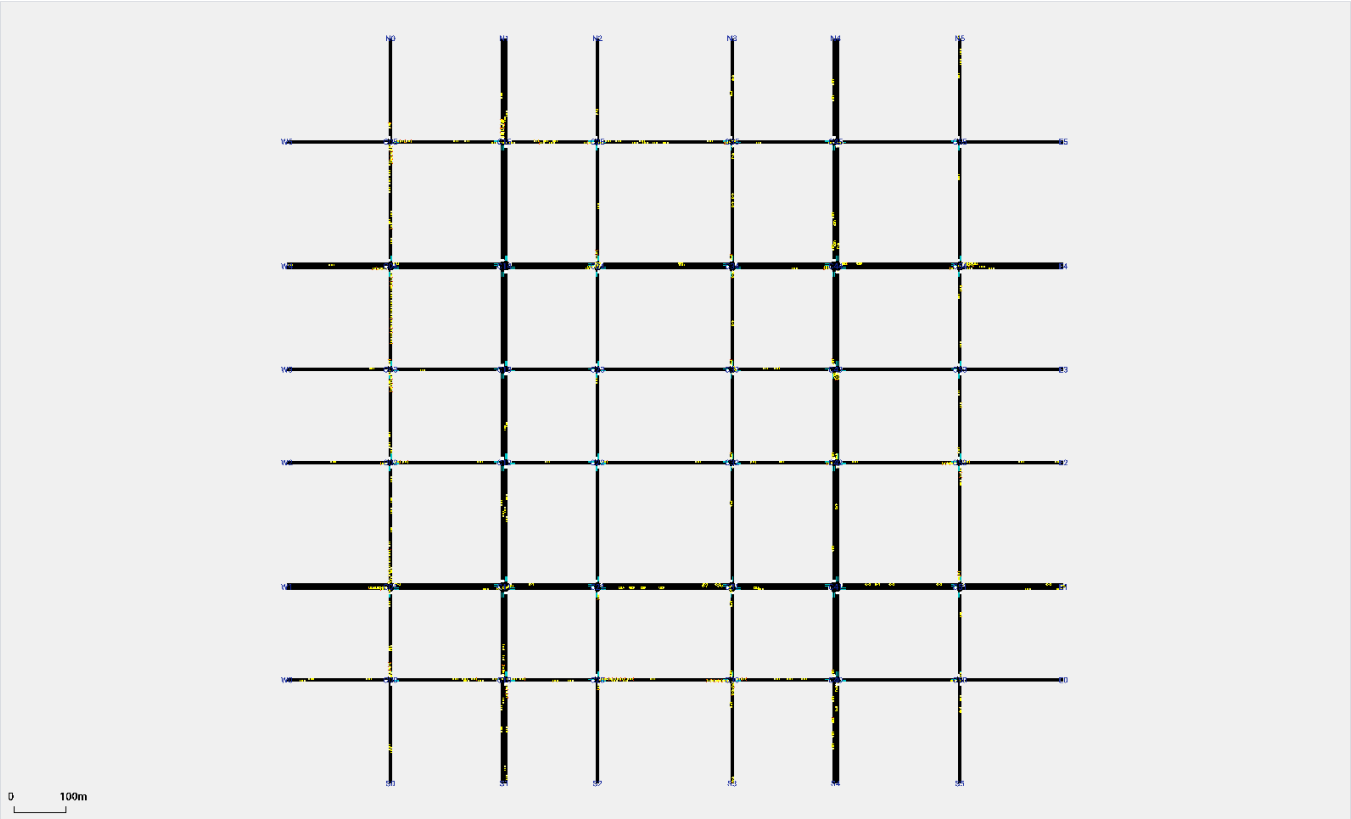
3.1 Network

The network comes from topology_report.json, which is validated from the SUMO file, not assumed:

check	value
junctions in the file	60: 36 traffic-light junctions and 24 dead-end fringe nodes
signal-controlled intersections	36 (C00–C55)
connected	strongly connected between the controlled junctions; each one is reachable from every entry

check	value
roads, lanes and links	168 directed edges, 224 lanes, 480 controlled links
routes	552 OD routes, all legal (528 with turns)
measurement	192 stop-line detectors
validation problems	0

The network is a synthetic regular grid, chosen because no real 30+ junction network with real demand was available. It is not a model of any city. Evidence E01 (the whole grid under VECTOR) and E02 (the topology validation) show it.



3.2 Baselines

No baseline was written to make VECTOR look better.

- **Native actuated** is SUMO's own actuated programme: gap-out on its own detectors, with min and max green matched to VECTOR's adapter. It is the strongest controller here.
- **Vehicle-actuated** and **fixed-time 30 s** are the repository's existing rule controllers. They run through the same `SignalAdapter`, the same harness and the same safety checks as VECTOR.

The harness was checked against the validated runner (`stus/sumo/multi`). On seed 91001, VECTOR and native give identical results on 13 compared metrics (`integrity/equivalence_91001.json`).

3.3 Scenarios

All windows are in simulated seconds. Demand runs from 0 to 2100 s at $1.15 \times$ base, and every trial ends at 2700 s. Emergency vehicles are named `amb1`, `fire1` and `pol1`.

id	what is injected
R0	reference: base demand, no event
A	demand surge $\times 4$ from W1 and E1 (row-1 corridor), 600–1500
B	three simultaneous congestion events: lane blocks on C12→C22 (600–1500) and C43→C44 (650–1450), plus a $\times 3$ surge from S0 and W0

id	what is injected
C	ambulance W1→E1 at 900 s, through 6 signalised junctions
D	ambulance W1→E1 at 900 s and fire engine S2→N2 at 909 s, converging on C21 from conflicting approaches
E	D plus police S4→N4 at 949 s, which conflicts at C41
F	capacity cut on C12→C13: one lane blocked and speed at 30%, 700–1500
G	sensor noise (SD 0.35), 15% dropout and 2% phantom spikes of 25 vehicles, at every junction, all run
H	concept shift from 900 s: side-street demand × 4, arterial × 0.4
I	downstream spillback: both lanes of C31→C41 blocked, plus a × 2 surge from W1, 600–1500
J	demand moves: the west half × 2 from 300–1200 s, then the east half × 2 from 1200–2100 s
K	conflicting objectives: side-street × 3 and arterial × 3 surges, plus ambulance W4→E4, fire engine N2→S3 and police E4→W1 at 950–970 s
L	stale (10 s delay), contradictory (approaches swapped at C11, C14, C41 and C44), dropped (10%) and invalid (1%) observations, 600–1500
M	VECTOR's forecasters fail at 900 s and the whole advisory orchestrator at 1500 s
N	recovery: both lanes of C21→C31 blocked with a × 4 surge, 600–1200; measured as the network recovers to 2100 s
O	combined stress: corridor surge, downstream blockage, capacity cut, ambulance, fire engine and 2 random emergency vehicles

Scenarios H, J and K saturate the network for **every** controller. The native completion rate there is 66–82%. Those scenarios are kept as defined.

4. Results (held-out seeds 95000–95009)

The full tables are in `RESULTS_TABLES_heldout.md` (T1–T9), generated from `artifacts/final-adversarial-audit/results/heldout/`. This section gives the results that decide the verdict.

4.1 Execution and integrity

- **Trials.** 800 attempted and 800 completed (`trials/heldout/attempts.jsonl`). 0 failed, 0 retried, 0 discarded.
- **Health gate.** 799 of 800 pass. The one failure is the vehicle-actuated arm in G, seed 95009: one SUMO rear-end collision and teleport (§4.6). It is kept in every analysis. Excluding it would move VECTOR – vehicle-actuated in G from –20.2 to –24.5 s/veh and would not change any conclusion.
- **Harness equivalence.** The harness gives the same results as the validated runner for both VECTOR and native (§3.2).
- **Shadow non-influence.** In 160 of 160 paired held-out trials, the signal sequence is identical with and without the shadow (§4.7).

4.2 Primary: VECTOR vs native actuated (P1, Holm over 16 scenarios)

VECTOR final adversarial audit rendered from results/heldout/comparisons.json, results/heldout/summary.json

Measured comparison, held-out seeds 95000–95009: VECTOR – native_actuated on P1_mean_time_loss_s

scenario	paired seeds	VECTOR	native_actuated	difference	95% CI (t)	relative	seeds better/worse	Wilcoxon p	Holm p	at 95%
R0	10	170.01	181.97	-11.96	[-69.93, 46.01]	-6.6%	4/6	0.9219	1.0000	not resolved
A	10	330.06	302.60	27.46	[-22.79, 77.72]	9.1%	4/6	0.2754	1.0000	not resolved
B	10	314.75	241.19	73.56	[26.61, 120.51]	30.5%	0/10	0.0020	0.0312	VECTOR worse
C	10	173.80	140.73	33.06	[16.31, 49.82]	23.5%	1/9	0.0039	0.0586	VECTOR worse
D	10	175.60	146.15	29.45	[12.67, 46.23]	20.2%	1/9	0.0059	0.0762	VECTOR worse
E	10	173.43	173.12	0.30	[-57.72, 58.33]	0.2%	2/8	0.4316	1.0000	not resolved
F	10	174.08	162.15	11.93	[-26.83, 50.70]	7.4%	3/7	0.3223	1.0000	not resolved
G	10	163.65	181.97	-18.32	[-73.39, 36.75]	-10.1%	5/5	0.8457	1.0000	not resolved
H	10	458.10	406.92	51.17	[19.86, 82.49]	12.6%	2/8	0.0098	0.1172	VECTOR worse
I	10	242.72	258.90	-16.18	[-65.62, 33.25]	-6.3%	4/6	0.7695	1.0000	not resolved
J	10	591.77	522.69	69.07	[6.96, 131.19]	13.2%	2/8	0.0371	0.3711	VECTOR worse
K	10	453.06	341.84	111.22	[67.52, 154.91]	32.5%	1/9	0.0039	0.0586	VECTOR worse
L	10	187.62	181.97	5.65	[-55.00, 66.31]	3.1%	3/7	0.3750	1.0000	not resolved
M	10	170.01	181.97	-11.96	[-69.93, 46.01]	-6.6%	4/6	0.9219	1.0000	not resolved
N	10	292.34	290.81	1.53	[-56.34, 59.39]	0.5%	4/6	0.9219	1.0000	not resolved
O	10	326.53	261.46	65.07	[11.52, 118.63]	24.9%	1/9	0.0273	0.3008	VECTOR worse

Difference = VECTOR – baseline per seed (same demand, same events). For delay, negative favours VECTOR. "not resolved": the interval contains 0.

scenario	VECTOR	native	VECTOR – native [95% t-CI]	seeds better/worse	Holm p
R0	170.0	182.0	-12.0 [-69.9, 46.0]	4/6	1.000
A	330.1	302.6	+27.5 [-22.8, 77.7]	4/6	1.000
B	314.7	241.2	+73.6 [26.6, 120.5]	0/10	0.031
C	173.8	140.7	+33.1 [16.3, 49.8]	1/9	0.059
D	175.6	146.2	+29.4 [12.7, 46.2]	1/9	0.076
E	173.4	173.1	+0.3 [-57.7, 58.3]	2/8	1.000
F	174.1	162.1	+11.9 [-26.8, 50.7]	3/7	1.000
G	163.6	182.0	-18.3 [-73.4, 36.7]	5/5	1.000
H	458.1	406.9	+51.2 [19.9, 82.5]	2/8	0.117
I	242.7	258.9	-16.2 [-65.6, 33.3]	4/6	1.000
J	591.8	522.7	+69.1 [7.0, 131.2]	2/8	0.371
K	453.1	341.8	+111.2 [67.5, 154.9]	1/9	0.059
L	187.6	182.0	+5.7 [-55.0, 66.3]	3/7	1.000
M	170.0	182.0	-12.0 [-69.9, 46.0]	4/6	1.000
N	292.3	290.8	+1.5 [-56.3, 59.4]	4/6	1.000
O	326.5	261.5	+65.1 [11.5, 118.6]	1/9	0.301

P1 is in seconds per requested vehicle. Bootstrap intervals agree with the t-intervals (T2).

Reading the table:

- M ≡ R0 for both arms, because the shadow changes nothing and native is not exposed to the faults.
- G and L share native's R0 runs, because native uses SUMO's detectors, not the audit's sensor model.
- VECTOR's point estimate is lower than native's in 4 scenarios (R0, G, I, M), and none of those is resolved.

- With 10 pairs, only effects of about one SD of the per-seed difference are resolvable.
- Secondary endpoints (T5) point the same way. Under the congestion scenarios B, H, J, K and O, VECTOR has more ineffective and wasted green, more spillback and lower completion. Under K, completion is -10.4 pp [-15.7, -5.1].

4.3 Secondary and exploratory baselines

Against vehicle-actuated (secondary family, T3). There is no Holm rejection; the smallest is 0.0625, in A, J, L and N.

- Unadjusted, VECTOR is better in R0 (-24.9), A (-61.6), G (-20.2), J (-127.3), L (-42.9), M (-24.9) and N (-95.9 s/veh).
- It is worse in K (+35.6 [6.2, 64.9]).
- The vehicle-actuated rule starves approaches badly: its max red wait is 500–1500 s. VECTOR's starvation guard prevents that.

Against fixed-time 30 s (exploratory, T4). VECTOR is better in 15 of 16 scenarios, by 84–218 s/veh, and on 8–10 of 10 seeds in each. J is not resolved (-33.4 [-81.3, 14.5]).

Against the simplest controllers, VECTOR is a working adaptive controller. Against a well-configured actuated controller, it is not better.

4.4 Emergency vehicles

Detection is the simulator's ground truth. Ordinary vehicles do not yield. Only the signal timing can help.

scenario	vehicle	VECTOR	native	veh-act	fixed	VECTOR – native [95% CI]
C	ambulance	121 (10/10)	168	191	227	-46.5 [-62.1, -30.9]
D	ambulance	126 (10/10)	168	197	226	-42.0 [-54.1, -29.9]
D	fire engine	138 (10/10)	204	209	278	-65.7 [-87.5, -43.9]
E	ambulance	127 (10/10)	171	194	227	-43.8 [-59.4, -28.2]
E	fire engine	139 (10/10)	205	208	299	-66.7 [-87.8, -45.6]
E	police	111 (10/10)	150	152	213	-39.3 [-56.3, -22.3]
K	ambulance	169 (10/10)	171	225	422	-1.8 [-28.9, 25.3]
K	fire engine	253 (10/10)	391	474	319 (9/10)	-138.0 [-193.0, -83.0]
K	police	1572 (5/10)	1553 (7/10)	— (0/10)	1576 (2/10)	+107 [-362, 576] (arrived pairs only)
O	ambulance	116 (10/10)	177	172	197	-61.1 [-76.7, -45.5]
O	fire engine	150 (10/10)	215	198	299 (8/10)	-64.8 [-85.5, -44.1]

Values are mean travel time in seconds, with arrivals out of 10 where not every trip finished.

- **This is VECTOR's clearest strength.** In C, D, E and O every emergency vehicle is faster under VECTOR than under every baseline. That includes the conflicting arrivals in D and E (evidence E10–E12, E22).
- **The general-traffic cost** of those overrides is not separately resolved. C differs from R0 only by the ambulance, but per-seed outcomes are too sensitive to compare across scenarios (§5.5).
- **Under saturation the override does not guarantee arrival.** In K the police car (E4→W1) is still in the network at 2700 s in 5 of 10 VECTOR runs. VECTOR made on average 1075 priority decisions for it, without effect: a green signal cannot move a vehicle through a gridlocked downstream link. Native got it through in 7 of 10 runs (E26).
- The input contract carries **one** emergency vehicle per junction. When two are on a junction's approaches, the harness shows VECTOR only the one dispatched most recently (`max` over requested departure). The other is invisible to VECTOR until then. This is a harness choice forced by the single slot.

4.5 Runtime

The figures are from T8, measured on one 4-vCPU cloud machine with Python timing. There is no real-time guarantee.

arm	decide p50	p95	p99	max	36-junction cycle p95	wall per trial
VECTOR (+ shadow)	0.175 ms	1.983 ms	2.803 ms	224.7 ms	24.39 ms	55.9 s
VECTOR (no shadow)	0.021 ms	0.056 ms	0.082 ms	20.4 ms	1.73 ms	22.3 s

Control itself is cheap. The advisory shadow multiplies decide latency roughly 35-fold at p95, and it produces the 225 ms worst case. A cycle at p95 uses 2.4% of the 1 s step.

4.6 Safety

These checks run in every trial and cover all 800.

- **Signal timing.** 0 timing violations (min green, yellow, all-red, max green), 0 illegal states and 0 invalid requests. Every interval was rebuilt from SUMO's reported state, not from the command.
- **Every step,** SUMO's reported state equalled the commanded state.
- **Teleporting** is disabled, and no route was changed.
- **Collisions.** VECTOR had 0 collisions and 0 teleports in 320 trials.

The one collision. The vehicle-actuated arm in G, seed 95009, had two colliding vehicles and one collision teleport:

```
Teleporting vehicle 'v1341'; collision with vehicle 'v1330', lane='C41_C31_1', gap=-0.10, time=876.00, stage=move
```

- It reproduces byte for byte (same signal-sequence hash `1b903a85...`, same P1).
- It is a rear-end gap violation within a lane, not a junction conflict: illegal states are 0.
- The SUMO log is kept in `integrity/sumo_log_G_vehicle_actuated_95009.log`.

It is a car-following event in a baseline arm, and it is reported, not removed.

What this does not establish. The signal-safety constraints are enforced by the harness's `SignalAdapter`, and VECTOR's own `SignalSafetyGovernor` is not in the path. The result therefore shows that VECTOR's requests are safe under an external interlock, not that VECTOR is safe without one.

4.7 Shadow non-influence and module degradation

- **Non-influence.** The signal-sequence sha256 of `vector` equals that of `vector_noshadow` in 160 of 160 held-out pairs, and in 21 of 21 smoke pairs.
- **Scenario M.** The forecasters raise from 900 s, and the whole orchestrator raises from 1500 s.
 - 255 217 orchestrator failures were caught and counted, and the control sequence was unchanged (E19).
 - This shows **isolation, not graceful degradation**. The failed modules never influenced control, so there was nothing to degrade.
- **The stack does not see its own failure.** From 900 s to 1500 s, all 10 058 shadow decisions on the trace seed report `degraded: FULL, fallback: false` with **0 forecasts** (§5.9).

Module degradation: scenario M, seed 95000, junction C21

Injected: ☒ forecasters at t=900 ☐ orchestrator at t=1500 (every junction). Shadow failures counted: **25523**.

t (s)	advisory stack (shadow)	control law reason	executed
605	PREDATOR; 4 forecasts; degraded FULL; fallback false	VEHICLE_DEMAND (1 vehicles in lane_2; gap_out)	end NS green (yellow)
899	SENTRY; 1 forecasts; degraded FULL; fallback false	VEHICLE_DEMAND (1 vehicles in lane_1; gap_out)	end EW green (yellow)
912	PREDATOR; 0 forecasts; degraded FULL; fallback false	NO_DEMAND_REST_IN_GREEN	extend NS
1202	SENTRY; 0 forecasts; degraded FULL; fallback false	NO_DEMAND_LONGEST_RED	green NS
1504	RuntimeError: audit fault injection: advisory orchestrator unavailable	VEHICLE_DEMAND (2 vehicles in lane_2; gap_out)	end NS green (yellow)
1509	RuntimeError: audit fault injection: advisory orchestrator unavailable	VEHICLE_DEMAND (4 vehicles in lane_2)	green EW
1800	RuntimeError: audit fault injection: advisory orchestrator unavailable	NO_DEMAND_REST_IN_GREEN	extend EW

Signal sequence with the faults vs. production VECTOR without any shadow, same seed: **identical**
 25c6cbda9f43b8011fa3c26b... vs 25c6cbda9f43b8011fa3c26b...

What this shows: the advisory stack fails and is recorded, and control continues unchanged. What it does not show: a fallback that improves anything. VECTOR has no supported fallback for advice it never used.

4.8 Prediction (shadow only)

The forecast is the total queue at the junction, 8 s ahead, scored against what SUMO then measured (T7). Persistence is "the reading now".

- The shadow ensemble's MAE is **worse than persistence in 15 of 16 scenarios, on every seed** (0/10 each). Examples:
 - R0: 1.60 against 1.18 vehicles;
 - J: 2.34 against 1.50.
- The exception is G, on 10 of 10 seeds: 1.78 against 2.62. There the phantom spikes corrupt the reading that persistence repeats, and the ensemble smooths them.
- Concept shift, H.** Shadow MAE is 1.25 before 900 s and 2.60 after. Persistence goes from 1.11 to 1.64. Nothing adapts online (E05, E06).

No model is authorised to influence control (the training gate holds them all at UNTRAINED). The forecasts cost most of the decide latency and add no predictive skill over the trivial forecast.

The shadow is not reproducible. The orchestrator runs forecasters only while a 50 ms wall-clock compute budget lasts (`ComputeBudget` in `extremis/intelligence/orchestrator.py`). Which forecasts exist therefore depends on machine load.

- Re-running held-out `vector` trials reproduces every control, traffic, safety and emergency field exactly.
- The shadow's forecast counts and scores differ slightly between runs (`integrity/reproduction_check.json`).
- The T7 prediction numbers are thus valid for this machine and load, not bit-for-bit.
- Control is unaffected, because nothing reads the shadow.

4.9 Observation faults

- **Sensor noise, G.** VECTOR is not resolved against native (-18.3 [$-73.4, 36.7$]). It is better than vehicle-actuated (-20.2 [$-39.7, -0.8$]), which sees the same faults. The observer's ± 0.5 veh/s bound absorbs the phantom spikes. In E04, at C00, a shown jump to 33 vehicles (true value 6) moves the belief by 0.5. The same bound also delays real changes: in the same window the EW belief lags the true queue by 10–15 vehicles in both directions (§5.2).
- **Stale, swapped and invalid observations, L.** VECTOR is not resolved against native, and better than vehicle-actuated (-42.9 [$-60.3, -25.6$]). But **nothing in VECTOR detects the corruption.**
 - The adapter drops the stale flag.
 - At swapped junctions VECTOR serves the wrong approach. In E23, at C44, 32–42 vehicles queue on EW while VECTOR, shown them on NS, extends NS green to max green and then serves NS again.
 - Only the max-green and starvation limits bound the damage.

5. Failure analysis

5.1 Where VECTOR loses

Against native, VECTOR is worse (unadjusted) in seven scenarios:

- the congestion scenarios B, H, J, K and O, where queues stop discharging because a downstream link is full or blocked;
- the emergency scenarios C and D, which add one or two emergency vehicles to R0's demand. Part of that gap is per-seed sensitivity, not mechanism (§5.5). The fresh-seed replication tests it (§5.7).

It is not resolved in R0, A, E, F, G, I, L, M and N.

5.2 Mechanism: green held on a queue that cannot move

- **How the law reads demand.** `decide_signals` ranks approaches by halting vehicle count, the `QueueObserver` belief. It extends green while that approach still "has demand". A queue that cannot discharge stays a large halting count, so it keeps winning.
- **What the law lacks.** It has no term for discharge, downstream occupancy or spillback. No junction knows its neighbour's state.
- **How native differs.** Native gap-out logic ends green when its detectors see no vehicle crossing, so it yields a blocked approach early.
- **The held-out data are consistent with this.** Over the 160 paired cells (`DIAGNOSIS.md` D4):
 - VECTOR – native in P1 correlates with VECTOR – native in ineffective green at $r = 0.91$, in wasted green at $r = 0.92$, and in gridlock events at $r = 0.79$;
 - the fitted slope is about 15 s/veh more time loss per 1000 extra junction-seconds of ineffective green;
 - in R0, where nothing blocks, VECTOR's ineffective green is *lower* than native's.

This is correlation across cells, labelled exploratory. §5.7 tests the mechanism directly with VECTOR's own ablations. * **One junction, decision by decision** (scenario I, C41, evidence E15 and E27). Vehicles stalled at the stop line block both lanes of the C31→C41 approach from 641/662 s. * VECTOR serves that approach EW green for the full 45 s. * After clearance it serves EW again, because 29–34 vehicles are still halting there. * The queue grows during its own green, and the reason string labels it "discharging 29". * NS, with 2–7 vehicles that could move, waits. * **Observer lag adds to it** (`DIAGNOSIS.md` D5). The `QueueObserver` bounds every change to ± 0.5 veh/s. * On the trace seed, its belief is off the true queue by ≥ 5 vehicles in 3.2% of approach-decisions in R0. That rises to 6.0% in B, 8.2% in J and 10.1% in L. * The lag works both ways. It can hold a green after a queue has discharged, and it is slow to see a queue that has formed.

5.3 Starvation cap

VECTOR's longest red wait is 105 s in every trial; native's is 55 s (T5: +50 s in every scenario). That bound is a design choice: starvation at red ≥ 90 s, plus clearance. It prevents the vehicle-actuated rule's 500–1500 s starvation. But it lets VECTOR hold a side approach red twice as long as native.

5.4 Emergency priority under saturation

In K the override fires (on average 1075 priority decisions for the police car), but a green cannot clear a gridlocked downstream link. VECTOR has no routing, so it cannot send the vehicle around the jam. The other two emergency vehicles in K still arrive: the fire engine is 138 s faster than under native; the ambulance is not resolved.

5.5 Sensitivity of per-seed outcomes

C is R0 plus one ambulance. Under native, adding that single vehicle changes P1 by -214 to $+29$ s/veh across seeds; under VECTOR, by -40 to $+49$. The mechanism is that one extra vehicle shifts SUMO's random stream and the network's gridlock-or-not trajectory.

The consequences:

- Paired differences *within* a scenario are valid.
- Comparisons *across* scenarios that share demand are not: for example, "native improves when an ambulance is added" is not a real effect.
- Ten seeds sample this sensitivity coarsely, which is why the post-hoc replication (§5.7) matters.

5.6 Controller-dependent incident exposure (a validity threat, checked)

The blockages in B, F, I, N and O stall *the first vehicle found halted at the lane's stop line* inside the window. Their onset therefore depends on the controller (DIAGNOSIS.md D6).

- In B, VECTOR's trials had more blocked lane-steps than native's: 1498 against 1222 on average, more in 8 of 10 seeds. If that explained VECTOR's loss, B would partly measure exposure rather than control.
- It does not. Within B, VECTOR's excess time loss is *negatively* correlated with its excess exposure ($r = -0.44$), and the fitted loss at equal exposure is $+89$ s/veh. On seed 95009, VECTOR's C12→C22 blockage never engaged, and VECTOR still lost by 177 s/veh.
- In F, I, N and O, exposure differs by less than 6% on average.

This design choice is recorded as a limitation (§8). A fixed-onset blockage would remove it.

5.7 Post-hoc diagnosis on fresh seeds

Post-hoc and exploratory (AUDIT_PROTOCOL.md §13, DIAGNOSIS.md). The diagnosis ran after the held-out results were seen:

- on 10 fresh seeds, 96000–96009;
- in five scenarios: R0, B, C, J and K;
- on four arms: VECTOR without the shadow (identical control to `vector`), native, and VECTOR's two registered ablations, used unchanged.

200 of 200 trials completed, and all passed the health gate. VECTOR was not modified. No multiplicity correction is applied.

Replication: VECTOR – native, P1, fresh seeds.

scenario	held-out (§4.2)	fresh seeds	seeds better/worse (fresh)
R0	-12.0 [$-69.9, 46.0$]	$+21.1$ [4.3, 37.9]	1/9
B	$+73.6$ [26.6, 120.5]	$+23.7$ [$-22.6, 70.0$]	4/6

scenario	held-out (§4.2)	fresh seeds	seeds better/worse (fresh)
C	+33.1 [16.3, 49.8]	+17.9 [-1.8, 37.6]	2/8
J	+69.1 [7.0, 131.2]	+67.0 [16.9, 117.2]	2/8
K	+111.2 [67.5, 154.9]	+116.3 [92.5, 140.2]	0/10

- **The direction replicates everywhere.** On fresh seeds VECTOR is never better than native, and it is worse in R0, a result the held-out seeds had left unresolved.
- **The size does not always replicate.** K and J are stable. B is about a third of its held-out size, with an interval that includes 0. Ten seeds sample the network's sensitivity (§5.5) coarsely, so single-scenario magnitudes should be read with that in mind.

Ablations: what each VECTOR mechanism contributes (ablation – VECTOR, P1, s/veh; positive means the ablation is worse):

scenario	remove the queue observer	remove green commitment	without commitment, vs native
R0	+16.7 [8.8, 24.5]	-15.1 [-26.6, -3.7]	+6.0 [-6.1, 18.1]
B	+2.8 [-34.3, 40.0]	+8.9 [-43.0, 60.8]	+32.6 [-25.8, 91.0]
C	+14.2 [7.0, 21.5]	-17.1 [-27.8, -6.3]	+0.8 [-22.3, 23.9]
J	+13.0 [-27.0, 52.9]	+102.9 [54.1, 151.7]	+170.0 [125.5, 214.4]
K	+37.1 [13.0, 61.3]	+65.6 [44.3, 86.8]	+181.9 [157.1, 206.7]

- **The queue observer helps**, by 14–37 s/veh where resolved. It is not the cause of the losses.
- **Green commitment is a trade-off.**
 - At moderate demand (R0, C) it costs about 15–17 s/veh, and it is the main source of VECTOR's excess ineffective green there (–858 and –924 junction-seconds without it, D3). Without commitment, VECTOR cannot be distinguished from native in R0 and C.
 - Under saturation (J, K) commitment is worth 65–103 s/veh. Presumably it stops the law from switching on every queue fluctuation. That mechanism was not separately tested.
- **Under congestion no variant comes close to native.** In K every VECTOR variant is 116–182 s/veh worse, on 10 of 10 seeds. Neither mechanism is the problem. The law has no notion of discharge or downstream capacity (§5.2), and no ablation adds one.

These findings locate the deficit; they do not fix it. Any change to VECTOR motivated by them would have to be evaluated on new, pre-registered seeds.

5.8 Corrupted input is not detected

In L, VECTOR serves swapped approaches because it has no consistency check between approaches, over time or against neighbours (§4.9, E23). The `stale` flag the harness provides is dropped by the `ExtremisCanonical` adapter.

5.9 The advisory stack misreports its health

- **The symptom.** When every forecaster raises (M, 900–1500 s), the orchestrator still reports `degraded: FULL`, `fallback: false` and emits 0 forecasts.
- **What gets counted.** Only orchestrator-level exceptions (after 1500 s) are counted as `shadow_failures`. Forecaster failures are recorded per model in the registry, but they are not surfaced as a degraded state.
- **Consequence today:** none, because nothing reads the stack.
- **Consequence if the shadow were ever promoted:** its health signal could not be trusted.

6. Acceptance gates

gate	requirement	result	evidence
A environment	≥ 30 validated signalised junctions, connected, legal routes, reproducible scenarios, baselines verified	PASS: 36 junctions, strongly connected, 552 of 552 legal routes, 0 validation problems, harness = validated runner	topology/topology_report.json , integrity/equivalence_91001.json , E01, E02
B integration	modules mapped to code; control path traced end to end; influence verified where claimed; gaps listed	PASS. The audit gate passes; the integration finding is largely negative: 4 components influence control, the shadow has none (verified), and most named modules are absent (§2)	integrity/module_inventory.json , traces, E18, §2.2
C adversarial coverage	all 16 scenarios on all arms; ≥ 2 emergency kinds; combined stress; failures kept	PASS: 16 × 5 × 10, 3 emergency kinds plus random emergency vehicles, 0 combined, 0 failures, every attempt logged	trials/heldout/attempts.jsonl
D measurable outcomes	traffic, emergency, prediction, runtime and safety metrics, paired, with uncertainty	PASS	results/heldout/*.json , RESULTS_TABLES_heldout.md
E evidence	≥ 15 genuine screenshots with IDs, run links and hashes; raw logs kept	PASS: 27 genuine screenshots, E01–E27 (14 <code>sumo-gui</code> images from 8 held-out trial runs, 13 dashboard renderings of recorded results), each with an ID, run link, sim time, claim, limitation and a sha256 re-verified on disk; all 8 GUI runs reproduce their headless trials exactly; raw trials, traces and logs kept	evidence/EVIDENCE_MANIFEST.json , EVIDENCE_GALLERY.md
F reproducibility	commands, environment, commit, configuration and seeds recorded	PASS: commands, environment, commits (freeze 63bc12a), network hashes, scenario config hashes and seeds recorded. 8 held-out trials (5 arms) re-run at 8558862 : every control, traffic, safety and emergency field identical. Only VECTOR's shadow forecast counts vary, by 1–2 in ~50 000, because its compute budget is wall-clock (§4.8). One command runs the whole audit (<code>run_audit.py</code>)	§10, trials/heldout/preflight.json , integrity/reproduction_check.json

6.1 Evidence notes

- **Two kinds of image.**
 - `sumo-gui` screenshots, taken over TraCI during a held-out trial on seed
 1. Each GUI run was checked against the headless trial of the same scenario, arm and seed: the signal-sequence hash and P1 are equal, so each image shows exactly the reported experiment.
 - Screenshots of the audit dashboard rendering recorded result files. The sha256 of every source file is recorded.
- **No image is drawn, edited or composited.** Each is re-hashed on disk when the gallery is generated.
- **Captions were checked against the data and corrected where they were wrong.** The GUI captions were written in the capture plan *before* the runs. On inspection, six did not match the recorded state, and each was rewritten from the recorded state:
 - E07: the queue was at C01, not on every approach.
 - E08: the change was to green, not to yellow.
 - E13 and E14: the queue shown was on a different street.
 - E15: no spillback was visible at C31.

- E20: the network had not drained.

E13, E14 and E15 were also re-framed and re-captured, to show the blockage they describe. Re-captured images whose view did not change (E01, E07–E09, E16 and E20) are byte-identical to the first capture, because rendering is deterministic. The superseded first-capture files are not part of the evidence set. *

The full list with captions, claims, limitations, run links and hashes is in `EVIDENCE_GALLERY.md` and `evidence/EVIDENCE_MANIFEST.json`. * Simulator images: E01, E03, E07–E16, E20, E21. * Dashboard images: E02, E04–E06, E17–E19, E22–E27.

7. What VECTOR can and cannot do (simulation evidence only)

It can:

- control all 36 junctions of a connected grid for 45 simulated minutes, under 16 adversarial conditions, with no signal-safety violation behind an external interlock;
- beat fixed-time control almost everywhere;
- match or beat a simple vehicle-actuated rule in 15 of 16 scenarios. It is worse in K;
- give emergency vehicles trips 39–138 s faster than native control, and faster still than the other baselines. Under saturation (K) this does not hold for every vehicle;
- survive the failure of its whole advisory stack with no change to control;
- make decisions in about 0.06 ms without the shadow, or about 2 ms with it, at p95.

It cannot, or has not been shown to:

- outperform a well-configured actuated controller, in any scenario. It is worse under congestion, significantly so in B;
- handle downstream blockage: it has no discharge or spillback term and no coordination;
- detect stale, swapped or invalid observations;
- route or reroute any vehicle;
- forecast better than persistence (its shadow models);
- report its own degradation correctly, or produce reproducible advisory output (the shadow is wall-clock dependent);
- use perception: no camera or detector path was in the loop;
- run with any of the named higher-level modules (Sentinel, Serpentine, Aegis, Atlas, Punisher, Phoenix, Hyperion, Terminus, Anarchy, ORACLE, PULSAR, the fusion engines, the learned models) in the loop;
- run on real data or a real network. Neither was used.

8. Limitations

This is simulation, and only simulation. Nothing here establishes real-world safety, regulatory compliance, deployment readiness or operational effectiveness.

- **No perception in the loop.** "Observation" is SUMO's halting count per approach. In G and L it is corrupted by the audit's sensor model. The camera, detector and `perception_adapter` path exists only in `run_extremis.py` and was not exercised.
- **Emergency detection is ideal.** It is the simulator's ground truth, with confidence 0.95, no false positives and no missed vehicles. The input admits one emergency vehicle per junction. SUMO's blue-light device is off, so ordinary vehicles never yield; only the signals can help an emergency vehicle.
- **No rerouting exists**, in any arm. Every vehicle keeps its assigned route, which SUMO's `vehroute` output confirms. Scenario F's "rerouting" is therefore a test of its absence.
- **No coordination exists.** There are 36 independent controllers with no offsets or shared state. A corridor result is the sum of local decisions.
- **One synthetic network.** It is a regular grid with uniform turning rules, two lane counts and Poisson demand from fixed OD classes. There are no pedestrians, buses, protected-left phases, detectors with

real failure physics, or real counts.

- **The demand level was set with the baselines on two calibration seeds.** Several scenarios, by design, push the network past saturation (H, J, K for every controller).
- **Ten paired seeds per cell** resolve paired differences of about 1 SD. Smaller effects are "not resolved". Per-seed outcomes are sensitive to small perturbations (§5.5).
- **The native baseline is not exposed to injected sensor faults.** It uses SUMO's own detectors, so in G and L it is a reference, not a like-for-like comparison.
- **Latency is measured on one 4-vCPU cloud machine** with Python-level timing. There is no hard real-time guarantee, and there is no hardware-in-the-loop.
- **Prediction exists only in shadow**, with an 8 s horizon on total junction queue. No model is authorised to influence control. Nothing updates its parameters online. The forecasters keep a short history; that is state, not learning.
- **Shadow forecasts depend on wall-clock time.** The advisory orchestrator skips forecasters when its 50 ms budget runs out, so the prediction scores vary slightly with machine load (§4.8). Control is deterministic.
- **The safety interlock is the harness's.** VECTOR's own `SignalSafetyGovernor` was not in the path.
- **Incident onset depends on the controller.** A blockage stalls the first vehicle halted at its stop line, so exposure differs between arms. It was checked and does not explain the B result (§5.6), but a fixed-onset design would be cleaner.
- **The diagnosis is post-hoc** (§5.7). It uses fresh seeds and VECTOR's existing ablations, but it was designed after seeing the held-out results, so it is exploratory.

9. Pilot-readiness assessment

The evidence supports **neither connection to live public signals nor an uncontrolled field trial**. VECTOR is outperformed by standard actuated control under congestion. It cannot detect corrupted input. Its safety interlock in these runs was external. And none of this was tested on a real network or with real data.

The defensible next step is a **supervised research collaboration**, offline and shadow-only, in this order (`docs/pilot/PILOT_PLAN.md`):

1. **A data-sharing agreement** with a road authority or university transport group. Historical detector counts and signal timing plans for a corridor or small district would be enough.
2. **A calibrated simulation of that site.** Build the network, calibrate demand to counts, and validate the native or field controller against the field data before VECTOR is run at all.
3. **The comparison:** VECTOR against the site's own controller, on pre-registered seeds, with this audit's harness, safety layer and statistics, reviewed by the partner's engineers.
4. **Fix the known defects first**, each re-evaluated on fresh seeds:
 - a discharge or downstream term in the law (§5.2);
 - observation consistency checks (§5.8);
 - truthful degradation reporting (§5.9);
 - VECTOR's own safety governor in the path.
5. **Only after that, and with authorisation:** a closed test track or a hardware-in-the-loop bench with a real signal controller. VECTOR would still run in shadow, beside the certified controller.

Contacting partners, sending outreach and signing agreements are decisions for the project owner. None was done by this audit.

10. Reproduction

Environment.

- Ubuntu 24.04, Python 3.11.

- SUMO 1.28.0 from `requirements-sumo.txt`, installed with `pip install --target ~/.vector-deps/sumo-1.28.0 -r requirements-sumo.txt` (`stus/sumo/env.py` finds it).
- numpy 2.4 and scipy 1.17.
- For the GUI evidence: `xvfb-run`.
- For the dashboard captures and the PDF: Chromium with Playwright on Node 22, and python-markdown 3.7.
- The reference machine had 4 vCPUs and 15 GB RAM. The held-out run took 1.6 h of wall time with 4 workers (6.4 trial-hours).

Commands, from the repository root. One command runs every stage, in order, and logs each stage's exit code and duration to `artifacts/final-adversarial-audit/audit_run_log.jsonl`:

```
python audit/final_adversarial/run_audit.py          # all stages; --list shows them; --stages picks
some
```

The stages are these individual commands:

```
# 0. preflight: network hashes, >= 30 junctions, equivalence with the validated runner
python audit/final_adversarial/topology_report.py
python audit/final_adversarial/scenario_matrix.py
python audit/final_adversarial/equivalence.py --seed 91001
python audit/final_adversarial/module_inventory.py

# 1. held-out evaluation (refuses a dirty tree; records the commit in preflight.json)
python audit/final_adversarial/run_suite.py --phase heldout --seeds 95000-95009 --trace-seed 95000 --
workers 4

# 2. analysis and tables
python audit/final_adversarial/analyze.py --phase heldout
python audit/final_adversarial/make_tables.py --phase heldout

# 3. post-hoc diagnosis (fresh seeds)
python audit/final_adversarial/run_suite.py --phase diagnostic --seeds 96000-96009 \
    --scenarios R0 B C K J --arms vector_noshadow diag_no_observer diag_no_commitment native_actuated
python audit/final_adversarial/diagnose.py
python audit/final_adversarial/reproduce_check.py          # re-runs 8 held-out trials, compares every
field

# 4. evidence
xvfb-run -a -s "-screen 0 1920x1300x24" python audit/final_adversarial/capture_gui.py \
    --seed 95000 --out artifacts/final-adversarial-audit/evidence
python3 -m http.server 8899 --bind 127.0.0.1 &             # serves the repository for the dashboard
node audit/final_adversarial/capture_dashboard.js audit/final_adversarial/dashboard_plan.json \
    artifacts/final-adversarial-audit/evidence
python audit/final_adversarial/make_gallery.py

# 5. dossier PDF
python audit/final_adversarial/build_pdf.py
```

Determinism. A trial's control and traffic outcome is fully determined by its scenario, arm and seed. Trials run in separate processes and do not share state. SUMO runs with a fixed seed and teleporting disabled.

`reproduce_check.py` re-ran eight held-out trials, covering all five arms, at commit 8558862. Every control, traffic, safety and emergency field was identical to the stored result (`integrity/reproduction_check.json`).

The only differences were in VECTOR's shadow forecast scoring: 1–2 forecasts in about 50 000, changing scores in the third decimal. The cause is the advisory orchestrator's wall-clock compute budget (§4.8).

Each `sumo-gui` evidence run also reproduced its headless trial exactly: the same signal-sequence hash and the same P1.

11. Artifact index

what	where
protocol (frozen), trace schema, scenario matrix	docs/audit/final-adversarial/AUDIT_PROTOCOL.md, TRACE_SCHEMA.md, scenario_matrix.json
results tables (held-out, smoke), post-hoc diagnosis	RESULTS_TABLES_heldout.md, RESULTS_TABLES_smoke.md, DIAGNOSIS.md
evidence gallery and manifest	EVIDENCE_GALLERY.md; artifacts/final-adversarial-audit/evidence/EVIDENCE_MANIFEST.json
this dossier as PDF (committed; the HTML beside it is regenerated by <code>build_pdf.py</code>)	artifacts/final-adversarial-audit/dossier/FINAL_DOSSIER.pdf
trial summaries (every attempt)	artifacts/final-adversarial-audit/trials/{smoke, heldout, diagnostic}/
control-cycle traces (trace seed 95000, every arm)	artifacts/final-adversarial-audit/traces/heldout/
analysis outputs	artifacts/final-adversarial-audit/results/{heldout, smoke, diagnostic}/
network, integrity and calibration	artifacts/final-adversarial-audit/{topology, integrity, calibration}/
audit code	audit/final_adversarial/ (harness, scenarios, runner, analysis, evidence, dashboard)

Results tables (heldout)

Generated by `audit/final_adversarial/make_tables.py` from `artifacts/final-adversarial-audit/results/heldout/` and the trial files. 800 trials ok, 0 failed.

T1. Traffic outcomes by scenario and arm (mean $\pm$ SD over seeds)

P1 = mean time loss over all requested ordinary vehicles (s/veh); completion = share of ordinary trips finished by 2700 s; stranded = still running at the end; gridlock = junction episodes of ≥ 120 s with ≥ 5 halting and no crossing; ineff./wasted green in junction-seconds (definitions: AUDIT_PROTOCOL.md §8, §11).

scenario	arm	n	P1 time loss (s)	completion	stranded	gridlock	spillback (lane-s)	max red wait (s)	ineff. green (s)	wasted green (s)
R0	VECTOR	10	170.0 $\pm$ 32.0	97.7 $\pm$ 2.5%	78 $\pm$ 87	1.4 $\pm$ 2.5	4209 $\pm$ 2371	105 $\pm$ 0	1514 $\pm$ 1024	2465 $\pm$ 1073
R0	native	10	182.0 $\pm$ 66.3	92.3 $\pm$ 7.7%	252 $\pm$ 244	5.9 $\pm$ 5.1	6653 $\pm$ 4710	55 $\pm$ 0	2792 $\pm$ 2575	3553 $\pm$ 2618
R0	veh-act	10	194.9 $\pm$ 41.1	96.1 $\pm$ 4.4%	131 $\pm$ 149	2.5 $\pm$ 4.3	5083 $\pm$ 3010	514 $\pm$ 95	1992 $\pm$ 1755	2985 $\pm$ 1777
R0	fixed	10	382.2 $\pm$ 50.6	79.2 $\pm$ 6.5%	615 $\pm$ 182	12.5 $\pm$ 7.2	16198 $\pm$ 3390	40 $\pm$ 0	7698 $\pm$ 2834	17990 $\pm$ 2048
A	VECTOR	10	330.1 $\pm$ 39.0	83.9 $\pm$ 3.0%	503 $\pm$ 104	3.3 $\pm$ 2.8	13796 $\pm$ 2591	105 $\pm$ 0	5517 $\pm$ 1252	6745 $\pm$ 1313
A	native	10	302.6 $\pm$ 72.9	80.7 $\pm$ 7.5%	642 $\pm$ 255	6.3 $\pm$ 4.0	14184 $\pm$ 4429	55 $\pm$ 0	5226 $\pm$ 2719	5964 $\pm$ 2503
A	veh-act	10	391.7 $\pm$ 83.0	76.7 $\pm$ 8.6%	713 $\pm$ 242	15.1 $\pm$ 10.8	17102 $\pm$ 5146	845 $\pm$ 236	8728 $\pm$ 4340	9634 $\pm$ 3685
A	fixed	10	455.1 $\pm$ 60.3	74.4 $\pm$ 7.0%	765 $\pm$ 233	7.6 $\pm$ 4.8	19901 $\pm$ 4409	40 $\pm$ 0	8152 $\pm$ 2988	18842 $\pm$ 2378
B	VECTOR	10	314.7 $\pm$ 56.2	84.9 $\pm$ 7.1%	476 $\pm$ 200	7.2 $\pm$ 7.5	13171 $\pm$ 3318	105 $\pm$ 0	6590 $\pm$ 3144	7390 $\pm$ 3042
B	native	10	241.2 $\pm$ 34.0	89.4 $\pm$ 4.5%	338 $\pm$ 134	6.1 $\pm$ 5.4	10027 $\pm$ 2709	55 $\pm$ 0	3325 $\pm$ 1652	4262 $\pm$ 1588
B	veh-act	10	357.0 $\pm$ 85.5	79.9 $\pm$ 10.6%	644 $\pm$ 310	13.7 $\pm$ 9.4	15625 $\pm$ 5499	893 $\pm$ 342	8714 $\pm$ 4509	9507 $\pm$ 4199
B	fixed	10	477.9 $\pm$ 77.3	71.6 $\pm$ 7.4%	747 $\pm$ 198	16.8 $\pm$ 6.2	19607 $\pm$ 4799	40 $\pm$ 0	10438 $\pm$ 3608	20387 $\pm$ 3020
C	VECTOR	10	173.8 $\pm$ 32.4	97.1 $\pm$ 4.0%	98 $\pm$ 135	1.9 $\pm$ 3.8	4485 $\pm$ 2528	105 $\pm$ 0	1784 $\pm$ 1596	2699 $\pm$ 1500
C	native	10	140.7 $\pm$ 30.4	99.1 $\pm$ 1.5%	30 $\pm$ 52	0.8 $\pm$ 1.9	3468 $\pm$ 2323	55 $\pm$ 0	711 $\pm$ 622	1540 $\pm$ 700
C	veh-act	10	183.2 $\pm$ 29.9	98.5 $\pm$ 1.7%	52 $\pm$ 59	0.6 $\pm$ 1.3	4077 $\pm$ 1916	496 $\pm$ 70	1283 $\pm$ 713	2315 $\pm$ 905
C	fixed	10	375.4 $\pm$ 43.2	80.5 $\pm$ 6.1%	578 $\pm$ 164	11.8 $\pm$ 5.9	15744 $\pm$ 2584	40 $\pm$ 0	7279 $\pm$ 2515	17706 $\pm$ 1883
D	VECTOR	10	175.6 $\pm$ 32.5	97.4 $\pm$ 3.0%	90 $\pm$ 100	1.3 $\pm$ 2.8	4792 $\pm$ 2438	105 $\pm$ 0	1819 $\pm$ 1287	2736 $\pm$ 1286
D	native	10	146.2 $\pm$ 33.9	97.7 $\pm$ 2.9%	79 $\pm$ 99	1.3 $\pm$ 2.5	3903 $\pm$ 2509	55 $\pm$ 0	1100 $\pm$ 1024	1882 $\pm$ 991
D	veh-act	10	184.5 $\pm$ 37.6	97.9 $\pm$ 2.8%	72 $\pm$ 98	1.1 $\pm$ 2.1	4329 $\pm$ 2630	524 $\pm$ 50	1399 $\pm$ 956	2378 $\pm$ 1161
D	fixed	10	369.4 $\pm$ 49.9	81.7 $\pm$ 6.6%	541 $\pm$ 180	9.3 $\pm$ 6.8	15021 $\pm$ 3179	40 $\pm$ 0	6572 $\pm$ 2768	17256 $\pm$ 2045
E	VECTOR	10	173.4 $\pm$ 33.6	97.5 $\pm$ 3.2%	86 $\pm$ 108	1.2 $\pm$ 2.8	4584 $\pm$ 2510	105 $\pm$ 0	1788 $\pm$ 1427	2724 $\pm$ 1472

scenario	arm	n	P1 time loss (s)	completion	stranded	gridlock	spillback (lane-s)	max red wait (s)	ineff. green (s)	wasted green (s)
E	native	10	173.1 ± 72.1	94.0 ± 8.6%	192 ± 271	4.3 ± 5.8	5619 ± 5016	55 ± 0	2234 ± 2756	3015 ± 2784
E	veh-act	10	185.5 ± 38.4	97.0 ± 3.5%	102 ± 113	1.6 ± 2.4	4353 ± 2591	566 ± 95	1760 ± 1242	2678 ± 1283
E	fixed	10	367.6 ± 46.1	81.6 ± 6.0%	541 ± 161	9.8 ± 6.8	14943 ± 2876	40 ± 0	6743 ± 2599	17371 ± 2001
F	VECTOR	10	174.1 ± 38.6	98.1 ± 1.9%	66 ± 65	0.3 ± 0.7	4629 ± 2948	105 ± 0	1592 ± 1117	2574 ± 1242
F	native	10	162.1 ± 68.2	95.8 ± 7.8%	137 ± 253	2.7 ± 5.3	5070 ± 5145	55 ± 0	1734 ± 2611	2491 ± 2628
F	veh-act	10	188.0 ± 48.4	97.5 ± 3.7%	85 ± 126	0.8 ± 2.2	4407 ± 3247	527 ± 75	1526 ± 1354	2500 ± 1523
F	fixed	10	387.3 ± 54.4	79.2 ± 6.8%	628 ± 185	11.4 ± 7.9	16606 ± 3500	40 ± 0	7850 ± 2993	18129 ± 2149
G	VECTOR	10	163.6 ± 22.6	98.6 ± 2.1%	47 ± 69	0.7 ± 2.2	3698 ± 1673	105 ± 0	1232 ± 905	2385 ± 817
G	native	10	182.0 ± 66.3	92.3 ± 7.7%	252 ± 244	5.9 ± 5.1	6653 ± 4710	55 ± 0	2792 ± 2575	3553 ± 2618
G	veh-act	10	183.9 ± 34.7	96.8 ± 4.1%	109 ± 137	2.6 ± 4.8	4251 ± 2509	504 ± 47	1774 ± 1568	2854 ± 1434
G	fixed	10	382.2 ± 50.6	79.2 ± 6.5%	615 ± 182	12.5 ± 7.2	16198 ± 3390	40 ± 0	7698 ± 2834	17990 ± 2048
H	VECTOR	10	458.1 ± 34.5	65.4 ± 3.2%	1110 ± 49	18.8 ± 7.5	23609 ± 1836	105 ± 0	12015 ± 1917	13464 ± 1682
H	native	10	406.9 ± 34.7	71.3 ± 4.4%	892 ± 102	14.9 ± 7.6	20555 ± 2252	55 ± 0	6816 ± 2034	8056 ± 1814
H	veh-act	10	480.7 ± 21.8	63.0 ± 2.9%	1160 ± 54	29.4 ± 8.2	23036 ± 1665	810 ± 160	13557 ± 2220	14584 ± 1683
H	fixed	10	565.5 ± 43.1	59.8 ± 4.2%	1020 ± 69	12.7 ± 6.4	25088 ± 2414	40 ± 0	11279 ± 2755	21048 ± 2134
I	VECTOR	10	242.7 ± 78.7	92.4 ± 8.8%	250 ± 260	4.2 ± 10.5	8655 ± 4750	105 ± 0	4013 ± 3760	4899 ± 3577
I	native	10	258.9 ± 105.3	87.4 ± 11.6%	398 ± 334	7.4 ± 7.1	11364 ± 6376	55 ± 0	4687 ± 3731	5522 ± 3779
I	veh-act	10	242.2 ± 36.1	94.1 ± 3.3%	206 ± 117	2.2 ± 2.7	7836 ± 2650	550 ± 73	3099 ± 1306	4237 ± 1367
I	fixed	10	388.2 ± 50.8	79.1 ± 5.2%	648 ± 178	9.9 ± 5.1	16924 ± 3722	40 ± 0	7592 ± 2339	18035 ± 1987
J	VECTOR	10	591.8 ± 47.7	61.1 ± 4.5%	1284 ± 96	29.0 ± 9.6	34669 ± 2937	105 ± 0	17564 ± 3313	18316 ± 2676
J	native	10	522.7 ± 64.0	66.5 ± 6.5%	1135 ± 187	19.5 ± 10.4	31963 ± 4877	55 ± 0	10463 ± 4123	11574 ± 3611
J	veh-act	10	719.0 ± 77.3	50.6 ± 6.4%	1515 ± 127	38.8 ± 6.7	40756 ± 4160	1491 ± 377	24781 ± 4305	24570 ± 3408
J	fixed	10	625.2 ± 39.4	60.4 ± 3.8%	1208 ± 112	18.5 ± 6.5	32844 ± 2933	40 ± 0	13721 ± 2498	23412 ± 2170
K	VECTOR	10	453.1 ± 33.8	71.9 ± 3.6%	1036 ± 120	23.2 ± 8.4	27019 ± 2543	105 ± 0	13939 ± 1841	14543 ± 1540
K	native	10	341.8 ± 39.1	82.3 ± 4.9%	654 ± 175	9.0 ± 6.3	19343 ± 4014	55 ± 0	6136 ± 2346	6939 ± 2088
K	veh-act	10	417.5 ± 49.8	77.2 ± 6.3%	881 ± 194	22.0 ± 7.7	22696 ± 3580	892 ± 187	11137 ± 2997	12679 ± 2655
K	fixed	10	537.0 ± 52.7	66.9 ± 4.7%	1093 ± 124	21.6 ± 5.4	29667 ± 3641	40 ± 0	13395 ± 2691	23192 ± 2183

scenario	arm	n	P1 time loss (s)	completion	stranded	gridlock	spillback (lane-s)	max red wait (s)	ineff. green (s)	wasted green (s)
L	VECTOR	10	187.6 ± 37.3	97.7 ± 3.7%	78 ± 126	1.4 ± 3.5	4019 ± 2738	105 ± 0	1508 ± 1493	5221 ± 1549
L	native	10	182.0 ± 66.3	92.3 ± 7.7%	252 ± 244	5.9 ± 5.1	6653 ± 4710	55 ± 0	2792 ± 2575	3553 ± 2618
L	veh-act	10	230.6 ± 28.8	98.8 ± 1.7%	42 ± 57	2.5 ± 2.3	4840 ± 1747	925 ± 10	1624 ± 731	5046 ± 883
L	fixed	10	382.2 ± 50.6	79.2 ± 6.5%	615 ± 182	12.5 ± 7.2	16198 ± 3390	40 ± 0	7698 ± 2834	17990 ± 2048
M	VECTOR	10	170.0 ± 32.0	97.7 ± 2.5%	78 ± 87	1.4 ± 2.5	4209 ± 2371	105 ± 0	1514 ± 1024	2465 ± 1073
M	native	10	182.0 ± 66.3	92.3 ± 7.7%	252 ± 244	5.9 ± 5.1	6653 ± 4710	55 ± 0	2792 ± 2575	3553 ± 2618
M	veh-act	10	194.9 ± 41.1	96.1 ± 4.4%	131 ± 149	2.5 ± 4.3	5083 ± 3010	514 ± 95	1992 ± 1755	2985 ± 1777
M	fixed	10	382.2 ± 50.6	79.2 ± 6.5%	615 ± 182	12.5 ± 7.2	16198 ± 3390	40 ± 0	7698 ± 2834	17990 ± 2048
N	VECTOR	10	292.3 ± 32.2	88.0 ± 3.3%	400 ± 120	1.7 ± 3.1	12127 ± 2730	105 ± 0	4615 ± 1527	5760 ± 1522
N	native	10	290.8 ± 78.2	84.5 ± 8.4%	520 ± 284	7.8 ± 6.6	13812 ± 5644	55 ± 0	4730 ± 3149	5714 ± 2992
N	veh-act	10	388.2 ± 93.2	78.4 ± 8.9%	698 ± 252	15.4 ± 11.2	17039 ± 5805	875 ± 345	8973 ± 4532	10112 ± 4240
N	fixed	10	447.6 ± 69.2	75.5 ± 6.7%	752 ± 198	8.5 ± 4.0	20656 ± 4243	40 ± 0	8329 ± 2921	19017 ± 2202
O	VECTOR	10	326.5 ± 68.3	81.7 ± 8.4%	610 ± 286	8.2 ± 7.9	14478 ± 5025	105 ± 0	7446 ± 4146	8528 ± 3990
O	native	10	261.5 ± 44.1	86.7 ± 4.8%	445 ± 176	6.8 ± 5.1	11694 ± 2863	55 ± 0	3746 ± 1796	4714 ± 1689
O	veh-act	10	372.1 ± 94.3	78.5 ± 10.7%	663 ± 306	15.5 ± 11.7	15782 ± 5573	882 ± 448	8636 ± 5122	9671 ± 4812
O	fixed	10	455.0 ± 100.2	74.3 ± 10.1%	733 ± 262	9.7 ± 8.6	19858 ± 5578	40 ± 0	8926 ± 4469	19412 ± 3384

T2. Primary: VECTOR – native actuated, P1 (paired by seed, Holm over 16 scenarios)

scenario	n	VECTOR	baseline	difference	95% CI (t)	95% CI (bootstrap)	relative	better/worse seeds	Wilcoxon p	Holm p	at 95%
R0	10	170.0	182.0	-12.0	[-69.9, 46.0]	[-63.6, 29.9]	-6.6%	4/6	0.9219	1.0000	not resolved
A	10	330.1	302.6	27.5	[-22.8, 77.7]	[-14.5, 66.4]	9.1%	4/6	0.2754	1.0000	not resolved
B	10	314.7	241.2	73.6	[26.6, 120.5]	[37.9, 113.8]	30.5%	0/10	0.0020	0.0312	VECTOR worse
C	10	173.8	140.7	33.1	[16.3, 49.8]	[19.4, 46.5]	23.5%	1/9	0.0039	0.0586	VECTOR worse
D	10	175.6	146.2	29.4	[12.7, 46.2]	[15.9, 43.3]	20.2%	1/9	0.0059	0.0762	VECTOR worse
E	10	173.4	173.1	0.3	[-57.7, 58.3]	[-51.8, 40.8]	0.2%	2/8	0.4316	1.0000	not resolved
F	10	174.1	162.1	11.9	[-26.8, 50.7]	[-21.2, 42.4]	7.4%	3/7	0.3223	1.0000	not resolved
G	10	163.6	182.0	-18.3	[-73.4, 36.7]	[-68.8, 18.9]	-10.1%	5/5	0.8457	1.0000	not resolved

scenario	n	VECTOR	baseline	difference	95% CI (t)	95% CI (bootstrap)	relative	better/worse seeds	Wilcoxon p	Holm p	at 95%
H	10	458.1	406.9	51.2	[19.9, 82.5]	[24.4, 76.9]	12.6%	2/8	0.0098	0.1172	VECTOR worse
I	10	242.7	258.9	-16.2	[-65.6, 33.3]	[-61.4, 16.7]	-6.2%	4/6	0.7695	1.0000	not resolved
J	10	591.8	522.7	69.1	[7.0, 131.2]	[16.0, 116.9]	13.2%	2/8	0.0371	0.3711	VECTOR worse
K	10	453.1	341.8	111.2	[67.5, 154.9]	[73.5, 144.7]	32.5%	1/9	0.0039	0.0586	VECTOR worse
L	10	187.6	182.0	5.7	[-55.0, 66.3]	[-47.8, 50.5]	3.1%	3/7	0.3750	1.0000	not resolved
M	10	170.0	182.0	-12.0	[-69.9, 46.0]	[-63.6, 29.9]	-6.6%	4/6	0.9219	1.0000	not resolved
N	10	292.3	290.8	1.5	[-56.3, 59.4]	[-47.2, 48.5]	0.5%	4/6	0.9219	1.0000	not resolved
O	10	326.5	261.5	65.1	[11.5, 118.6]	[20.8, 108.3]	24.9%	1/9	0.0273	0.3008	VECTOR worse

T3. Secondary: VECTOR – vehicle-actuated, P1 (Holm over 16 scenarios)

scenario	n	VECTOR	baseline	difference	95% CI (t)	95% CI (bootstrap)	relative	better/worse seeds	Wilcoxon p	Holm p	at 95%
R0	10	170.0	194.9	-24.9	[-47.3, -2.6]	[-42.6, -5.8]	-12.8%	8/2	0.0488	0.5371	VECTOR better
A	10	330.1	391.7	-61.6	[-104.8, -18.4]	[-97.5, -27.4]	-15.7%	9/1	0.0039	0.0625	VECTOR better
B	10	314.7	357.0	-42.3	[-105.3, 20.7]	[-97.0, 7.2]	-11.8%	5/5	0.3223	1.0000	not resolved
C	10	173.8	183.2	-9.4	[-26.5, 7.8]	[-22.4, 5.3]	-5.1%	7/3	0.2324	1.0000	not resolved
D	10	175.6	184.5	-8.8	[-28.0, 10.3]	[-24.5, 7.0]	-4.8%	6/4	0.4316	1.0000	not resolved
E	10	173.4	185.5	-12.1	[-34.3, 10.1]	[-29.2, 7.1]	-6.5%	7/3	0.2754	1.0000	not resolved
F	10	174.1	188.0	-13.9	[-42.3, 14.4]	[-35.1, 11.2]	-7.4%	7/3	0.2324	1.0000	not resolved
G	10	163.6	183.9	-20.2	[-39.7, -0.8]	[-35.8, -3.7]	-11.0%	8/2	0.0645	0.5801	VECTOR better
H	10	458.1	480.7	-22.6	[-56.6, 11.4]	[-51.1, 4.9]	-4.7%	6/4	0.2754	1.0000	not resolved
I	10	242.7	242.2	0.5	[-54.0, 55.0]	[-35.2, 51.3]	0.2%	7/3	0.3223	1.0000	not resolved
J	10	591.8	719.0	-127.3	[-186.7, -67.8]	[-178.7, -81.1]	-17.7%	9/1	0.0039	0.0625	VECTOR better
K	10	453.1	417.5	35.6	[6.2, 64.9]	[11.2, 59.2]	8.5%	2/8	0.0273	0.3281	VECTOR worse
L	10	187.6	230.6	-42.9	[-60.3, -25.6]	[-56.3, -27.8]	-18.6%	9/1	0.0039	0.0625	VECTOR better
M	10	170.0	194.9	-24.9	[-47.3, -2.6]	[-42.6, -5.8]	-12.8%	8/2	0.0488	0.5371	VECTOR better
N	10	292.3	388.2	-95.9	[-154.2, -37.6]	[-146.0, -51.0]	-24.7%	8/2	0.0098	0.1270	VECTOR better
O	10	326.5	372.1	-45.5	[-132.8, 41.7]	[-119.2, 24.5]	-12.2%	5/5	0.3750	1.0000	not resolved

T4. VECTOR – fixed-time, P1 (exploratory)

scenario	n	VECTOR	baseline	difference	95% CI (t)	95% CI (bootstrap)	relative	better/worse seeds	Wilcoxon p	Holm p	at 95%
R0	10	170.0	382.2	-212.2	[-260.9, -163.5]	[-253.9, -174.4]	-55.5%	10/0	0.0020	—	VECTOR better
A	10	330.1	455.1	-125.1	[-152.1, -98.1]	[-148.2, -104.0]	-27.5%	10/0	0.0020	—	VECTOR better
B	10	314.7	477.9	-163.1	[-220.6, -105.6]	[-207.6, -113.5]	-34.1%	10/0	0.0020	—	VECTOR better
C	10	173.8	375.4	-201.6	[-249.7, -153.4]	[-242.3, -164.0]	-53.7%	10/0	0.0020	—	VECTOR better
D	10	175.6	369.4	-193.8	[-247.7, -139.9]	[-240.7, -153.0]	-52.5%	10/0	0.0020	—	VECTOR better
E	10	173.4	367.6	-194.2	[-247.1, -141.3]	[-240.7, -154.7]	-52.8%	10/0	0.0020	—	VECTOR better
F	10	174.1	387.3	-213.2	[-258.1, -168.3]	[-252.3, -178.9]	-55.0%	10/0	0.0020	—	VECTOR better
G	10	163.6	382.2	-218.5	[-263.3, -173.8]	[-255.4, -182.5]	-57.2%	10/0	0.0020	—	VECTOR better
H	10	458.1	565.5	-107.4	[-141.5, -73.3]	[-136.5, -81.1]	-19.0%	10/0	0.0020	—	VECTOR better
I	10	242.7	388.2	-145.5	[-217.1, -73.8]	[-193.8, -79.2]	-37.5%	9/1	0.0059	—	VECTOR better
J	10	591.8	625.2	-33.4	[-81.3, 14.5]	[-72.7, 6.4]	-5.3%	8/2	0.1602	—	not resolved
K	10	453.1	537.0	-84.0	[-113.9, -54.0]	[-107.3, -58.7]	-15.6%	10/0	0.0020	—	VECTOR better
L	10	187.6	382.2	-194.6	[-245.7, -143.4]	[-237.4, -154.4]	-50.9%	10/0	0.0020	—	VECTOR better
M	10	170.0	382.2	-212.2	[-260.9, -163.5]	[-253.9, -174.4]	-55.5%	10/0	0.0020	—	VECTOR better
N	10	292.3	447.6	-155.3	[-201.6, -108.9]	[-196.4, -120.7]	-34.7%	10/0	0.0020	—	VECTOR better
O	10	326.5	455.0	-128.5	[-207.5, -49.5]	[-198.5, -66.0]	-28.2%	8/2	0.0098	—	VECTOR better

T5. Secondary endpoints: VECTOR – native actuated, mean difference [95% t-CI] (exploratory)

scenario	completion (pp)	p95 time loss (s)	gridlock	spillback (lane-s)	max red wait (s)	ineff. green (s)	wasted green (s)
R0	5.5 [-1.0, 11.9]	-159.6 [-442.7, 123.5]	-4.5 [-9.0, -0.0]	-2444 [-6585, 1697]	50 [50, 50]	-1279 [-3517, 959]	-1088 [-3365, 1188]
A	3.2 [-2.2, 8.6]	-42.2 [-196.5, 112.0]	-3.0 [-6.6, 0.6]	-388 [-3652, 2876]	50 [50, 50]	291 [-1768, 2350]	780 [-1133, 2693]
B	-4.6 [-11.3, 2.1]	96.7 [-84.6, 277.9]	1.1 [-6.4, 8.6]	3144 [-257, 6545]	50 [50, 50]	3265 [571, 5959]	3129 [530, 5728]
C	-2.0 [-4.9, 0.8]	79.6 [-15.7, 174.8]	1.1 [-2.1, 4.3]	1016 [-563, 2596]	50 [50, 50]	1072 [8, 2137]	1160 [245, 2074]
D	-0.3 [-2.9, 2.2]	16.6 [-104.3, 137.5]	0.0 [-2.9, 2.9]	889 [-374, 2153]	50 [50, 50]	719 [-261, 1700]	854 [12, 1696]
E	3.4 [-3.6, 10.5]	-103.1 [-391.9, 185.7]	-3.1 [-8.0, 1.8]	-1035 [-5078, 3007]	50 [50, 50]	-445 [-2863, 1972]	-291 [-2740, 2158]
F	2.3 [-2.5, 7.0]	9.2 [-155.9, 174.2]	-2.4 [-6.3, 1.5]	-440 [-3513, 2632]	50 [50, 50]	-142 [-1751, 1466]	82 [-1445, 1609]

scenario	completion (pp)	p95 time loss (s)	gridlock	spillback (lane - s)	max red wait (s)	ineff. green (s)	wasted green (s)
G	6.4 [1.0, 11.7]	-192.3 [-453.1, 68.4]	-5.2 [-8.6, -1.8]	-2955 [-6730, 820]	50 [50, 50]	-1560 [-3443, 323]	-1168 [-3157, 821]
H	-5.9 [-9.9, -1.9]	36.1 [-2.5, 74.8]	3.9 [-4.8, 12.6]	3054 [1046, 5061]	50 [50, 50]	5199 [3055, 7343]	5408 [3524, 7291]
I	5.0 [-1.3, 11.3]	-203.4 [-385.0, -21.8]	-3.2 [-10.6, 4.2]	-2709 [-5981, 564]	50 [50, 50]	-674 [-3160, 1813]	-622 [-2942, 1697]
J	-5.3 [-12.0, 1.4]	86.1 [-42.8, 215.0]	9.5 [-1.7, 20.7]	2706 [-2281, 7694]	50 [50, 50]	7100 [2709, 11491]	6742 [3049, 10436]
K	-10.4 [-15.7, -5.1]	112.1 [28.4, 195.7]	14.2 [4.8, 23.6]	7676 [3502, 11850]	50 [50, 50]	7803 [5099, 10508]	7604 [5348, 9859]
L	5.4 [-1.7, 12.5]	-150.0 [-437.6, 137.5]	-4.5 [-9.9, 0.9]	-2633 [-7010, 1743]	50 [50, 50]	-1284 [-3759, 1190]	1668 [-848, 4183]
M	5.5 [-1.0, 11.9]	-159.6 [-442.7, 123.5]	-4.5 [-9.0, -0.0]	-2444 [-6585, 1697]	50 [50, 50]	-1279 [-3517, 959]	-1088 [-3365, 1188]
N	3.5 [-3.3, 10.3]	-179.1 [-400.5, 42.3]	-6.1 [-12.1, -0.1]	-1685 [-6012, 2641]	50 [50, 50]	-115 [-2820, 2589]	47 [-2537, 2630]
O	-5.0 [-12.2, 2.1]	116.4 [-96.7, 329.5]	1.4 [-6.1, 8.9]	2783 [-1210, 6776]	50 [50, 50]	3700 [423, 6977]	3814 [716, 6912]

T6. Named emergency vehicles: travel time (s), mean over seeds; arrived/n

scenario	vehicle	VECTOR	native	veh-act	fixed	VECTOR – native [95% CI]	VECTOR priority decisions (mean)
C	amb1	121 (10/10)	168 (10/10)	191 (10/10)	227 (10/10)	-46.5 [-62.1, -30.9]	48
D	amb1	126 (10/10)	168 (10/10)	197 (10/10)	226 (10/10)	-42.0 [-54.1, -29.9]	46
D	fire1	138 (10/10)	204 (10/10)	209 (10/10)	278 (10/10)	-65.7 [-87.5, -43.9]	60
E	amb1	127 (10/10)	171 (10/10)	194 (10/10)	227 (10/10)	-43.8 [-59.4, -28.2]	43
E	fire1	139 (10/10)	205 (10/10)	208 (10/10)	299 (10/10)	-66.7 [-87.8, -45.6]	61
E	pol1	111 (10/10)	150 (10/10)	152 (10/10)	213 (10/10)	-39.3 [-56.3, -22.3]	42
K	amb1	169 (10/10)	171 (10/10)	225 (10/10)	422 (10/10)	-1.8 [-28.9, 25.3]	85
K	fire1	253 (10/10)	391 (10/10)	474 (10/10)	319 (9/10)	-138.0 [-193.0, -83.0]	149
K	pol1	1572 (5/10)	1553 (7/10)	— (0/10)	1576 (2/10)	107.0 [-362.3, 576.3]	1075
O	amb1	116 (10/10)	177 (10/10)	172 (10/10)	197 (10/10)	-61.1 [-76.7, -45.5]	43
O	fire1	150 (10/10)	215 (10/10)	198 (10/10)	299 (8/10)	-64.8 [-85.5, -44.1]	52

T7. VECTOR's shadow forecasts against SUMO (total queue at the junction, horizon 8 s; means over seeds)

scenario	forecasts	MAE	RMSE	bias	persistence MAE	persistence RMSE	seeds where shadow beats persistence
R0	497144	1.595	3.289	0.066	1.183	2.210	0/10
A	519710	2.235	4.491	0.127	1.488	2.717	0/10
B	525085	1.952	4.074	0.031	1.325	2.505	0/10

scenario	forecasts	MAE	RMSE	bias	persistence MAE	persistence RMSE	seeds where shadow beats persistence
C	496845	1.592	3.349	0.048	1.177	2.220	0/10
D	497230	1.637	3.391	0.078	1.206	2.256	0/10
E	497477	1.589	3.295	0.061	1.180	2.203	0/10
F	498280	1.617	3.290	0.089	1.198	2.224	0/10
G	499000	1.777	3.604	-0.136	2.620	6.239	10/10
H	548014	2.246	4.535	-0.085	1.500	2.789	0/10
I	510767	1.859	3.836	0.122	1.288	2.388	0/10
J	567307	2.344	4.682	-0.016	1.500	2.808	0/10
K	537981	2.011	4.013	-0.063	1.381	2.564	0/10
L	560941	2.008	3.882	-0.110	1.724	3.268	0/10
M	144341	1.254	2.194	-0.015	1.113	1.878	0/10
N	519907	2.310	4.716	0.170	1.504	2.749	0/10
O	520755	1.872	3.834	0.008	1.286	2.403	0/10

Concept shift (H, from 900 s): shadow MAE 1.254 before → 2.601 after; persistence 1.113 → 1.639. By 300 s bin: 0: 0.668, 1200: 3.645, 1500: 3.259, 1800: 2.790, 2100: 1.955, 2400: 1.694, 300: 1.442, 600: 1.725, 900: 2.760.

T8. Runtime (per trial, then mean over trials) and safety totals by arm

arm	trials	decide p50 (ms)	decide p95	decide p99	decide max	36-junction cycle p95 (ms)	wall (s)	timing viol.	illegal states	invalid req.	teleports	collisions	route changes	health not ok
VECTOR	160	0.175	1.983	2.803	224.7	24.390	55.9	0	0	0	0	0	0	0
native	160	—	—	—	0.0	—	18.8	0	0	0	0	0	0	0
veh-act	160	0.000	0.001	0.002	6.8	0.640	20.7	0	0	0	1	2	0	1
fixed	160	0.000	0.001	0.002	6.1	0.986	21.9	0	0	0	0	0	0	0
VECTOR (no shadow)	160	0.021	0.056	0.082	20.4	1.728	22.3	0	0	0	0	0	0	0

T9. Integrity

- Trials ok: 800; failed: 0.
- Shadow non-influence: identical signal sequences in 160 of 160 paired trials (vector vs vector_noshadow).
- Shadow failures by scenario: {"A": 0, "B": 0, "C": 0, "D": 0, "E": 0, "F": 0, "G": 0, "H": 0, "I": 0, "J": 0, "K": 0, "L": 0, "M": 255217, "N": 0, "O": 0, "R0": 0}.
- Health not ok: [['G', 'vehicle_actuated', 95009]].

Post-hoc diagnosis (after the held-out evaluation)

Generated by `audit/final_adversarial/diagnose.py`. **Post-hoc and exploratory.** Run after the held-out results were seen, on fresh seeds 96000–96009 that no earlier step used, with VECTOR's own registered ablations (`extremis_canonical_no_observer`, `extremis_canonical_no_commitment`) unchanged. Nothing here changes or re-tests a held-out result; no multiplicity correction is applied. Protocol: `AUDIT_PROTOCOL.md` §13.

Trials whose health gate failed: none.

D1. Means by arm (fresh seeds)

scenario	arm	n	P1 time loss (s)	completion	gridlock	ineff. green (s)	wasted green (s)
R0	vector_noshadow	10	154.3	98.4%	0.7	1146	2043
R0	diag_no_observer	10	170.9	97.7%	1.4	1435	2367
R0	diag_no_commitment	10	139.1	99.9%	0.0	288	1004
R0	native_actuated	10	133.1	99.1%	0.8	600	1377
B	vector_noshadow	10	294.6	85.7%	8.2	6051	6963
B	diag_no_observer	10	297.4	87.3%	3.5	5323	6395
B	diag_no_commitment	10	303.5	84.4%	3.0	5227	5856
B	native_actuated	10	270.9	85.9%	8.1	4356	5227
C	vector_noshadow	10	160.5	98.3%	0.8	1413	2385
C	diag_no_observer	10	174.8	97.7%	0.0	1350	2310
C	diag_no_commitment	10	143.5	99.4%	0.0	489	1205
C	native_actuated	10	142.7	97.5%	2.7	1033	1772
J	vector_noshadow	10	579.0	61.9%	24.7	16774	17562
J	diag_no_observer	10	592.0	61.7%	22.2	16180	17424
J	diag_no_commitment	10	681.9	53.0%	17.8	20299	20252
J	native_actuated	10	512.0	67.5%	19.3	9720	10912
K	vector_noshadow	10	449.0	72.5%	23.7	13461	14339
K	diag_no_observer	10	486.1	69.1%	29.3	15063	16026
K	diag_no_commitment	10	514.6	65.4%	10.7	16258	16624
K	native_actuated	10	332.7	83.6%	10.8	5834	6826

D2. Paired differences on P1 (s/veh): mean [95% t-CI] (seeds where the first arm is better/worse)

scenario	VECTOR – native (replication)	no observer – VECTOR	no commitment – VECTOR	no observer – native	no commitment – native
R0	+21.1 [4.3, 37.9] (1/9)	+16.7 [8.8, 24.5] (1/9)	-15.1 [-26.6, -3.7] (9/1)	+37.8 [17.9, 57.8] (1/9)	+6.0 [-6.1, 18.1] (1/9)
B	+23.7 [-22.6, 70.0] (4/6)	+2.8 [-34.3, 40.0] (5/5)	+8.9 [-43.0, 60.8] (5/5)	+26.5 [-28.4, 81.5] (4/6)	+32.6 [-25.8, 91.0] (3/7)
C	+17.9 [-1.8, 37.6] (2/8)	+14.2 [7.0, 21.5] (1/9)	-17.1 [-27.8, -6.3] (9/1)	+32.1 [10.6, 53.5] (1/9)	+0.8 [-22.3, 23.9] (4/6)
J	+67.0 [16.9, 117.2] (2/8)	+13.0 [-27.0, 52.9] (5/5)	+102.9 [54.1, 151.7] (1/9)	+80.0 [17.0, 143.0] (2/8)	+170.0 [125.5, 214.4] (0/10)
K	+116.3 [92.5, 140.2] (0/10)	+37.1 [13.0, 61.3] (1/9)	+65.6 [44.3, 86.8] (0/10)	+153.5 [137.4, 169.5] (0/10)	+181.9 [157.1, 206.7] (0/10)

D3. Same contrasts on ineffective green (junction-seconds)

scenario	VECTOR – native (replication)	no observer – VECTOR	no commitment – VECTOR	no observer – native	no commitment – native
R0	+545.9 [5.9, 1085.9] (2/8)	+289.0 [-399.8, 977.8] (4/6)	-858.3 [-1411.8, -304.8] (10/0)	+834.9 [-200.2, 1870.0] (2/8)	-312.4 [-513.8, -111.0] (9/1)
B	+1695.3 [-410.2, 3800.8] (4/6)	-728.2 [-3543.7, 2087.3] (7/3)	-824.8 [-3980.8, 2331.2] (6/4)	+967.1 [-1802.9, 3737.1] (5/5)	+870.5 [-2133.7, 3874.7] (4/6)
C	+380.0 [-646.7, 1406.7] (4/6)	-62.7 [-532.7, 407.3] (5/5)	-923.5 [-1577.9, -269.1] (10/0)	+317.3 [-448.5, 1083.1] (3/7)	-543.5 [-1310.6, 223.6] (8/2)
J	+7053.8 [3677.1, 10430.5] (1/9)	-593.7 [-3226.7, 2039.3] (5/5)	+3524.9 [707.8, 6342.0] (2/8)	+6460.1 [2926.2, 9994.0] (0/10)	+10578.7 [8126.1, 13031.3] (0/10)
K	+7626.9 [5693.4, 9560.4] (0/10)	+1601.7 [133.8, 3069.6] (2/8)	+2796.6 [1495.6, 4097.6] (1/9)	+9228.6 [8117.9, 10339.3] (0/10)	+10423.5 [8904.1, 11942.9] (0/10)

D4. Mechanism check on the held-out data (no new runs)

Over the 160 paired (scenario, seed) cells, VECTOR – native:

- P1 vs ineffective green: $r = 0.906$; about 15.45 s/veh more time loss per 1000 extra junction-seconds of ineffective green (intercept 4.7 s).
- P1 vs wasted green: $r = 0.917$.
- P1 vs gridlock events: $r = 0.794$.

Correlation across cells is consistent with, not proof of, the mechanism.

Sensitivity: P1 in C minus P1 in R0, same seed (C adds one ambulance to R0's demand):

- vector: [1.2, -39.9, -0.9, 14.0, 25.4, -2.4, -2.9, 48.6, 4.1, -9.3] (range -39.9 to 48.6 s/veh)
- native_actuated: [-61.6, 9.4, -213.7, -37.5, 0.8, -42.5, -44.0, -51.6, 29.3, -1.0] (range -213.7 to 29.3 s/veh)
- vehicle_actuated: [2.2, 0.4, 11.0, -50.2, -22.8, -14.8, 9.7, -30.0, -21.9, -1.3] (range -50.2 to 11.0 s/veh)

D5. QueueObserver belief against the true queue (held-out traces, seed 95000)

Per approach and decision: $|\text{belief} - \text{SUMO halting}|$. The observer bounds every change to ± 0.5 veh/s, which absorbs phantom spikes and also delays real queue changes.

scenario	n	MAE (veh)	p95 (veh)	share off by ≥ 5 veh
R0	101318	0.48	3.0	3.2%
A	102936	0.76	5.0	5.3%
B	108590	0.81	5.5	6.0%
C	101264	0.41	2.5	2.9%
D	103390	0.41	2.5	2.7%
E	101444	0.40	2.5	2.7%
F	101928	0.42	2.5	2.9%
G	101928	0.68	4.0	3.8%
H	112134	0.63	4.0	4.4%
I	103466	0.68	4.5	4.8%
J	115332	1.08	7.5	8.2%
K	111420	0.59	4.0	4.2%
L	113162	1.59	9.0	10.1%
M	101318	0.48	3.0	3.2%
N	104340	0.75	5.0	5.2%

scenario	n	MAE (veh)	p95 (veh)	share off by ≥ 5 veh
O	106636	0.71	4.5	4.9%

D6. Blockage exposure (held-out seeds)

A blockage stalls the first vehicle found halted at the lane's stop line inside the window, so its onset depends on the controller. Mean blocked lane-steps per trial:

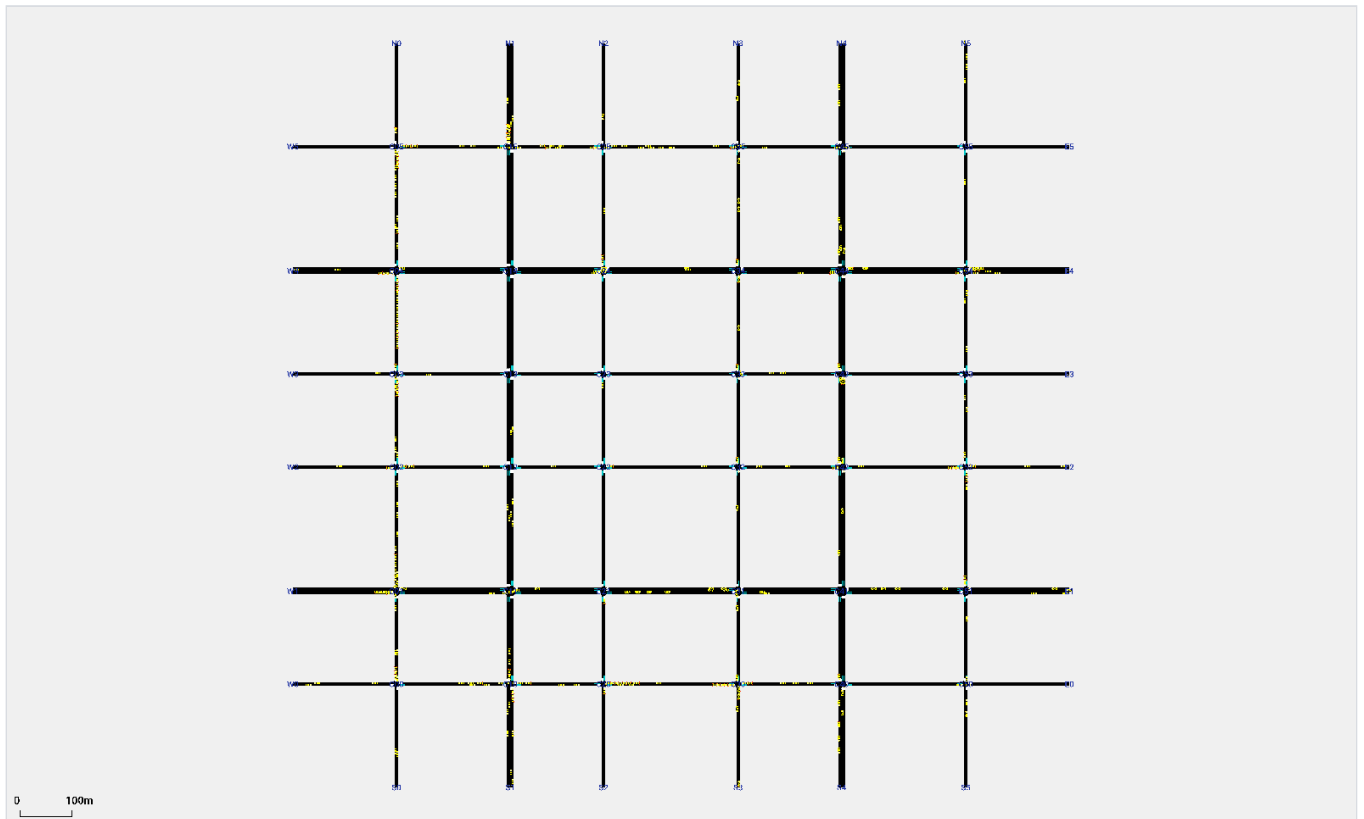
scenario	VECTOR	native	veh-act	fixed	VECTOR – native per seed	corr($\Delta P1$, $\Delta exposure$)	fitted $\Delta P1$ at equal exposure (s)
B	1498	1222	1462	1583	[401, 65, 44, 0, 483, 833, 282, 744, 812, -902]	-0.438	88.8
F	737	743	754	765	[142, -31, -132, -29, 73, -68, 26, 13, 14, -70]	-0.015	11.9
I	1700	1637	1726	1766	[80, -98, 10, -130, 69, 177, 254, 159, 106, -2]	0.015	-16.7
N	1131	1130	1103	1196	[-25, -15, 6, 12, -66, 132, 83, -68, 9, -59]	-0.147	1.7
O	1098	1117	1109	1150	[-257, 111, 46, -27, 0, 1, -49, -14, 101, -103]	-0.372	60.0

Evidence gallery

27 screenshots. Every image is either a `sumo-gui` screenshot of the running simulation, taken over TraCI during an audit trial, or a screenshot of the audit dashboard rendering recorded result files. None is drawn, edited or composited. Each image's sha256 was re-verified on disk when this page was generated (all match).

Reading the simulator images: the thin red, yellow and green lines at the stop lines are the signal states SUMO is showing; the cyan bars just past junctions are the lane-area detectors the harness uses to measure spillback; the ambulance is white with a cross icon, the fire engine red, the police car blue, ordinary cars yellow. A simulator image shows the state at `sim time`, one step after the capture call.

E01: city36, all 36 signalised junctions (C00-C55), VECTOR controlling every one, t=301 s



- **What it is:** sumo-gui screenshot of the running simulation.
- **Run:** `A/vector/95000` (scenario A, arm vector, seed 95000); sim time 301.0 s; captured 2026-10-25T18:17:19.175518+00:00.
- **Claim it supports:** the evaluated network is the 36-junction grid, under live control
- **Limitation:** an image shows layout and traffic; junction count is proven by `topology_report.json`
- **Linked data:** `artifacts/final-adversarial-audit/trials/heldout/A/vector/95000.json`, `artifacts/final-adversarial-audit/traces/heldout/A_vector_95000.jsonl.gz`
- **sha256:** `5707b9171dbdc31218683eb37fde0122f39c1e9a4a919cde14f2c0de45bfdb13`
- **State recorded at capture:** halting on approaches 56; C33: YELLOW EW (last decision: VEHICLE_DEMAND (1 vehicles in lane_1; gap_out)); C32: GREEN EW (last decision: NO_DEMAND_REST_IN_GREEN); C23: GREEN NS (last decision: NO_DEMAND_REST_IN_GREEN)

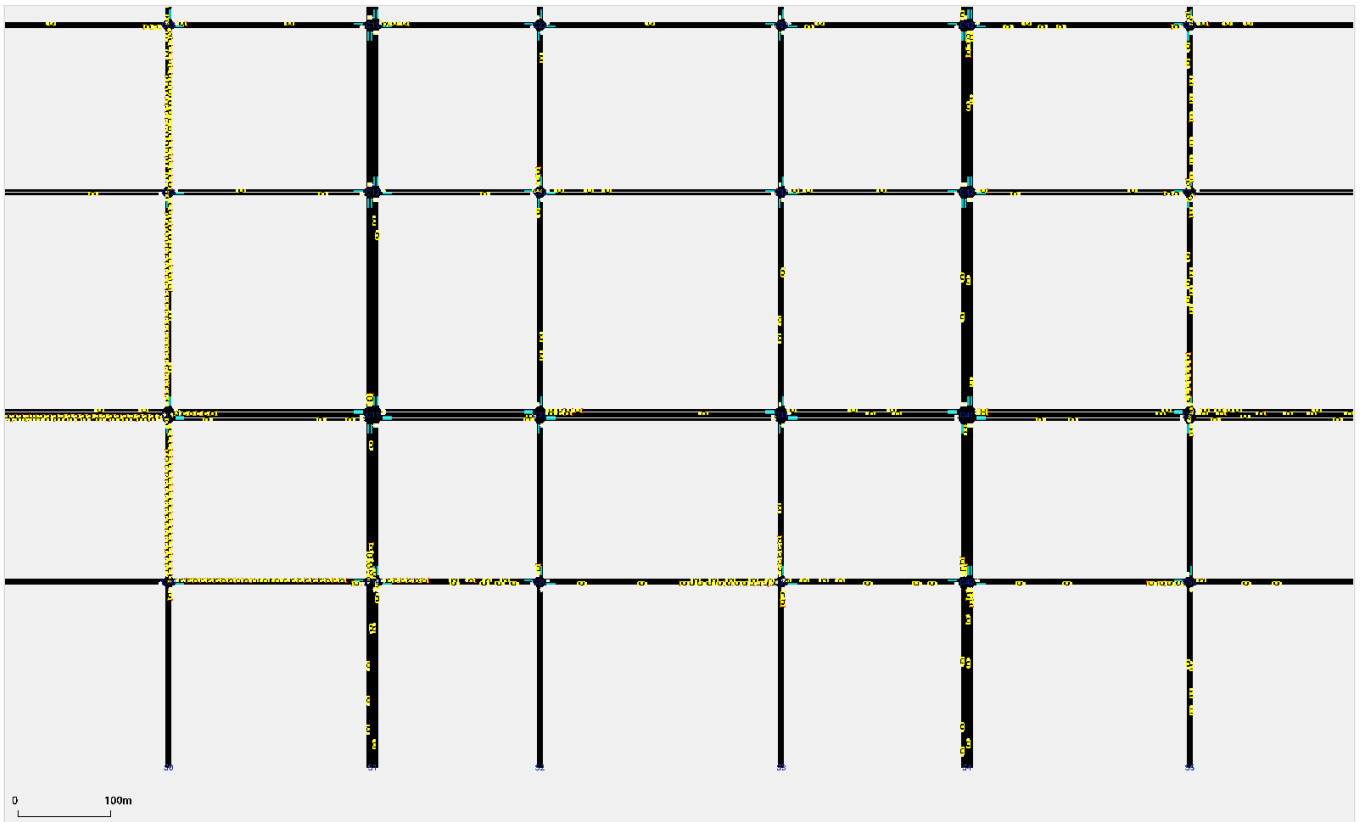
E02: Topology validation of city36: 36 signal-controlled intersections, gate A

VECTOR final adversarial audit rendered from topology/topology_report.json		
Topology validation: city36_static (SUMO 1.28.0) — gate A PASS		
check	controlled variant	native variant
junctions in the file	60	60
signal-controlled intersections	36	36
controlled links	480	480
roads (directed edges)	168	168
lanes	224	224
lane connections	480	480
stop-line detectors	192	192
OD routes legal	552	552
strongly connected (controlled junctions)	true	true
every controlled junction on some route	true	true
validation problems	0	0
junctions by type	{"dead_end":24,"traffic_light":36}	{"dead_end":24,"traffic_light":36}
excluded from control	24 fringe nodes: fringe nodes (type priority / dead end): where vehicles enter and leave the network; not intersections, no conflicting movements	
file canonical sha256	8a9ade80bb0d7c06...	6f2c50a9e2be4f03...

Per controlled junction						
junction	approaches	approach lanes	controlled links	movements	program	OD routes crossing
C00	4	4	12	l r s	static, 4 phases, 90 s	146
C01	4	6	14	l r s	static, 4 phases, 90 s	156
C02	4	4	12	l r s	static, 4 phases, 90 s	169
C03	4	4	12	l r s	static, 4 phases, 90 s	169
C04	4	6	14	l r s	static, 4 phases, 90 s	156
C05	4	4	12	l r s	static, 4 phases, 90 s	146
C10	4	6	14	l r s	static, 4 phases, 90 s	128
C11	4	8	16	l r s	static, 4 phases, 90 s	49
C12	4	6	14	l r s	static, 4 phases, 90 s	47
C13	4	6	14	l r s	static, 4 phases, 90 s	47
C14	4	8	16	l r s	static, 4 phases, 90 s	47
C15	4	6	14	l r s	static, 4 phases, 90 s	126

- **What it is:** audit dashboard rendering recorded results.
- **Run:** `—` (scenario `—`, arm `—`, seed `—`); captured 2026-10-25T18:33:20.839Z.
- **Claim it supports:** the executed network has 36 connected signal-controlled intersections, validated from the SUMO network files
- **Limitation:** a synthetic 6×6 grid, not a real city; validation is structural (junctions, links, programmes), not behavioural
- **Linked data:** artifacts/final-adversarial-audit/topology/topology_report.json
- **sha256:** d3f8ab9ab508884ab7eeb12a94a6d2abf607d8ad6d1e3be73c610d1dcb2d5bbc

E03: the same corridor, same seed and surge, under SUMO's native actuated baseline



- **What it is:** sumo-gui screenshot of the running simulation.
- **Run:** `A/native_actuated/95000` (scenario A, arm native_actuated, seed 95000); sim time 1201.0 s; captured 2026-10-25T18:04:42.728208+00:00.
- **Claim it supports:** the baseline faces identical demand at the same instant
- **Limitation:** one instant of one seed; the comparison is in the metrics, not in two pictures
- **Linked data:** `artifacts/final-adversarial-audit/trials/heldout/A/native_actuated/95000.json`, `artifacts/final-adversarial-audit/traces/heldout/A_native_actuated_95000.jsonl.gz`
- **sha256:** `e38b744a138a1b536165ad2bb2ba02a86d397d09fa88e31e4facdc204e0a7776`
- **State recorded at capture:** halting on approaches 220; C31: ALL_RED; C21: ALL_RED; C30: GREEN EW

E04: Scenario G, junction C00 (congested corner), t 1688-1726 s: true queue, what the sensor model showed VECTOR, the QueueObserver belief and the decision

VECTOR final adversarial audit rendered from traces/heldout/G_vector_95000.jsonl.gz

What VECTOR was shown at C00: scenario G, seed 95000, t 1688–1726 s

Each row is one real controller call. **truth** is SUMO's halting count (NS/EW); **shown** is what the controller received after the scenario's sensor model; **belief** is VECTOR's QueueObserver output that the law used. 10 of 28 decisions in the window had a sensor fault.

t (s)	truth NS/EW	shown NS/EW	sensor flags	belief NS/EW	reason (control_logic)	adapter effect
1688	6/42	5/56	dropout, stale	7.5/31.5	VEHICLE_DEMAND (8 vehicles in lane_1; discharging 8)	extend NS
1689	6/42	5/56	dropout, stale	7.0/32.0	VEHICLE_DEMAND (7 vehicles in lane_1; discharging 7)	extend NS
1690	6/42	5/47	exact	6.5/32.5	VEHICLE_DEMAND (6 vehicles in lane_1; discharging 6)	extend NS
1691	6/42	5/52	exact	6.0/33.0	VEHICLE_DEMAND (6 vehicles in lane_1; discharging 6)	extend NS
1692	6/42	5/52	dropout, stale	5.5/33.5	VEHICLE_DEMAND (6 vehicles in lane_1; discharging 6)	extend NS
1693	6/42	3/40	exact	5.0/34.0	VEHICLE_DEMAND (5 vehicles in lane_1; discharging 5)	extend NS
1694	6/42	3/40	dropout, stale	4.5/34.5	VEHICLE_DEMAND (4 vehicles in lane_1; discharging 4)	extend NS
1695	6/42	4/49	exact	4.0/35.0	VEHICLE_DEMAND (4 vehicles in lane_1; discharging 4)	extend NS
1696	6/43	5/62	exact	4.5/35.5	VEHICLE_DEMAND (4 vehicles in lane_1; discharging 4)	extend NS
1697	6/43	33/28	spike:NS	5.0/35.0	VEHICLE_DEMAND (5 vehicles in lane_1; discharging 5)	extend NS
1698	6/43	12/44	exact	5.5/35.5	VEHICLE_DEMAND (36 vehicles in lane_2; max_out at 45s)	end NS green (yellow)
1703	8/43	12/45	exact	8.0/38.0	VEHICLE_DEMAND (38 vehicles in lane_2)	green EW
1711	12/31	8/51	exact	8.0/42.0	VEHICLE_DEMAND (42 vehicles in lane_2; discharging 42)	extend EW
1712	12/29	17/34	exact	8.5/41.5	VEHICLE_DEMAND (42 vehicles in lane_2; discharging 42)	extend EW
1713	13/27	22/37	exact	9.0/41.0	VEHICLE_DEMAND (41 vehicles in lane_2; discharging 41)	extend EW
1714	13/26	22/37	dropout, stale	9.5/40.5	VEHICLE_DEMAND (40 vehicles in lane_2; discharging 40)	extend EW
1715	13/26	26/33	exact	10.0/40.0	VEHICLE_DEMAND (40 vehicles in lane_2; discharging 40)	extend EW
1716	13/25	40/24	spike:NS	10.5/39.5	VEHICLE_DEMAND (40 vehicles in lane_2; discharging 40)	extend EW
1717	13/23	40/24	dropout, stale	11.0/39.0	VEHICLE_DEMAND (39 vehicles in lane_2; discharging 39)	extend EW
1718	13/22	18/9	exact	11.5/38.5	VEHICLE_DEMAND (38 vehicles in lane_2; discharging 38)	extend EW
1719	13/21	17/23	exact	12.0/38.0	VEHICLE_DEMAND (38 vehicles in lane_2; discharging 38)	extend EW
1720	13/20	14/16	exact	12.5/37.5	VEHICLE_DEMAND (38 vehicles in lane_2; discharging 38)	extend EW
1721	13/21	20/25	exact	13.0/37.0	VEHICLE_DEMAND (37 vehicles in lane_2; discharging 37)	extend EW
1722	13/20	14/21	exact	13.5/36.5	VEHICLE_DEMAND (36 vehicles in lane_2; discharging 36)	extend EW
1723	13/21	12/15	exact	13.0/36.0	VEHICLE_DEMAND (36 vehicles in lane_2; discharging 36)	extend EW
1724	13/21	12/15	dropout, stale	12.5/35.5	VEHICLE_DEMAND (36 vehicles in lane_2; discharging 36)	extend EW
1725	13/20	16/25	exact	13.0/35.0	VEHICLE_DEMAND (35 vehicles in lane_2; discharging 35)	extend EW

- **What it is:** audit dashboard rendering recorded results.
- **Run:** G/vector/95000 (scenario G, arm vector, seed 95000); sim time 1688 s; captured 2026-10-25T18:33:21.378Z.
- **Claim it supports:** VECTOR decided on corrupted observations (noise, dropouts, phantom spikes); the observer clamps impossible jumps
- **Limitation:** the sensor model is the audit's own simple model (Gaussian noise, Bernoulli dropout, spikes), not a real detector's failure physics
- **Linked data:** artifacts/final-adversarial-audit/traces/heldout/G_vector_95000.jsonl.gz
- **sha256:** 418fe6a5392029f97eeb21fb459ba1da81ec2d494fd924b4ed87c7d50d2a78fc

E05: Scenario R0, junction C21: shadow forecast of total queue at t+8 s against what happened

VECTOR final adversarial audit rendered from traces/heldout/R0_vector_95000.jsonl.gz, trials/heldout/R0/vector/95000.json

VECTOR's shadow forecast at C21: scenario R0, seed 95000 (horizon 8 s)

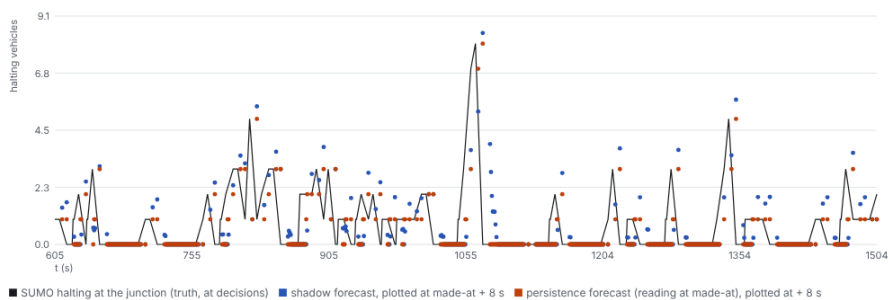
ONE FORECAST, AS RECORDED

made at	t = 1104 s
for	t = 1112 s (horizon 8 s)
forecast total queue	0.00 veh (mean of 1 shadow forecasters)
tier / regime	SENTRY / FREE_FLOW
escalations	none
disagreement	—
models authorised to influence	none (gate holds every model)
influence on the signal	none: computed after the phase is chosen

THIS RUN, ALL JUNCTIONS: FORECAST VS SUMO AT T+8 S

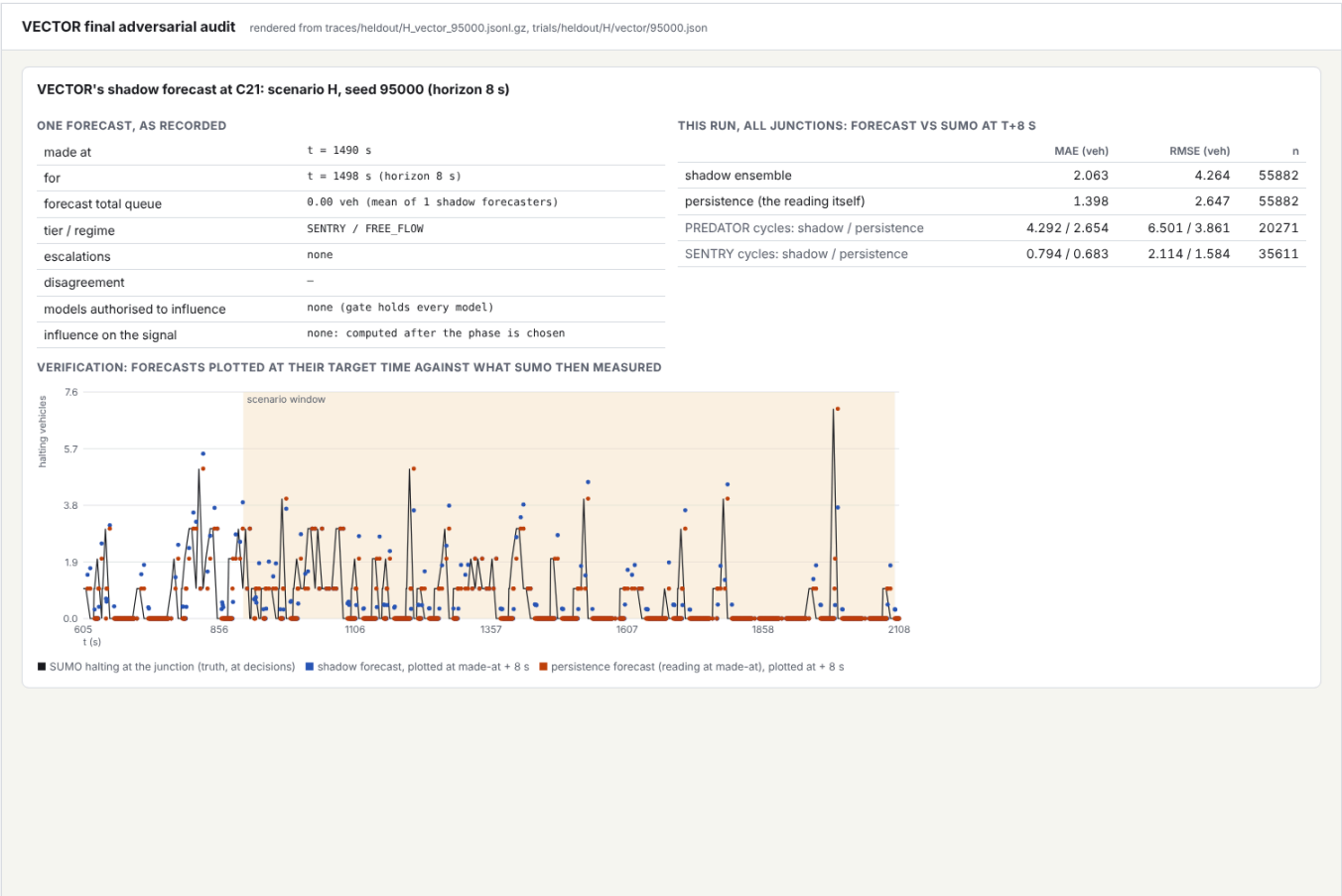
	MAE (veh)	RMSE (veh)	n
shadow ensemble	1.780	3.804	50461
persistence (the reading itself)	1.260	2.360	50461
PREDATOR cycles: shadow / persistence	3.651 / 2.362	5.780 / 3.406	18139
SENTRY cycles: shadow / persistence	0.729 / 0.641	1.960 / 1.477	32322

VERIFICATION: FORECASTS PLOTTED AT THEIR TARGET TIME AGAINST WHAT SUMO THEN MEASURED

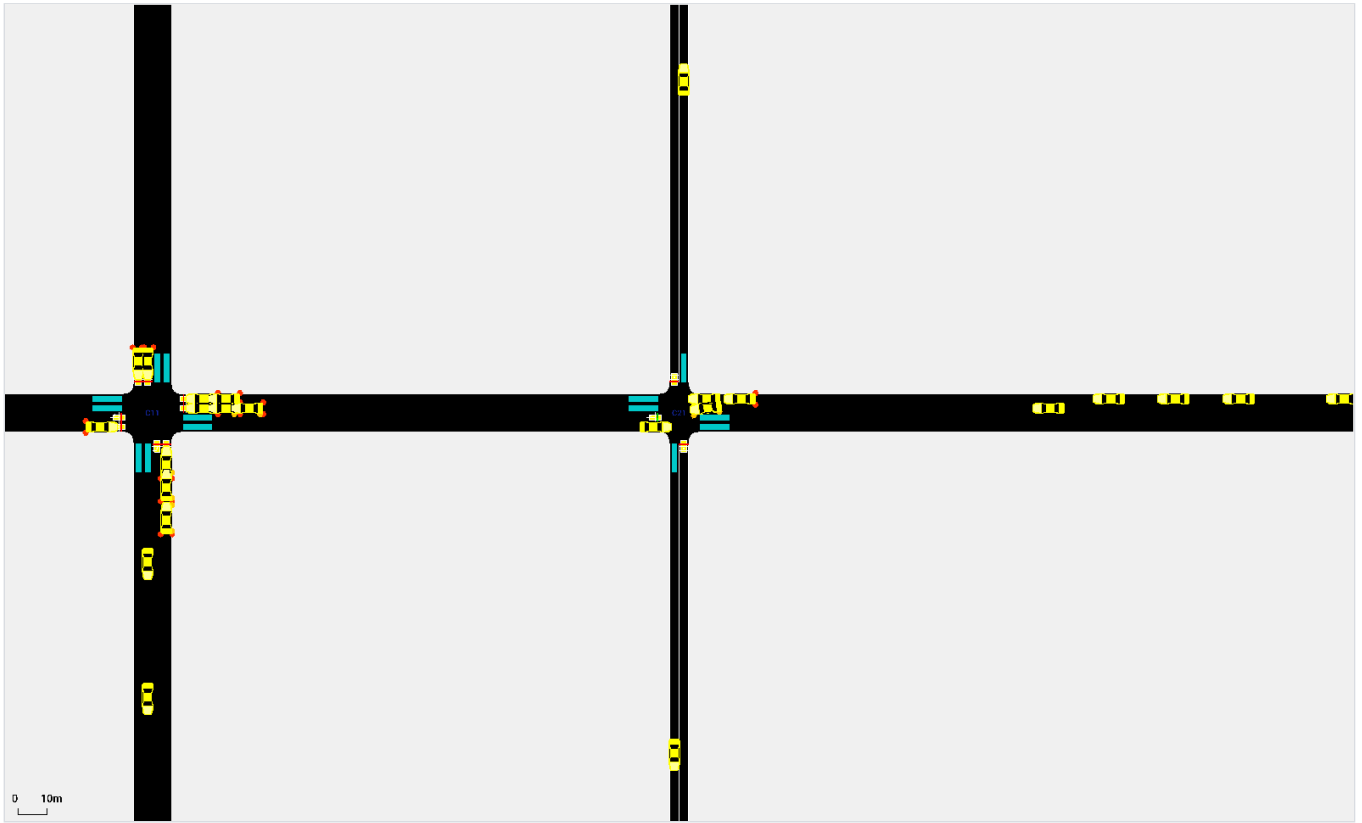


- **What it is:** audit dashboard rendering recorded results.
- **Run:** R0/vector/95000 (scenario R0, arm vector, seed 95000); captured 2026-10-25T18:33:21.991Z.
- **Claim it supports:** the shadow forecasters run in the live loop and are scored against the realised queue
- **Limitation:** shadow only: no forecast reaches a signal decision; held-out MAE is worse than persistence in 15 of 16 scenarios
- **Linked data:** artifacts/final-adversarial-audit/traces/heldout/R0_vector_95000.jsonl.gz
- **sha256:** ecf7915ba59981691bf3171958be62db07ae04926227b8e1015a4d0b61365d1e

E06: Scenario H, junction C21: shadow forecast across the concept shift at 900 s

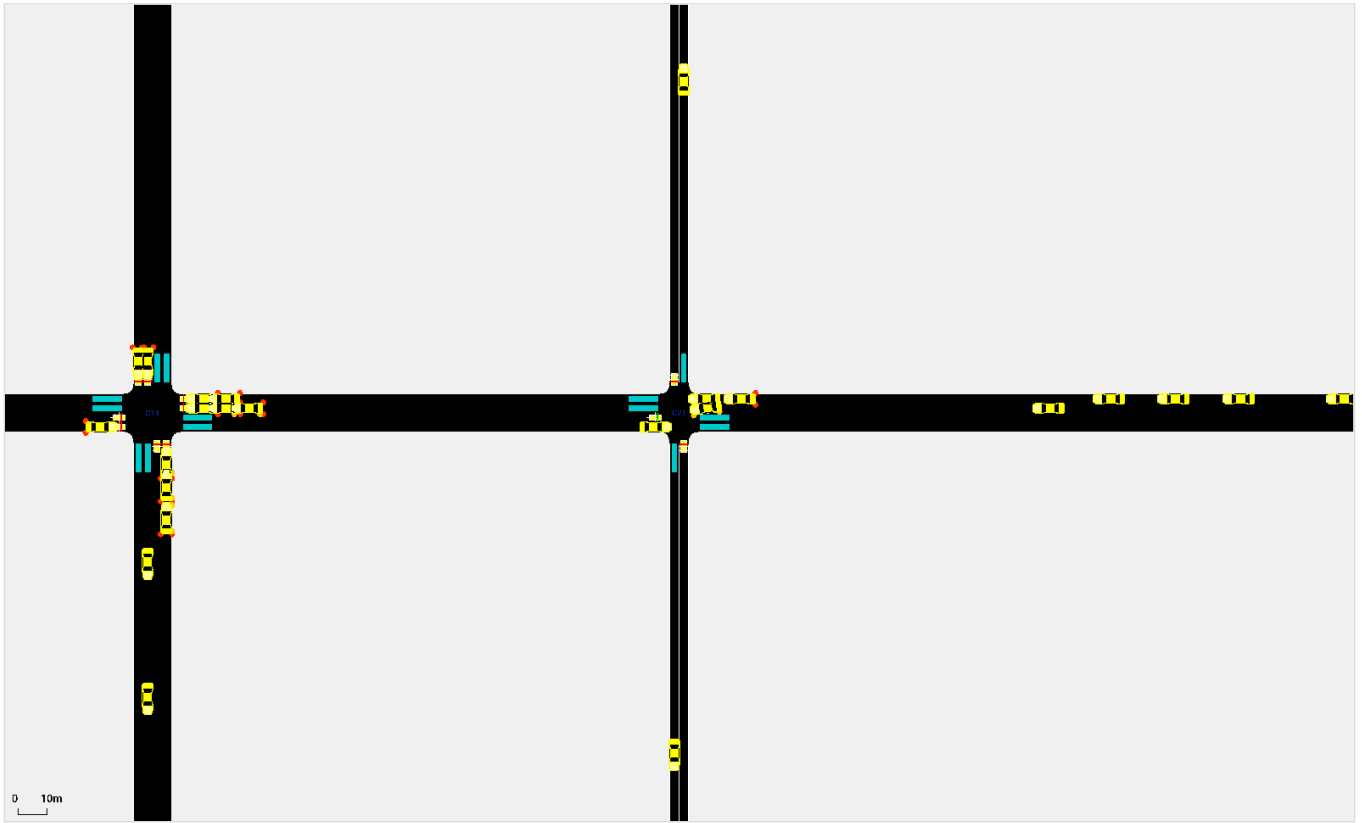


E07: row-1 corridor at 1201 s, during the x4 surge from W1 and E1 (600-1500 s): the queue is concentrated at the western entry C01 (52 halting on EW); the downstream corridor junctions hold 0-3



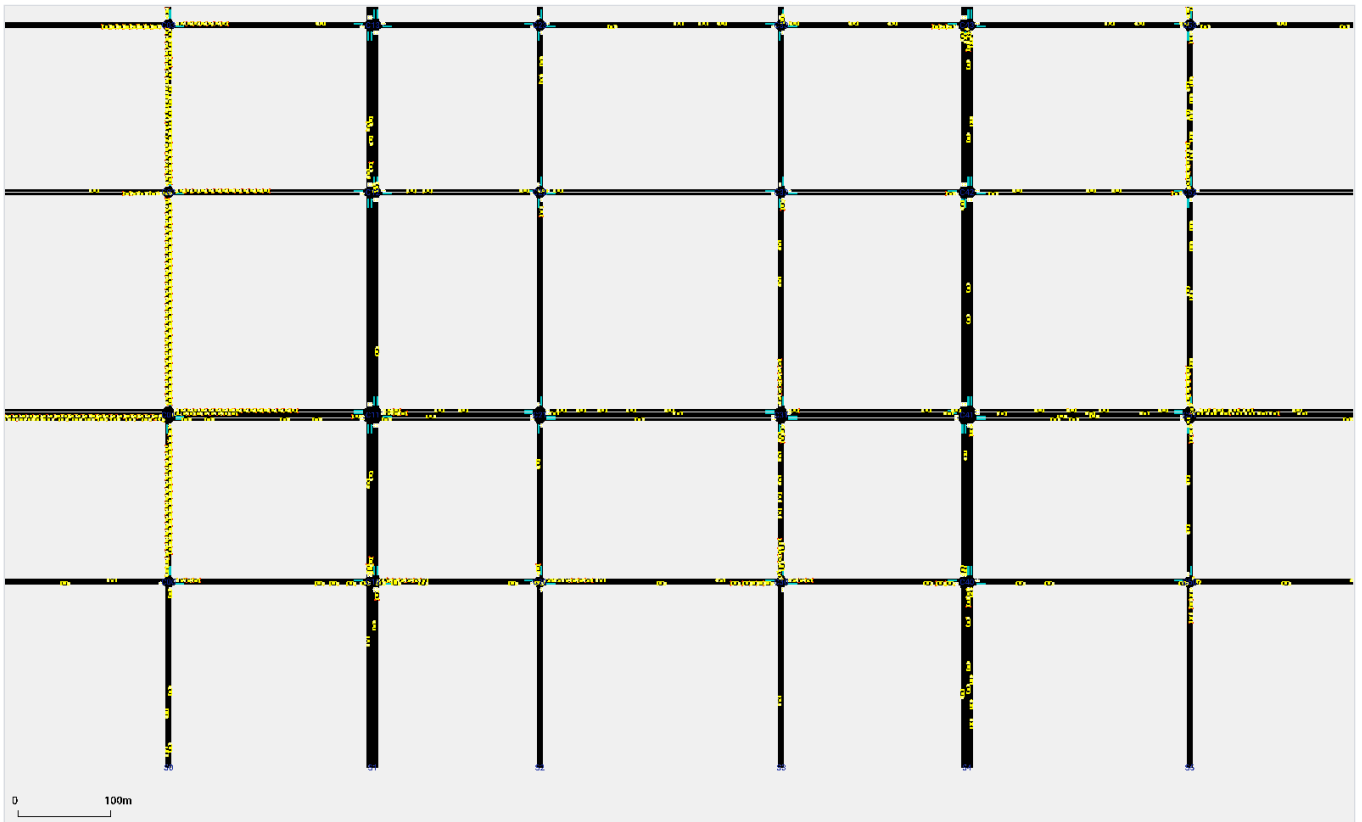
- **What it is:** sumo-gui screenshot of the running simulation.
- **Run:** `A/vector/95000` (scenario A, arm vector, seed 95000); sim time 1201.0 s; captured 2026-10-25T18:17:51.539483+00:00.
- **Claim it supports:** the surge congests the corridor entry, and VECTOR's input there is the halting count
- **Limitation:** VECTOR has no congestion detector; 'detection' here is the halting count it is shown
- **Linked data:** `artifacts/final-adversarial-audit/trials/heldout/A/vector/95000.json`, `artifacts/final-adversarial-audit/traces/heldout/A_vector_95000.jsonl.gz`
- **sha256:** `d690b5bf344b0d43277d13baf395189d83d83b08c1d062d6d2964a8bc875eecf`
- **State recorded at capture:** halting on approaches 297; C31: GREEN EW (last decision: VEHICLE_DEMAND (2 vehicles in lane_2; discharging 2)); C21: GREEN EW (last decision: VEHICLE_DEMAND (3 vehicles in lane_2)); C30: GREEN NS (last decision: VEHICLE_DEMAND (6 vehicles in lane_1))

E08: C21 at its first commanded change after 1200 s: after clearance VECTOR gives EW green to 3 halting vehicles (VEHICLE_DEMAND); C11, left, is in all-red after ending EW green on gap-out



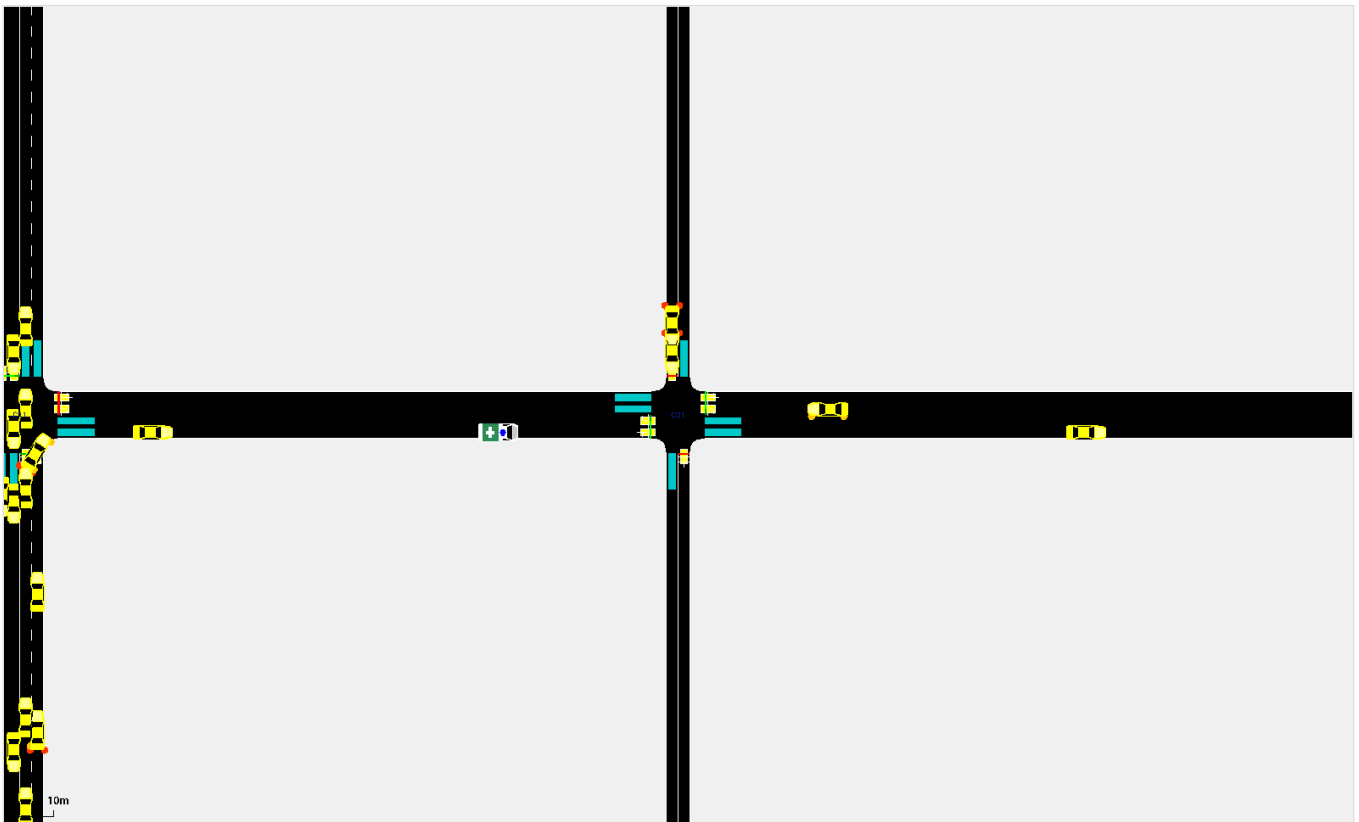
- **What it is:** sumo-gui screenshot of the running simulation.
- **Run:** `A/vector/95000` (scenario A, arm vector, seed 95000); sim time 1201.0 s; captured 2026-10-25T18:17:51.543470+00:00.
- **Claim it supports:** a VECTOR decision was executed by SUMO (commanded = reported state, checked each step)
- **Limitation:** the reason is in the decision record beside the image, not in the image
- **Linked data:** `artifacts/final-adversarial-audit/trials/heldout/A/vector/95000.json`, `artifacts/final-adversarial-audit/traces/heldout/A_vector_95000.jsonl.gz`
- **sha256:** `d690b5bf344b0d43277d13baf395189d83d83b08c1d062d6d2964a8bc875eecf`
- **State recorded at capture:** halting on approaches 297; C21: GREEN EW (last decision: VEHICLE_DEMAND (3 vehicles in lane_2)); C11: ALL_RED (last decision: VEHICLE_DEMAND (1 vehicles in lane_1; gap_out)); C20: GREEN NS (last decision: VEHICLE_DEMAND (1 vehicles in lane_1))

E09: row-1 corridor, same instant: six junctions, six independent decisions



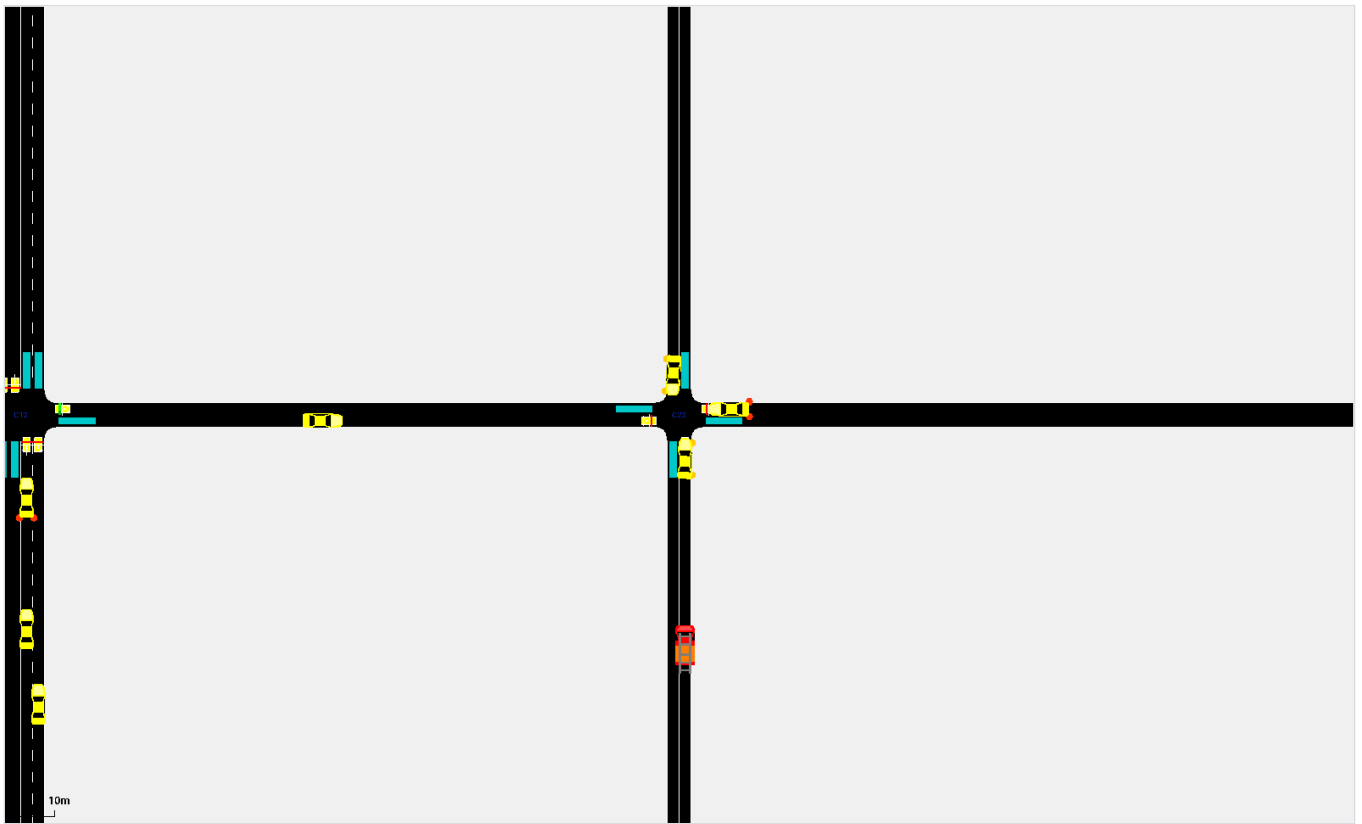
- **What it is:** sumo-gui screenshot of the running simulation.
- **Run:** `A/vector/95000` (scenario A, arm vector, seed 95000); sim time 1211.0 s; captured 2026-10-25T18:17:52.635789+00:00.
- **Claim it supports:** each corridor junction responds to its own queues
- **Limitation:** there is NO coordination mechanism (no offsets, no shared state); this shows its absence, not coordination
- **Linked data:** `artifacts/final-adversarial-audit/trials/heldout/A/vector/95000.json`, `artifacts/final-adversarial-audit/traces/heldout/A_vector_95000.jsonl.gz`
- **sha256:** `ee081fd837e930f83d5a418e1a934377f40078f0fe52c5d558596f80803d658c`
- **State recorded at capture:** halting on approaches 321; C31: YELLOW EW (last decision: VEHICLE_DEMAND (6 vehicles in lane_1; gap_out)); C21: GREEN EW (last decision: NO_DEMAND_REST_IN_GREEN); C30: GREEN NS (last decision: VEHICLE_DEMAND (3 vehicles in lane_1; discharging 3))

E10: ambulance amb1 (white, eastbound on row 1) approaching C21; the decision record beside it shows whether VECTOR gave its approach priority



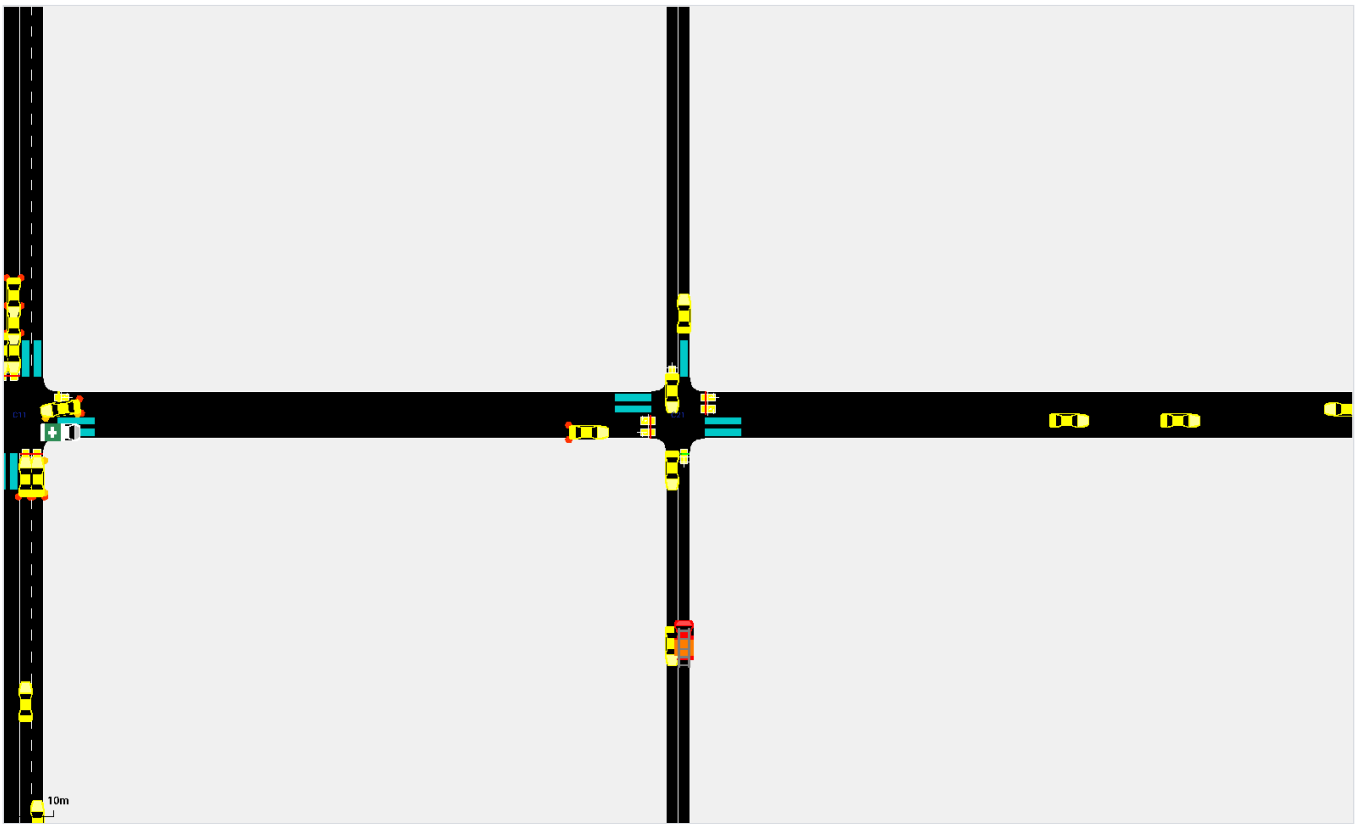
- **What it is:** sumo-gui screenshot of the running simulation.
- **Run:** `C/vector/95000` (scenario C, arm vector, seed 95000); sim time 947.0 s; captured 2026-10-25T18:05:29.913635+00:00.
- **Claim it supports:** VECTOR's emergency rule acts on an ambulance it is shown
- **Limitation:** detection is ideal (simulator ground truth, confidence 0.95); vehicles do not yield
- **Linked data:** `artifacts/final-adversarial-audit/trials/heldout/C/vector/95000.json`, `artifacts/final-adversarial-audit/traces/heldout/C_vector_95000.jsonl.gz`
- **sha256:** `0dbd4c40bcafc2aa8f49931c10bb571d5ed807c9ed11b77d4ce38763aaf845de`
- **State recorded at capture:** halting on approaches 130; C21: GREEN EW (last decision: EMERGENCY_OVERRIDE (0.95)); C11: GREEN NS (last decision: VEHICLE_DEMAND (6 vehicles in lane_1)); C20: GREEN NS (last decision: VEHICLE_DEMAND (5 vehicles in lane_1)); emergency vehicles `{"amb1": {"x": 542.6, "y": 375.2, "speed": 13.48, "lane": "C11_C21_0"}}`

E11: fire engine fire1 (red, northbound on column 2) after the C21 conflict, approaching C22



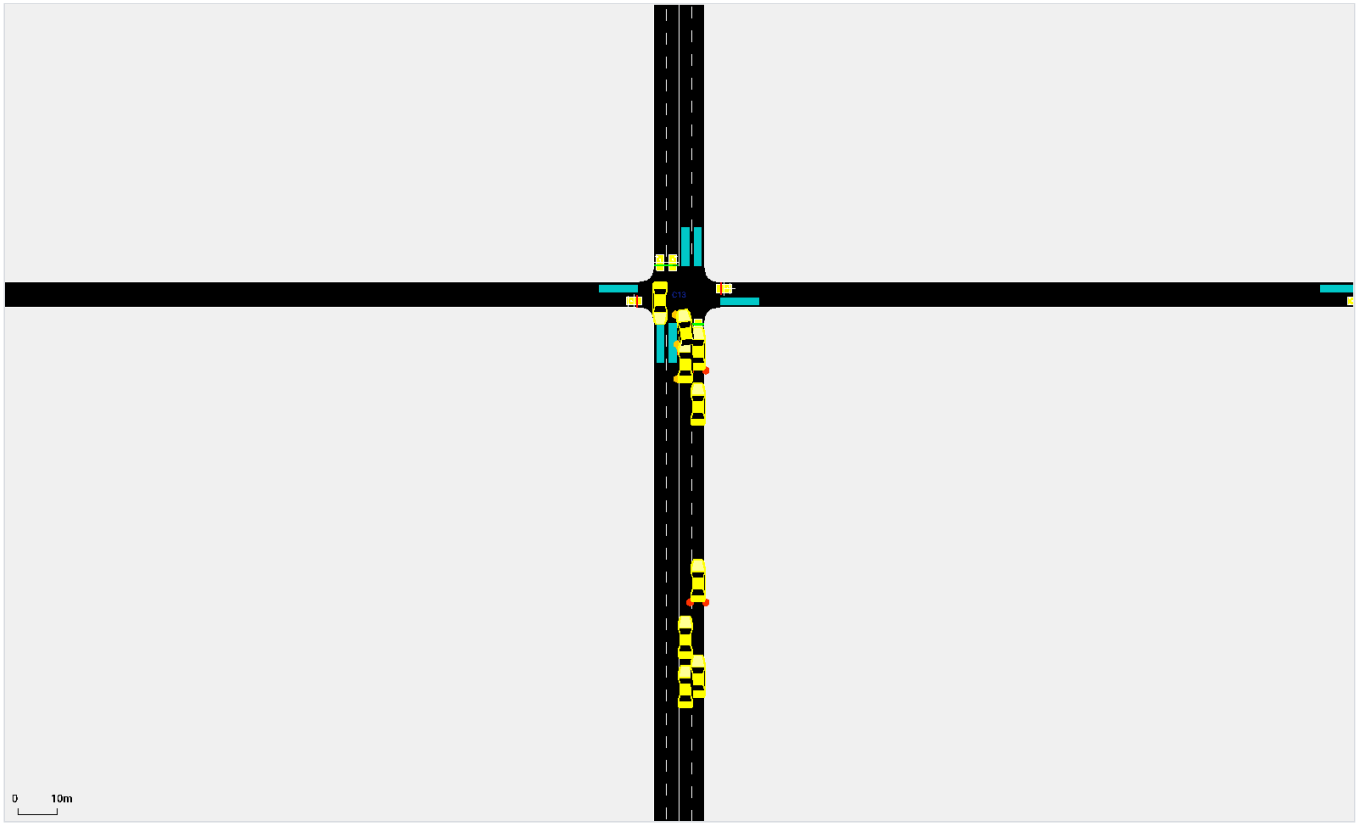
- **What it is:** sumo-gui screenshot of the running simulation.
- **Run:** `D/vector/95000` (scenario D, arm vector, seed 95000); sim time 960.0 s; captured 2026-10-25T18:06:40.863370+00:00.
- **Claim it supports:** the fire engine is handled by the same emergency rule
- **Limitation:** one vehicle at one instant; its full journey is in the EV table
- **Linked data:** `artifacts/final-adversarial-audit/trials/heldout/D/vector/95000.json`, `artifacts/final-adversarial-audit/traces/heldout/D_vector_95000.jsonl.gz`
- **sha256:** `d19491dfb6a859d43f97ebc698adda30e813ec0d7a3666979daa7c30ef9960c5`
- **State recorded at capture:** halting on approaches 136; C22: ALL_RED (last decision: EMERGENCY_OVERRIDE (0.95)); C12: GREEN EW (last decision: NO_DEMAND_REST_IN_GREEN); C23: GREEN EW (last decision: VEHICLE_DEMAND (1 vehicles in lane_2)); emergency vehicles {"amb1": {"x": 715.6, "y": 378.4, "speed": 13.01, "lane": "C21_C31_1"}, "fire1": {"x": 601.6, "y": 551.5, "speed": 10.91, "lane": "C21_C22_0"}}

E12: C21 as the ambulance (EW) and the fire engine (NS) converge on conflicting approaches



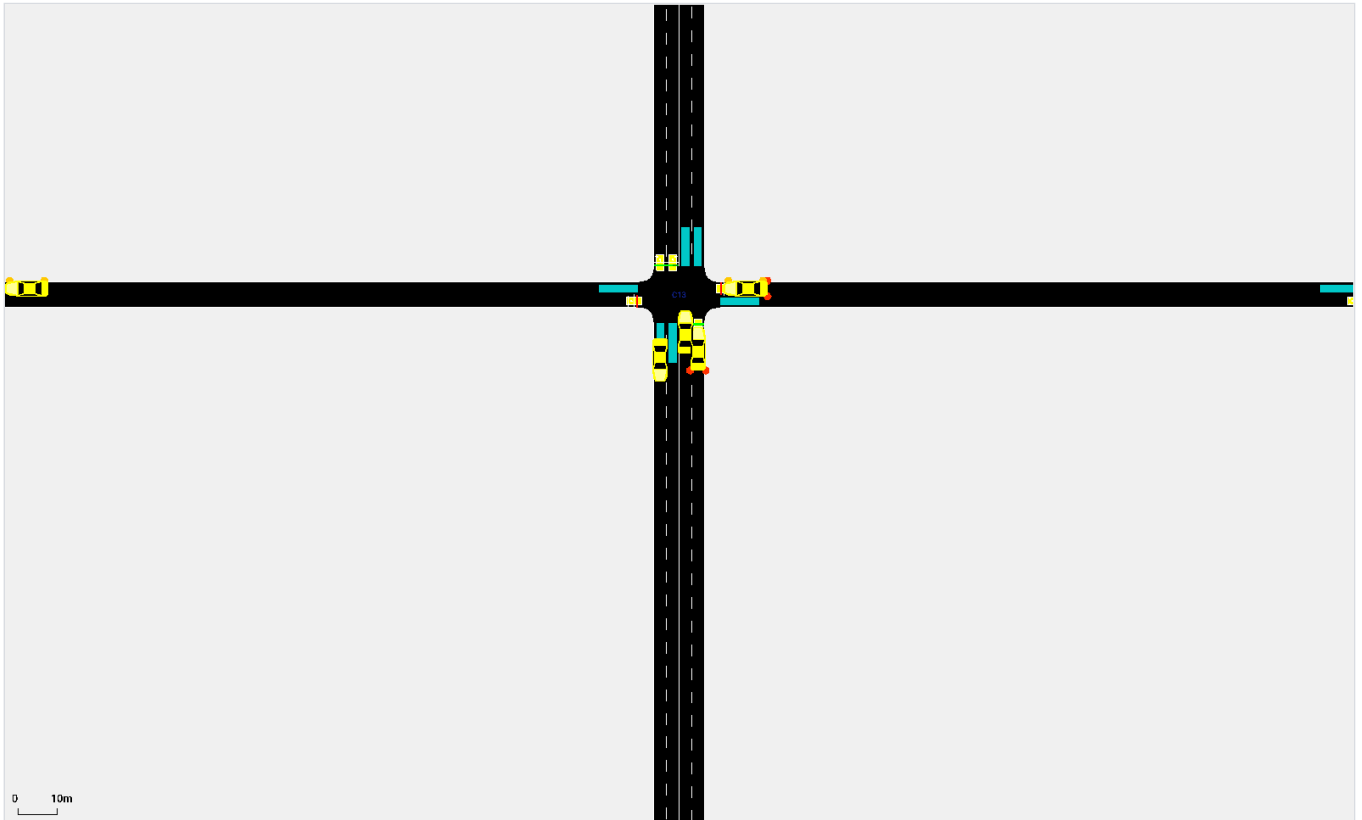
- **What it is:** sumo-gui screenshot of the running simulation.
- **Run:** `D/vector/95000` (scenario D, arm vector, seed 95000); sim time 938.0 s; captured 2026-10-25T18:06:39.989491+00:00.
- **Claim it supports:** competing requests at one junction: the record shows which approach was served
- **Limitation:** VECTOR's input admits ONE emergency per junction; the harness reports the most recently dispatched one, so the other gets no priority
- **Linked data:** `artifacts/final-adversarial-audit/trials/heldout/D/vector/95000.json`, `artifacts/final-adversarial-audit/traces/heldout/D_vector_95000.jsonl.gz`
- **sha256:** `0b3131306852c7fc8bcc83687f953e9ea4ce0336177b198ecde240d226a28443`
- **State recorded at capture:** halting on approaches 134; C21: GREEN NS (last decision: EMERGENCY_OVERRIDE (0.95)); C11: YELLOW EW (last decision: VEHICLE_DEMAND (5 vehicles in lane_1; gap_out)); C20: GREEN EW (last decision: VEHICLE_DEMAND (16 vehicles in lane_2)); emergency vehicles `{"amb1": {"x": 423.2, "y": 375.2, "speed": 12.81, "lane": ":C11_13_0"}, "fire1": {"x": 601.6, "y": 312.9, "speed": 10.73, "lane": "C20_C21_0"}}`

E13: capacity reduction on C12→C13 (column-1 arterial northbound), 1001 s: a vehicle stalled at C13's stop line in lane 0 since 700 s, speed limit 30% on the link; traffic passes in lane 1



- **What it is:** sumo-gui screenshot of the running simulation.
- **Run:** `F/vector/95000` (scenario F, arm vector, seed 95000); sim time 1001.0 s; captured 2026-10-25T18:44:35.959196+00:00.
- **Claim it supports:** the incident removes one of two lanes at C13's northbound stop line
- **Limitation:** VECTOR is not told of the incident; it sees only queues (the stalled vehicle is excluded from the count it is shown)
- **Linked data:** `artifacts/final-adversarial-audit/trials/heldout/F/vector/95000.json`, `artifacts/final-adversarial-audit/traces/heldout/F_vector_95000.jsonl.gz`
- **sha256:** `cadee3ee103686b52a32496708d6d7fc17572f244ee453e06157629b7b440b43`
- **State recorded at capture:** halting on approaches 124; C13: GREEN NS (last decision: VEHICLE_DEMAND (2 vehicles in lane_1)); C12: GREEN NS (last decision: NO_DEMAND_REST_IN_GREEN); C23: GREEN NS (last decision: NO_DEMAND_REST_IN_GREEN)

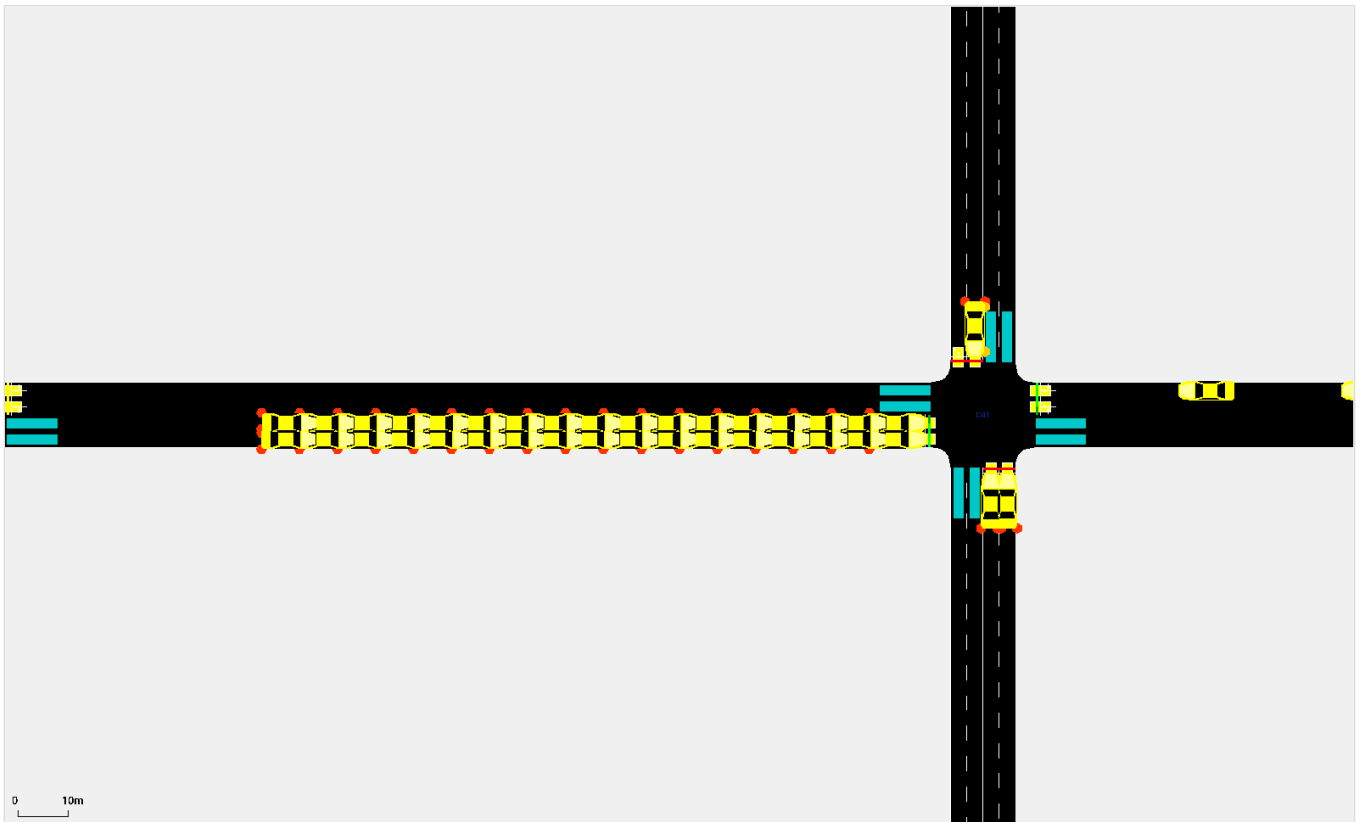
E14: the same view at 1401 s: lane 0 still blocked, traffic still on its assigned route



- **What it is:** sumo-gui screenshot of the running simulation.
- **Run:** `F/vector/95000` (scenario F, arm vector, seed 95000); sim time 1401.0 s; captured 2026-10-25T18:44:46.821943+00:00.
- **Claim it supports:** no vehicle was rerouted (route changes 0 across every run)
- **Limitation:** documents the ABSENCE of rerouting; there is no rerouting module in the SUMO path
- **Linked data:** `artifacts/final-adversarial-audit/trials/heldout/F/vector/95000.json`, `artifacts/final-adversarial-audit/traces/heldout/F_vector_95000.jsonl.gz`
- **sha256:** `4fb9dc94d7dbc94c5a05422beadb66603d2d91016ba5e411fecbead4ca79b496`
- **State recorded at capture:** halting on approaches 160; C13: GREEN NS (last decision: NO_DEMAND_REST_IN_GREEN); C12: GREEN NS (last decision: VEHICLE_DEMAND (1 vehicles in lane_1)); C23: GREEN NS (last decision: NO_DEMAND_REST_IN_GREEN)

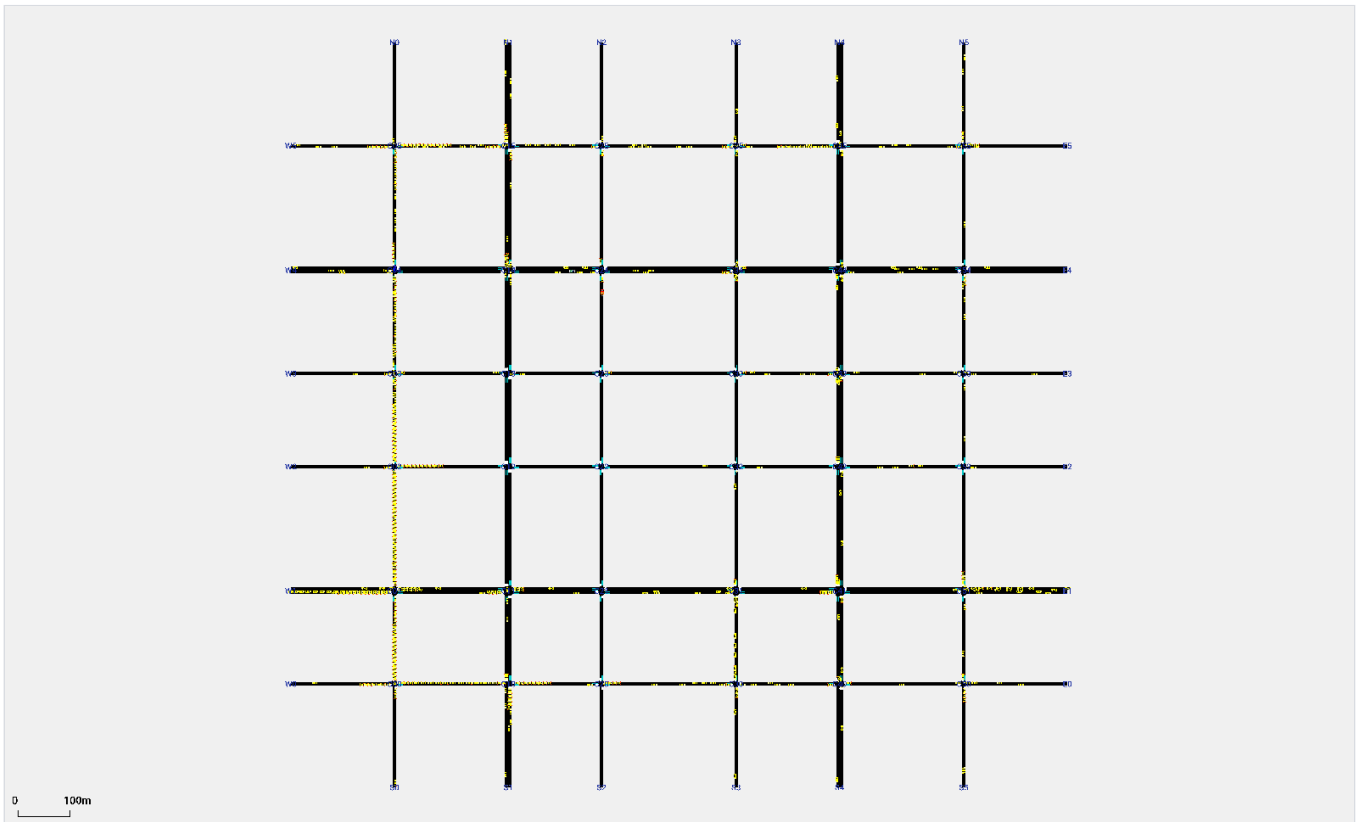
E15: C41 at 1291 s, scenario I: both lanes of the C31→C41 approach blocked by vehicles stalled at the stop line since 641/662 s; VECTOR has held EW green for

43 s on a 32-vehicle queue that cannot move (it ends at max green, 45 s, and is re-served)



- **What it is:** sumo-gui screenshot of the running simulation.
- **Run:** `I/vector/95000` (scenario I, arm vector, seed 95000); sim time 1291.0 s; captured 2026-10-25T18:20:46.242824+00:00.
- **Claim it supports:** VECTOR's law serves a queue that cannot discharge: halting count is its only demand signal (the mechanism of section 5.2)
- **Limitation:** one junction, one instant; the decision sequence is in the trace (E27) and the aggregate in ineffective green (S18)
- **Linked data:** `artifacts/final-adversarial-audit/trials/heldout/I/vector/95000.json`, `artifacts/final-adversarial-audit/traces/heldout/I_vector_95000.jsonl.gz`
- **sha256:** `dc41e4ba6740e1973b7946a634718328f420e8ae179710b88912238fe0bcf03c`
- **State recorded at capture:** halting on approaches 255; C41: GREEN EW (last decision: VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)); C31: GREEN EW (last decision: VEHICLE_DEMAND (2 vehicles in lane_2)); C40: GREEN EW (last decision: VEHICLE_DEMAND (7 vehicles in lane_2))

E16: combined stress at 1041 s: row-1 surge, C31→C41 blocked, ambulance and fire engine en route, a capacity cut due at 1100 s



- **What it is:** sumo-gui screenshot of the running simulation.
- **Run:** `0/vector/95000` (scenario O, arm vector, seed 95000); sim time 1041.0 s; captured 2026-10-25T18:21:51.414795+00:00.
- **Claim it supports:** VECTOR ran the whole network under simultaneous stresses
- **Limitation:** the effect of VECTOR, if any, is in the paired metrics
- **Linked data:** `artifacts/final-adversarial-audit/trials/heldout/0/vector/95000.json`, `artifacts/final-adversarial-audit/traces/heldout/0_vector_95000.jsonl.gz`
- **sha256:** `485c39939a69215c6ba2add926253cf4e5c7bf57ac7ebbb4b09c53dc35ad1676`
- **State recorded at capture:** halting on approaches 250; C33: GREEN NS (last decision: VEHICLE_DEMAND (1 vehicles in lane_1)); C32: GREEN EW (last decision: VEHICLE_DEMAND (1 vehicles in lane_2)); C23: GREEN NS (last decision: NO_DEMAND_REST_IN_GREEN); emergency vehicles {"amb1": {"x": 536.0, "y": 998.4, "speed": 12.74, "lane": "C14_C24_1"}, "ev1": {"x": 199.3, "y": 1001.3, "speed": 6.48, "lane": "C04_17_0"}, "fire1": {"x": 601.6, "y": 952.5, "speed": 10.8, "lane": "C23_C24_0"}}

E17: Primary endpoint on the held-out seeds: VECTOR minus SUMO's native actuated control, per scenario, paired 95% t-intervals; the coloured label is the unadjusted interval, the Holm column is the family-wise test (only B rejects)

VECTOR final adversarial audit rendered from results/heldout/comparisons.json, results/heldout/summary.json

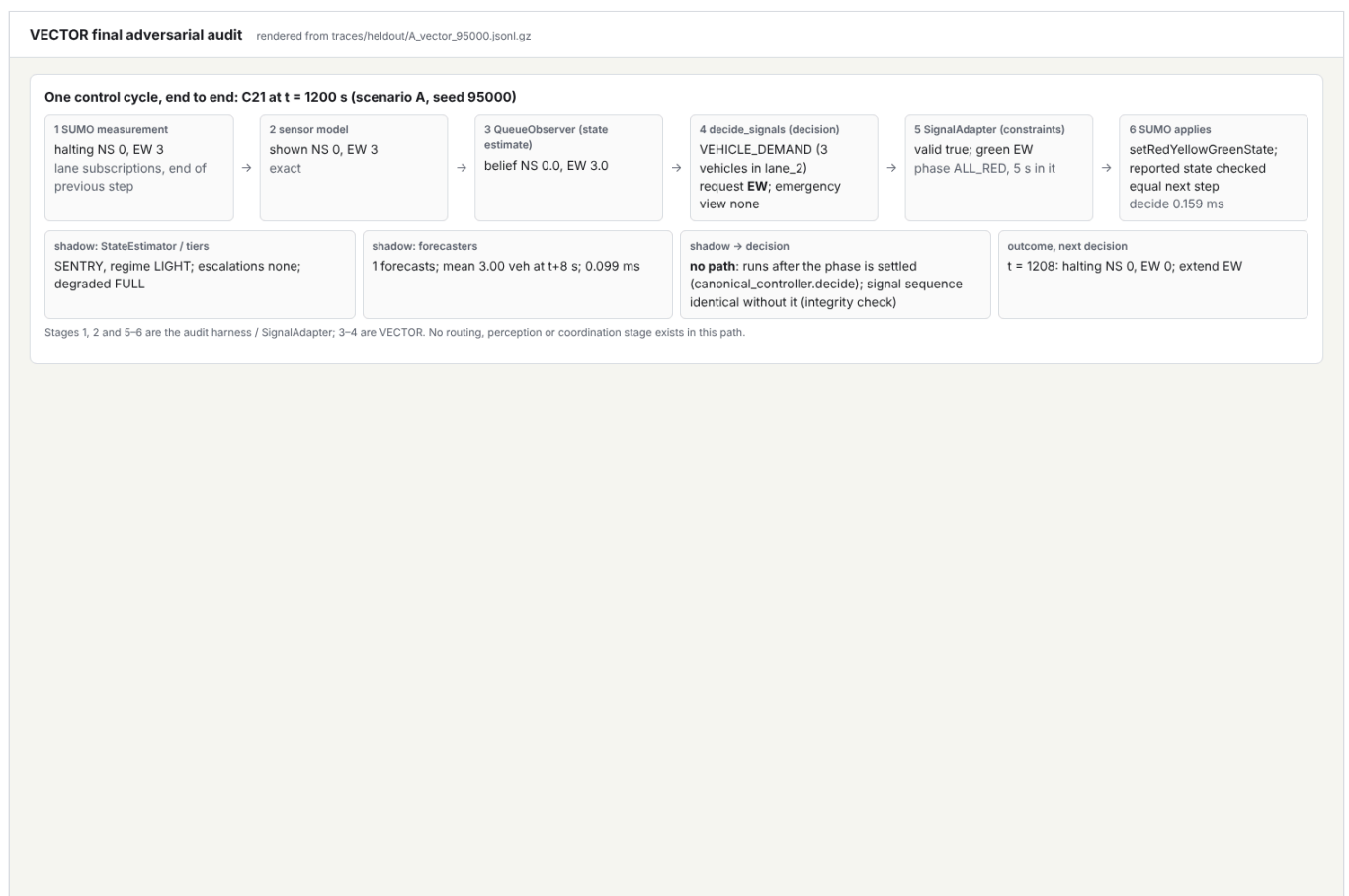
Measured comparison, held-out seeds 95000-95009: VECTOR - native_actuated on P1_mean_time_loss_s

scenario	paired seeds	VECTOR	native_actuated	difference	95% CI (t)	relative	seeds better/worse	Wilcoxon p	Holm p	at 95%
R0	10	170.01	181.97	-11.96	[-69.93, 46.01]	-6.6%	4/6	0.9219	1.0000	not resolved
A	10	330.06	302.60	27.46	[-22.79, 77.72]	9.1%	4/6	0.2754	1.0000	not resolved
B	10	314.75	241.19	73.56	[26.61, 120.51]	30.5%	0/10	0.0020	0.0312	VECTOR worse
C	10	173.80	140.73	33.06	[16.31, 49.82]	23.5%	1/9	0.0039	0.0586	VECTOR worse
D	10	175.60	146.15	29.45	[12.67, 46.23]	20.2%	1/9	0.0059	0.0762	VECTOR worse
E	10	173.43	173.12	0.30	[-57.72, 58.33]	0.2%	2/8	0.4316	1.0000	not resolved
F	10	174.08	162.15	11.93	[-26.83, 50.70]	7.4%	3/7	0.3223	1.0000	not resolved
G	10	163.65	181.97	-18.32	[-73.39, 36.75]	-10.1%	5/5	0.8457	1.0000	not resolved
H	10	458.10	406.92	51.17	[19.86, 82.49]	12.6%	2/8	0.0098	0.1172	VECTOR worse
I	10	242.72	258.90	-16.18	[-65.62, 33.25]	-6.3%	4/6	0.7695	1.0000	not resolved
J	10	591.77	522.69	69.07	[6.96, 131.19]	13.2%	2/8	0.0371	0.3711	VECTOR worse
K	10	453.06	341.84	111.22	[67.52, 154.91]	32.5%	1/9	0.0039	0.0586	VECTOR worse
L	10	187.62	181.97	5.65	[-55.00, 66.31]	3.1%	3/7	0.3750	1.0000	not resolved
M	10	170.01	181.97	-11.96	[-69.93, 46.01]	-6.6%	4/6	0.9219	1.0000	not resolved
N	10	292.34	290.81	1.53	[-56.34, 59.39]	0.5%	4/6	0.9219	1.0000	not resolved
O	10	326.53	261.46	65.07	[11.52, 118.63]	24.9%	1/9	0.0273	0.3008	VECTOR worse

Difference = VECTOR - baseline per seed (same demand, same events). For delay, negative favours VECTOR. "not resolved": the interval contains 0.

- **What it is:** audit dashboard rendering recorded results.
- **Run:** `—` (scenario `—`, arm `—`, seed `—`); captured 2026-10-25T18:33:23.064Z.
- **Claim it supports:** VECTOR does not beat the native actuated controller on mean time loss; it is significantly worse in B after Holm correction
- **Limitation:** 10 paired seeds per scenario; one synthetic network; the native controller uses SUMO's own detectors
- **Linked data:** `artifacts/final-adversarial-audit/results/heldout/comparisons.json`
- **sha256:** `0aba1b3daa80d6a66796c28c6150bd4e39da68bfb4cc5b5bb4379a5343e34564`

E18: One control cycle end to end at C21, t=1200 s, scenario A: measurement, sensor, observer, law, adapter, SUMO, shadow



- **What it is:** audit dashboard rendering recorded results.
- **Run:** A/vector/95000 (scenario A, arm vector, seed 95000); sim time 1200 s; captured 2026-10-25T18:33:24.559Z.
- **Claim it supports:** the live path is SUMO → observer → decide_signals → SignalAdapter → SUMO; the shadow has no path to the decision
- **Limitation:** a single decision record; that the shadow has no influence is proven by the paired signal-sequence check, not by this image
- **Linked data:** artifacts/final-adversarial-audit/traces/heldout/A_vector_95000.json.gz
- **sha256:** 141d5874a1ff9bf8669485ab41d93057aa5602cce2173a32f297882f0ac65d8f

E19: Scenario M: forecasters fail at 900 s, the whole advisory orchestrator at 1500 s; control continues unchanged

VECTOR final adversarial audit rendered from traces/heldout/M_vector_95000.jsonl.gz, trials/heldout/M/vector(_noshadow)/95000.json

Module degradation: scenario M, seed 95000, junction C21

Injected: `forecasters at t=900` `orchestrator at t=1500` (every junction). Shadow failures counted: **25523**.

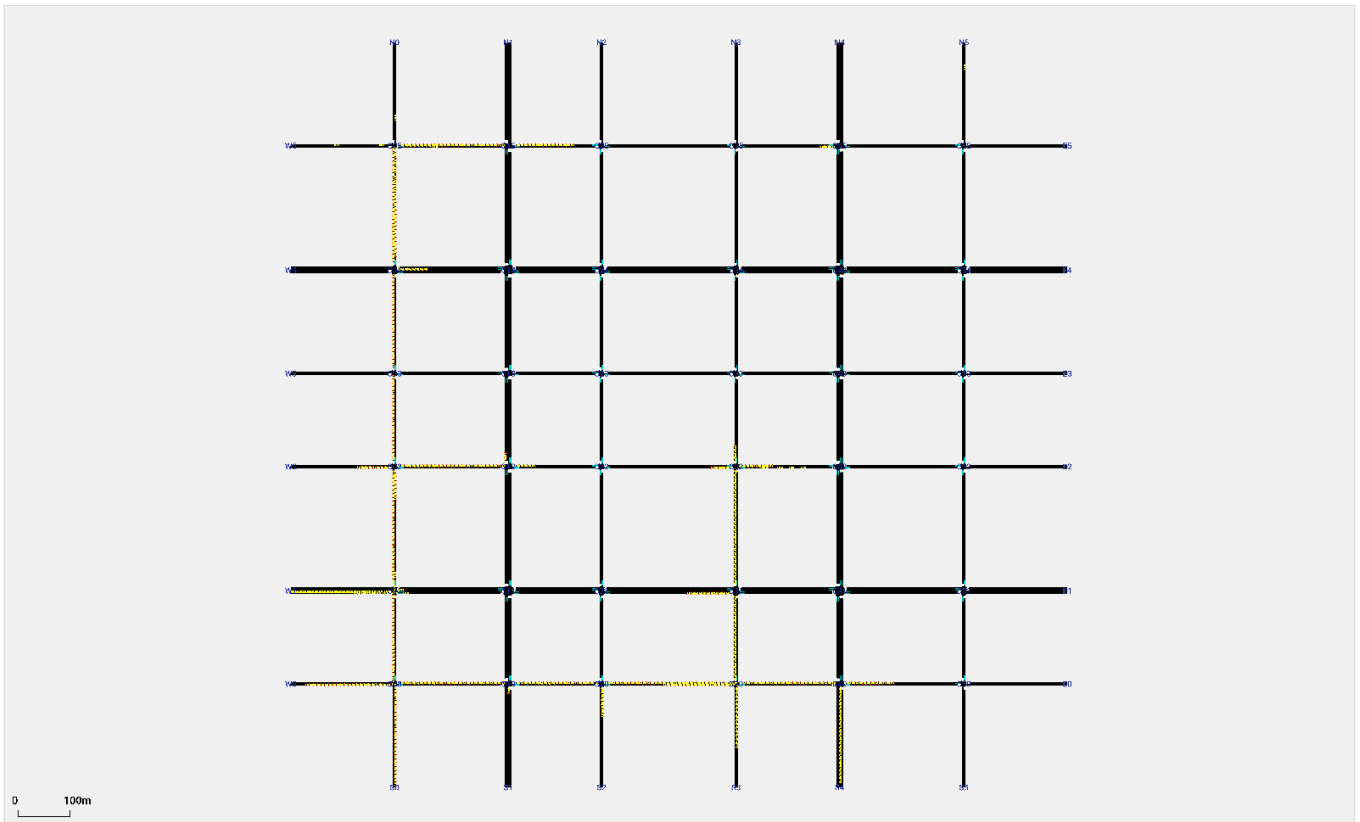
t (s)	advisory stack (shadow)	control law reason	executed
605	PREDATOR; 4 forecasts; degraded FULL; fallback false	VEHICLE_DEMAND (1 vehicles in lane_2; gap_out)	end NS green (yellow)
899	SENTRY; 1 forecasts; degraded FULL; fallback false	VEHICLE_DEMAND (1 vehicles in lane_1; gap_out)	end EW green (yellow)
912	PREDATOR; 0 forecasts; degraded FULL; fallback false	NO_DEMAND_REST_IN_GREEN	extend NS
1202	SENTRY; 0 forecasts; degraded FULL; fallback false	NO_DEMAND_LONGEST_RED	green NS
1504	RuntimeError: audit fault injection: advisory orchestrator unavailable	VEHICLE_DEMAND (2 vehicles in lane_2; gap_out)	end NS green (yellow)
1509	RuntimeError: audit fault injection: advisory orchestrator unavailable	VEHICLE_DEMAND (4 vehicles in lane_2)	green EW
1800	RuntimeError: audit fault injection: advisory orchestrator unavailable	NO_DEMAND_REST_IN_GREEN	extend EW

Signal sequence with the faults vs. production VECTOR without any shadow, same seed: **identical**
25c6cbda9f43b8011fa3c26b... vs 25c6cbda9f43b8011fa3c26b...

What this shows: the advisory stack fails and is recorded, and control continues unchanged. What it does not show: a fallback that improves anything. VECTOR has no supported fallback for advice it never used.

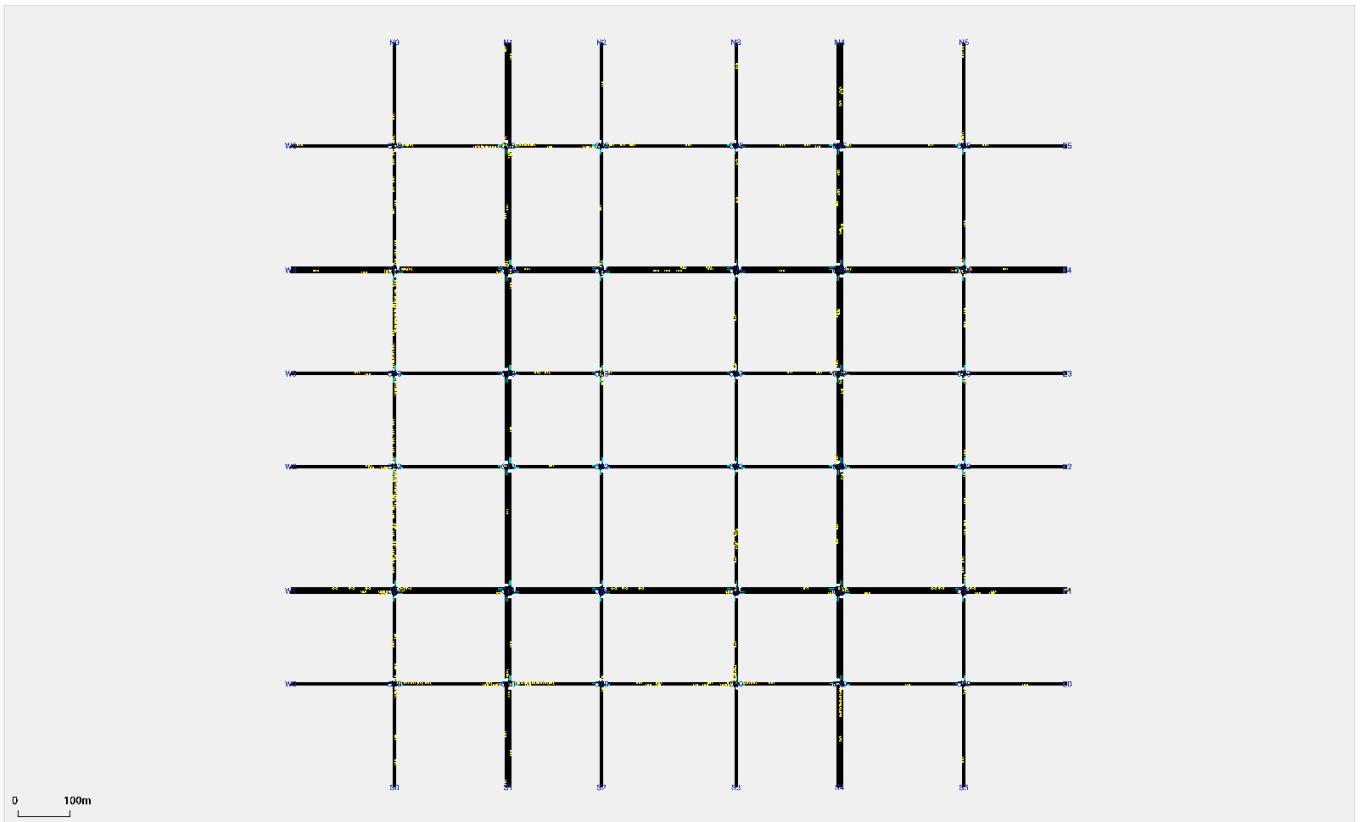
- **What it is:** audit dashboard rendering recorded results.
- **Run:** `M/vector/95000` (scenario M, arm vector, seed 95000); captured 2026-10-25T18:33:25.111Z.
- **Claim it supports:** advisory-module failure is caught and counted, and the executed signal sequence equals VECTOR without the shadow; but after the forecasters fail the stack still reports 'degraded FULL; fallback false' with 0 forecasts, so it does not detect its own degradation
- **Limitation:** shows isolation, not graceful degradation: the failed modules never influenced control in the first place
- **Linked data:** `artifacts/final-adversarial-audit/traces/heldout/M_vector_95000.jsonl.gz`, `artifacts/final-adversarial-audit/trials/heldout/M/vector/95000.json`, `artifacts/final-adversarial-audit/trials/heldout/M/vector_noshadow/95000.json`
- **sha256:** `b1da9eafe12575bcb933f1f685f7c929ebe397d65f90e4d221c5b002813d086a`

E20: end of the combined stress run (2691 s): demand stopped at 2100 s, yet 601 vehicles are still in the network (569 halting at approaches); it did not drain



- **What it is:** sumo-gui screenshot of the running simulation.
- **Run:** 0/vector/95000 (scenario O, arm vector, seed 95000); sim time 2691.0 s; captured 2026-10-25T18:23:11.914578+00:00.
- **Claim it supports:** under combined stress VECTOR leaves a residual queue at the end of the run
- **Limitation:** final metrics are in the results tables
- **Linked data:** artifacts/final-adversarial-audit/trials/heldout/0/vector/95000.json, artifacts/final-adversarial-audit/traces/heldout/0_vector_95000.jsonl.gz
- **sha256:** 3c31c958b8f865bf277141e6fdd016316f037582457843ece4a863c2bfd0cbe3
- **State recorded at capture:** halting on approaches 569; C33: GREEN EW (last decision: NO_DEMAND_REST_IN_GREEN); C32: YELLOW NS (last decision: VEHICLE_DEMAND (13 vehicles in lane_2; max_out at 45s)); C23: YELLOW NS (last decision: NO_DEMAND_REST_IN_GREEN)

E21: reference scenario R0 at 601 s under SUMO's native actuated baseline: the traffic state with no VECTOR and no event



- **What it is:** sumo-gui screenshot of the running simulation.
- **Run:** `R0/native_actuated/95000` (scenario R0, arm native_actuated, seed 95000); sim time 601.0 s; captured 2026-10-25T18:03:55.881273+00:00.
- **Claim it supports:** the baseline network state before any intervention or disturbance
- **Limitation:** one instant of one seed
- **Linked data:** `artifacts/final-adversarial-audit/trials/heldout/R0/native_actuated/95000.json`, `artifacts/final-adversarial-audit/traces/heldout/R0_native_actuated_95000.jsonl.gz`
- **sha256:** `9a75291eec5848fc0c71e35e16961473078273fcc594de903fdd9dcaf1136cb5`
- **State recorded at capture:** halting on approaches 45; C33: GREEN EW; C32: GREEN EW; C23: GREEN NS

E22: Scenario E, seed 95000: ambulance, fire engine and police in every arm

VECTOR final adversarial audit rendered from results/heldout/emergency.json

Emergency vehicles: scenario E, seed 95000

vehicle	kind	route	arm	seed	state	travel (s)	time loss (s)	waiting (s)	stops	priority decisions	waits by junction (s)
amb1	ambulance	W1→E1	fixed_time_30s	95000	arrived	226.0	118.1	90.0	3	0	C01:30 C21:31 C41:29
amb1	ambulance	W1→E1	native_actuated	95000	arrived	155.0	47.1	9.0	3	0	C01:1 C41:3 C51:5
amb1	ambulance	W1→E1	vector	95000	arrived	122.0	14.1	0.0	0	29	
amb1	ambulance	W1→E1	vector_noshadow	95000	arrived	122.0	14.1	0.0	0	29	
amb1	ambulance	W1→E1	vehicle_actuated	95000	arrived	257.0	149.4	116.0	4	0	C01:93 C11:6 C41:4 C51:13
fire1	fire_engine	S2→N2	fixed_time_30s	95000	arrived	195.0	66.0	36.0	2	0	C22:15 C24:21
fire1	fire_engine	S2→N2	native_actuated	95000	arrived	226.0	96.9	46.0	4	0	C20:5 C21:13 C24:7 C25:21
fire1	fire_engine	S2→N2	vector	95000	arrived	132.0	3.3	0.0	0	45	
fire1	fire_engine	S2→N2	vector_noshadow	95000	arrived	132.0	3.3	0.0	0	45	
fire1	fire_engine	S2→N2	vehicle_actuated	95000	arrived	188.0	58.6	29.0	3	0	C20:15 C22:6 C25:8
pol1	police	S4→N4	fixed_time_30s	95000	arrived	212.0	108.0	89.0	3	0	C40:21 C42:32 C44:36
pol1	police	S4→N4	native_actuated	95000	arrived	177.0	72.9	45.0	5	0	C40:23 C41:14 C42:1 C43:1 C45:6
pol1	police	S4→N4	vector	95000	arrived	113.0	8.9	0.0	0	36	
pol1	police	S4→N4	vector_noshadow	95000	arrived	113.0	8.9	0.0	0	36	
pol1	police	S4→N4	vehicle_actuated	95000	arrived	149.0	45.2	25.0	2	0	C43:9 C44:16

priority decisions: VECTOR controller calls whose reason was EMERGENCY_OVERRIDE for this vehicle (rule controllers and the native programme have no such reason). Detection is the simulator's ground truth; ordinary vehicles do not yield.

- **What it is:** audit dashboard rendering recorded results.
- **Run:** E/*/95000 (scenario E, arm —, seed 95000); captured 2026-10-25T18:33:25.595Z.
- **Claim it supports:** VECTOR's emergency override gives the shortest trip for all three emergency vehicles, against every baseline, on this seed and in the 10-seed means
- **Limitation:** ideal detection from simulator ground truth; ordinary vehicles do not yield; one seed shown, the 10-seed result is in emergency.json
- **Linked data:** artifacts/final-adversarial-audit/results/heldout/emergency.json
- **sha256:** 2ebc273701da36caacb753a98b29d2ac2cec58c776a1507dc772e4e1fa6a6ba9

E23: Scenario L, junction C44 (approaches swapped), t 1134-1172 s: stale (10 s delayed), contradictory, dropped and invalid observations

VECTOR final adversarial audit rendered from traces/heldout/L_vector_95000.jsonl.gz

What VECTOR was shown at C44: scenario L, seed 95000, t 1134-1172 s

Each row is one real controller call. **truth** is SUMO's halting count (NS/EW); **shown** is what the controller received after the scenario's sensor model; **belief** is VECTOR's QueueObserver output that the law used. 28 of 28 decisions in the window had a sensor fault.

t (s)	truth NS/EW	shown NS/EW	sensor flags	belief NS/EW	reason (control_logic)	adapter effect
1134	0/32	30/0	delayed_s:10, swapped	21.5/0.0	VEHICLE_DEMAND (22 vehicles in lane_1; discharging 22)	extend NS
1135	0/33	30/0	delayed_s:10, swapped	22.0/0.0	VEHICLE_DEMAND (22 vehicles in lane_1; discharging 22)	extend NS
1136	0/33	30/0	delayed_s:10, swapped	22.5/0.0	VEHICLE_DEMAND (22 vehicles in lane_1; discharging 22)	extend NS
1137	0/34	30/0	delayed_s:10, swapped	23.0/0.0	VEHICLE_DEMAND (23 vehicles in lane_1; discharging 23)	extend NS
1138	0/34	30/0	delayed_s:10, swapped	23.5/0.0	VEHICLE_DEMAND (24 vehicles in lane_1; discharging 24)	extend NS
1139	0/34	31/0	delayed_s:10, swapped	24.0/0.0	VEHICLE_DEMAND (24 vehicles in lane_1; discharging 24)	extend NS
1140	0/34	31/0	delayed_s:10, swapped	24.5/0.0	VEHICLE_DEMAND (24 vehicles in lane_1; discharging 24)	extend NS
1141	0/36	31/0	delayed_s:10, swapped	25.0/0.0	VEHICLE_DEMAND (25 vehicles in lane_1; discharging 25)	extend NS
1142	0/36	32/0	delayed_s:10, swapped	25.5/0.0	VEHICLE_DEMAND (26 vehicles in lane_1; discharging 26)	extend NS
1143	0/36	32/0	delayed_s:10, dropout, stale	26.0/0.0	VEHICLE_DEMAND (26 vehicles in lane_1; discharging 26)	extend NS
1144	0/36	32/0	delayed_s:10, swapped	26.5/0.0	VEHICLE_DEMAND (26 vehicles in lane_1; discharging 26)	extend NS
1145	0/36	33/0	delayed_s:10, swapped	27.0/0.0	VEHICLE_DEMAND (27 vehicles in lane_1; discharging 27)	extend NS
1146	0/36	33/0	delayed_s:10, swapped	27.5/0.0	VEHICLE_DEMAND (28 vehicles in lane_1; discharging 28)	extend NS
1147	0/37	34/0	delayed_s:10, swapped	28.0/0.0	VEHICLE_DEMAND (28 vehicles in lane_1; discharging 28)	max green: NS terminated (yellow)
1152	0/38	36/0	delayed_s:10, swapped	30.5/0.0	VEHICLE_DEMAND (30 vehicles in lane_1)	green NS
1160	0/40	38/0	delayed_s:10, swapped	34.5/0.0	VEHICLE_DEMAND (34 vehicles in lane_1; discharging 34)	extend NS
1161	0/41	38/0	delayed_s:10, swapped	35.0/0.0	VEHICLE_DEMAND (35 vehicles in lane_1; discharging 35)	extend NS
1162	0/41	38/0	delayed_s:10, swapped	35.5/0.0	VEHICLE_DEMAND (36 vehicles in lane_1; discharging 36)	extend NS
1163	0/41	38/0	delayed_s:10, swapped	36.0/0.0	VEHICLE_DEMAND (36 vehicles in lane_1; discharging 36)	extend NS
1164	0/41	38/0	delayed_s:10, swapped	36.5/0.0	VEHICLE_DEMAND (36 vehicles in lane_1; discharging 36)	extend NS
1165	0/41	38/0	delayed_s:10, dropout, stale	37.0/0.0	VEHICLE_DEMAND (37 vehicles in lane_1; discharging 37)	extend NS
1166	0/41	38/0	delayed_s:10, dropout, stale	37.5/0.0	VEHICLE_DEMAND (38 vehicles in lane_1; discharging 38)	extend NS
1167	0/41	39/0	delayed_s:10, swapped	38.0/0.0	VEHICLE_DEMAND (38 vehicles in lane_1; discharging 38)	extend NS
1168	0/41	40/0	delayed_s:10, swapped	38.5/0.0	VEHICLE_DEMAND (38 vehicles in lane_1; discharging 38)	extend NS
1169	0/41	40/0	delayed_s:10, swapped	39.0/0.0	VEHICLE_DEMAND (39 vehicles in lane_1; discharging 39)	extend NS
1170	0/42	40/0	delayed_s:10, swapped	39.5/0.0	VEHICLE_DEMAND (40 vehicles in lane_1; discharging 40)	extend NS
1171	0/42	41/0	delayed_s:10, swapped	40.0/0.0	VEHICLE_DEMAND (40 vehicles in lane_1; discharging 40)	extend NS

- **What it is:** audit dashboard rendering recorded results.
- **Run:** L/vector/95000 (scenario L, arm vector, seed 95000); sim time 1134 s; captured 2026-10-25T18:33:26.688Z.
- **Claim it supports:** VECTOR acts on contradictory input without detecting it: at C44, 32-42 vehicles queue on EW while the swapped, 10 s-delayed view shows them on NS; VECTOR extends NS green until max green ends it, then serves NS again
- **Limitation:** the corruption is the audit's model; the ExtremisCanonical adapter drops the stale flag, so VECTOR cannot see it; one junction, one window
- **Linked data:** artifacts/final-adversarial-audit/traces/heldout/L_vector_95000.jsonl.gz
- **sha256:** 1b92b662602c82e3e75dfdd8dace31e58d774a2b627f9fc945c575c682c2a8e0

E24: Maximum red wait: VECTOR minus native, per scenario

VECTOR final adversarial audit rendered from results/heldout/comparisons.json, results/heldout/summary.json

Measured comparison, held-out seeds 95000-95009: VECTOR - native_actuated on S14_max_red_wait_s

scenario	paired seeds	VECTOR	native_actuated	difference	95% CI (t)	relative	seeds better/worse	Wilcoxon p	Holm p	at 95%
R0	10	105.00	55.00	50.00	[50.00, 50.00]	90.9%	0/10	0.0020		VECTOR worse
A	10	105.00	55.00	50.00	[50.00, 50.00]	90.9%	0/10	0.0020		VECTOR worse
B	10	105.00	55.00	50.00	[50.00, 50.00]	90.9%	0/10	0.0020		VECTOR worse
C	10	105.00	55.00	50.00	[50.00, 50.00]	90.9%	0/10	0.0020		VECTOR worse
D	10	105.00	55.00	50.00	[50.00, 50.00]	90.9%	0/10	0.0020		VECTOR worse
E	10	105.00	55.00	50.00	[50.00, 50.00]	90.9%	0/10	0.0020		VECTOR worse
F	10	105.00	55.00	50.00	[50.00, 50.00]	90.9%	0/10	0.0020		VECTOR worse
G	10	105.00	55.00	50.00	[50.00, 50.00]	90.9%	0/10	0.0020		VECTOR worse
H	10	105.00	55.00	50.00	[50.00, 50.00]	90.9%	0/10	0.0020		VECTOR worse
I	10	105.00	55.00	50.00	[50.00, 50.00]	90.9%	0/10	0.0020		VECTOR worse
J	10	105.00	55.00	50.00	[50.00, 50.00]	90.9%	0/10	0.0020		VECTOR worse
K	10	105.00	55.00	50.00	[50.00, 50.00]	90.9%	0/10	0.0020		VECTOR worse
L	10	105.00	55.00	50.00	[50.00, 50.00]	90.9%	0/10	0.0020		VECTOR worse
M	10	105.00	55.00	50.00	[50.00, 50.00]	90.9%	0/10	0.0020		VECTOR worse
N	10	105.00	55.00	50.00	[50.00, 50.00]	90.9%	0/10	0.0020		VECTOR worse
O	10	105.00	55.00	50.00	[50.00, 50.00]	90.9%	0/10	0.0020		VECTOR worse

Difference = VECTOR - baseline per seed (same demand, same events). For delay, negative favours VECTOR. "not resolved": the interval contains 0.

- **What it is:** audit dashboard rendering recorded results.
- **Run:** `—` (scenario `—`, arm `—`, seed `—`); captured 2026-10-25T18:33:24.077Z.
- **Claim it supports:** VECTOR's starvation guard (90 s) lets some approaches wait about 50 s longer than under native control, in every scenario
- **Limitation:** a worst-case statistic; it says nothing about how many vehicles waited that long
- **Linked data:** `artifacts/final-adversarial-audit/results/heldout/comparisons.json`
- **sha256:** `fb1210785ca6817007a130bbd1b0e28b0a9e612728fa734e555f3e5540ed817c`

E25: Primary endpoint: VECTOR minus the simple vehicle-actuated rule controller

VECTOR final adversarial audit rendered from results/heldout/comparisons.json, results/heldout/summary.json

Measured comparison, held-out seeds 95000-95009: VECTOR - vehicle_actuated on P1_mean_time_loss_s										
scenario	paired seeds	VECTOR	vehicle_actuated	difference	95% CI (t)	relative	seeds better/worse	Wilcoxon p	Holm p	at 95%
R0	10	170.01	194.92	-24.92	[-47.27, -2.56]	-12.8%	8/2	0.0488		VECTOR better
A	10	330.06	391.67	-61.60	[-104.79, -18.41]	-15.7%	9/1	0.0039		VECTOR better
B	10	314.75	357.05	-42.30	[-105.31, 20.71]	-11.8%	5/5	0.3223		not resolved
C	10	173.80	183.17	-9.37	[-26.53, 7.79]	-5.1%	7/3	0.2324		not resolved
D	10	175.60	184.45	-8.85	[-28.04, 10.34]	-4.8%	6/4	0.4316		not resolved
E	10	173.43	185.51	-12.09	[-34.31, 10.14]	-6.5%	7/3	0.2754		not resolved
F	10	174.08	188.00	-13.92	[-42.26, 14.41]	-7.4%	7/3	0.2324		not resolved
G	10	163.65	183.89	-20.24	[-39.71, -0.77]	-11.0%	8/2	0.0645		VECTOR better
H	10	458.10	480.71	-22.62	[-56.65, 11.41]	-4.7%	6/4	0.2754		not resolved
I	10	242.72	242.23	0.49	[-53.99, 54.98]	0.2%	7/3	0.3223		not resolved
J	10	591.77	719.03	-127.26	[-186.75, -67.77]	-17.7%	9/1	0.0039		VECTOR better
K	10	453.06	417.47	35.58	[6.23, 64.94]	8.5%	2/8	0.0273		VECTOR worse
L	10	187.62	230.56	-42.94	[-60.28, -25.60]	-18.6%	9/1	0.0039		VECTOR better
M	10	170.01	194.92	-24.92	[-47.27, -2.56]	-12.8%	8/2	0.0488		VECTOR better
N	10	292.34	388.23	-95.89	[-154.23, -37.56]	-24.7%	8/2	0.0098		VECTOR better
O	10	326.53	372.07	-45.55	[-132.81, 41.71]	-12.2%	5/5	0.3750		not resolved

Difference = VECTOR - baseline per seed (same demand, same events). For delay, negative favours VECTOR. "not resolved": the interval contains 0.

- **What it is:** audit dashboard rendering recorded results.
- **Run:** `— (scenario —, arm —, seed —);` captured 2026-10-25T18:33:23.560Z.
- **Claim it supports:** against a simple actuated rule VECTOR is better in several scenarios unadjusted, with no Holm rejection
- **Limitation:** secondary family; the vehicle-actuated baseline is a simple gap-out rule on the same harness
- **Linked data:** `artifacts/final-adversarial-audit/results/heldout/comparisons.json`
- **sha256:** `16b4a1db6448c13543748d683f13e93b8760bdb3e162df317e165abccbf01ff1`

E26: Scenario K, seed 95000: three emergency vehicles in a saturated network

VECTOR final adversarial audit rendered from results/heldout/emergency.json

Emergency vehicles: scenario K, seed 95000

vehicle	kind	route	arm	seed	state	travel (s)	time loss (s)	waiting (s)	stops	priority decisions	waits by junction (s)
amb1	ambulance	W4→E4	fixed_time_30s	95000	arrived	246.0	138.2	99.0	4	0	C04:40 C24:29 C44:30
amb1	ambulance	W4→E4	native_actuated	95000	arrived	138.0	29.6	0.0	0	0	
amb1	ambulance	W4→E4	vector	95000	arrived	152.0	43.9	7.0	2	97	C04:7
amb1	ambulance	W4→E4	vector_noshadow	95000	arrived	152.0	43.9	7.0	2	97	C04:7
amb1	ambulance	W4→E4	vehicle_actuated	95000	arrived	248.0	140.1	96.0	5	0	C04:76 C24:12 C34:8
fire1	fire_engine	N2→S3	fixed_time_30s	95000	arrived	317.0	162.1	105.0	5	0	C20:2 C21:28 C22:33 C24:34 C25:8
fire1	fire_engine	N2→S3	native_actuated	95000	arrived	331.0	175.8	97.0	7	0	C20:4 C21:19 C22:2 C23:7 C25:65
fire1	fire_engine	N2→S3	vector	95000	arrived	304.0	149.5	58.0	6	164	C20:11 C24:16 C25:30 C30:1
fire1	fire_engine	N2→S3	vector_noshadow	95000	arrived	304.0	149.5	58.0	6	164	C20:11 C24:16 C25:30 C30:1
fire1	fire_engine	N2→S3	vehicle_actuated	95000	arrived	528.0	372.7	257.0	10	0	C20:99 C22:16 C24:14 C25:46 C30:82
pol1	police	E4→W1	fixed_time_30s	95000	running	—	1643.8	1313.0	37	0	C04:824 C14:364 C24:25 C44:26 C54:74
pol1	police	E4→W1	native_actuated	95000	arrived	1606.0	1441.3	1175.0	27	0	C01:44 C02:53 C03:42 C04:600 C14:423 C24:2 C44:11
pol1	police	E4→W1	vector	95000	running	—	1646.0	1469.0	24	1126	C04:1099 C14:354 C44:1 C54:15
pol1	police	E4→W1	vector_noshadow	95000	running	—	1646.0	1469.0	24	1126	C04:1099 C14:354 C44:1 C54:15
pol1	police	E4→W1	vehicle_actuated	95000	running	—	1647.2	1430.0	31	0	C04:671 C14:720 C24:2 C34:1 C44:25 C54:11

priority decisions: VECTOR controller calls whose reason was EMERGENCY_OVERRIDE for this vehicle (rule controllers and the native programme have no such reason). Detection is the simulator's ground truth; ordinary vehicles do not yield.

- **What it is:** audit dashboard rendering recorded results.
- **Run:** K/*/95000 (scenario K, arm —, seed 95000); captured 2026-10-25T18:33:26.082Z.
- **Claim it supports:** under saturation the override does not guarantee arrival: on this seed the police car (pol1, E4→W1) is still in the network at the end under VECTOR, vehicle-actuated and fixed-time control; only the native programme gets it through (over 10 seeds: VECTOR 5/10, native 7/10, vehicle-actuated 0/10, fixed 2/10)
- **Limitation:** one seed shown; arrival counts over 10 seeds are in emergency.json
- **Linked data:** artifacts/final-adversarial-audit/results/heldout/emergency.json
- **sha256:** 4cb54e8ca439002a52ac1872c0e4cfee3a085a7b5c0cb2b3d99719e4666c0862

E27: Scenario I, junction C41, 1243-1320 s: every decision while both lanes of the C31→C41 approach are blocked at the stop line

VECTOR final adversarial audit rendered from traces/heldout/l_vector_95000.jsonl.gz

What VECTOR was shown at C41: scenario I, seed 95000, t 1243–1320 s

Each row is one real controller call. **truth** is SUMO's halting count (NS/EW); **shown** is what the controller received after the scenario's sensor model; **belief** is VECTOR's QueueObserver output that the law used. 0 of 56 decisions in the window had a sensor fault.

t (s)	truth NS/EW	shown NS/EW	sensor flags	belief NS/EW	reason (control_logic)	adapter effect
1243	0/29	0/29	exact	0.5/29.0	VEHICLE_DEMAND (29 vehicles in lane_2; gap_out)	end NS green (yellow)
1248	0/31	0/31	exact	0.0/31.0	VEHICLE_DEMAND (31 vehicles in lane_2)	green EW
1256	1/29	1/29	exact	1.0/29.0	VEHICLE_DEMAND (29 vehicles in lane_2; discharging 29)	extend EW
1257	1/29	1/29	exact	1.0/29.0	VEHICLE_DEMAND (29 vehicles in lane_2; discharging 29)	extend EW
1258	1/29	1/29	exact	1.0/29.0	VEHICLE_DEMAND (29 vehicles in lane_2; discharging 29)	extend EW
1259	1/29	1/29	exact	1.0/29.0	VEHICLE_DEMAND (29 vehicles in lane_2; discharging 29)	extend EW
1260	1/29	1/29	exact	1.0/29.0	VEHICLE_DEMAND (29 vehicles in lane_2; discharging 29)	extend EW
1261	1/29	1/29	exact	1.0/29.0	VEHICLE_DEMAND (29 vehicles in lane_2; discharging 29)	extend EW
1262	1/29	1/29	exact	1.0/29.0	VEHICLE_DEMAND (29 vehicles in lane_2; discharging 29)	extend EW
1263	1/29	1/29	exact	1.0/29.0	VEHICLE_DEMAND (29 vehicles in lane_2; discharging 29)	extend EW
1264	1/29	1/29	exact	1.0/29.0	VEHICLE_DEMAND (29 vehicles in lane_2; discharging 29)	extend EW
1265	1/29	1/29	exact	1.0/29.0	VEHICLE_DEMAND (29 vehicles in lane_2; discharging 29)	extend EW
1266	2/29	2/29	exact	1.5/29.0	VEHICLE_DEMAND (29 vehicles in lane_2; discharging 29)	extend EW
1267	2/29	2/29	exact	2.0/29.0	VEHICLE_DEMAND (29 vehicles in lane_2; discharging 29)	extend EW
1268	2/29	2/29	exact	2.0/29.0	VEHICLE_DEMAND (29 vehicles in lane_2; discharging 29)	extend EW
1269	2/29	2/29	exact	2.0/29.0	VEHICLE_DEMAND (29 vehicles in lane_2; discharging 29)	extend EW
1270	2/29	2/29	exact	2.0/29.0	VEHICLE_DEMAND (29 vehicles in lane_2; discharging 29)	extend EW
1271	2/29	2/29	exact	2.0/29.0	VEHICLE_DEMAND (29 vehicles in lane_2; discharging 29)	extend EW
1272	2/30	2/30	exact	2.0/29.5	VEHICLE_DEMAND (30 vehicles in lane_2; discharging 30)	extend EW
1273	2/30	2/30	exact	2.0/30.0	VEHICLE_DEMAND (30 vehicles in lane_2; discharging 30)	extend EW
1274	2/30	2/30	exact	2.0/30.0	VEHICLE_DEMAND (30 vehicles in lane_2; discharging 30)	extend EW
1275	2/30	2/30	exact	2.0/30.0	VEHICLE_DEMAND (30 vehicles in lane_2; discharging 30)	extend EW
1276	2/30	2/30	exact	2.0/30.0	VEHICLE_DEMAND (30 vehicles in lane_2; discharging 30)	extend EW
1277	2/30	2/30	exact	2.0/30.0	VEHICLE_DEMAND (30 vehicles in lane_2; discharging 30)	extend EW
1278	2/30	2/30	exact	2.0/30.0	VEHICLE_DEMAND (30 vehicles in lane_2; discharging 30)	extend EW
1279	2/30	2/30	exact	2.0/30.0	VEHICLE_DEMAND (30 vehicles in lane_2; discharging 30)	extend EW
1280	2/30	2/30	exact	2.0/30.0	VEHICLE_DEMAND (30 vehicles in lane_2; discharging 30)	extend EW
1281	2/30	2/30	exact	2.0/30.0	VEHICLE_DEMAND (30 vehicles in lane_2; discharging 30)	extend EW
1282	2/31	2/31	exact	2.0/30.5	VEHICLE_DEMAND (30 vehicles in lane_2; discharging 30)	extend EW
1283	2/31	2/31	exact	2.0/31.0	VEHICLE_DEMAND (31 vehicles in lane_2; discharging 31)	extend EW
1284	2/31	2/31	exact	2.0/31.0	VEHICLE_DEMAND (31 vehicles in lane_2; discharging 31)	extend EW
1285	2/32	2/32	exact	2.0/31.5	VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)	extend EW
1286	2/32	2/32	exact	2.0/32.0	VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)	extend EW
1287	2/32	2/32	exact	2.0/32.0	VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)	extend EW
1288	2/32	2/32	exact	2.0/32.0	VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)	extend EW
1289	2/32	2/32	exact	2.0/32.0	VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)	extend EW
1290	2/32	2/32	exact	2.0/32.0	VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)	extend EW
1291	2/32	2/32	exact	2.0/32.0	VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)	extend EW
1292	2/32	2/32	exact	2.0/32.0	VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)	extend EW
1293	3/32	3/32	exact	2.5/32.0	VEHICLE_DEMAND (2 vehicles in lane_1; max_out at 45s)	end EW green (yellow)
1298	3/32	3/32	exact	3.0/32.0	VEHICLE_DEMAND (32 vehicles in lane_2)	green EW
1306	4/32	4/32	exact	4.0/32.0	VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)	extend EW
1307	5/32	5/32	exact	4.5/32.0	VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)	extend EW
1308	5/32	5/32	exact	5.0/32.0	VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)	extend EW
1309	5/32	5/32	exact	5.0/32.0	VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)	extend EW
1310	5/32	5/32	exact	5.0/32.0	VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)	extend EW
1311	5/32	5/32	exact	5.0/32.0	VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)	extend EW
1312	5/32	5/32	exact	5.0/32.0	VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)	extend EW
1313	5/32	5/32	exact	5.0/32.0	VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)	extend EW
1314	5/32	5/32	exact	5.0/32.0	VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)	extend EW
1315	5/33	5/33	exact	5.0/32.5	VEHICLE_DEMAND (32 vehicles in lane_2; discharging 32)	extend EW
1316	5/33	5/33	exact	5.0/33.0	VEHICLE_DEMAND (33 vehicles in lane_2; discharging 33)	extend EW
1317	5/33	5/33	exact	5.0/33.0	VEHICLE_DEMAND (33 vehicles in lane_2; discharging 33)	extend EW
1318	5/33	5/33	exact	5.0/33.0	VEHICLE_DEMAND (33 vehicles in lane_2; discharging 33)	extend EW
1319	5/33	5/33	exact	5.0/33.0	VEHICLE_DEMAND (33 vehicles in lane_2; discharging 33)	extend EW
1320	5/33	5/33	exact	5.0/33.0	VEHICLE_DEMAND (33 vehicles in lane_2; discharging 33)	extend EW

There is no camera or detector in this loop: "observation" is SUMO's halting count, optionally corrupted by the audit's sensor model.

- **What it is:** audit dashboard rendering recorded results.

- **Run:** `I/vector/95000` (scenario I, arm vector, seed 95000); sim time 1243 s; captured 2026-10-25T18:33:27.413Z.
- **Claim it supports:** VECTOR serves EW green to max green (45 s) on a 29-34 vehicle queue that does not discharge, then re-serves EW after clearance; its reason string labels the queue 'discharging'
- **Limitation:** one junction, one seed; the aggregate is ineffective green (S18) and the cross-cell correlation in DIAGNOSIS.md D4
- **Linked data:** `artifacts/final-adversarial-audit/traces/heldout/I_vector_95000.jsonl.gz`
- **sha256:** `19b9176ddc0358b684ce8328372651323a2aa44ec9856827b302dc92ac88813c`