

Broadcast-Gain: Two-Byte Junction Control under Partial Information

Dario Fumarola
Amazon Web Services
New York, New York

Abstract

A controller that receives little information should not withhold service longer than its no-report fallback allows. We examine this requirement in Broadcast-Gain (BG), a junction controller using two-byte reports of nearby demand. Its uncertainty-dependent switching rule violates the requirement: with an empty served axis and one waiting agent, improving report delivery can postpone the next switch from 24 to 48 steps. We derive the expected switch time and replace the conditional fallback with a deadline that reports cannot postpone. In the high-load, high-loss workload, this correction reduces 95th-percentile waiting from 173.4 to 164.9 steps while increasing throughput from 678.9 to 682.8 crossings per 1,000 steps. Phase traces explain the change: 46.0% of original phases exceed the 24-step fallback, whereas the corrected rule admits no such extension. A targeted intervention shows that the strictly low-support branch explains only a small part of the workload penalty; longer holds with inconclusive reports also matter. A separately selected 12-step timer reaches 156.8 waiting steps, but loses its advantage when clearance becomes expensive or demand is persistently unbalanced. The resulting design rule is to enforce the service deadline outside the report-dependent decision logic.

Keywords: junction control; sparse communication; partial observation; service deadlines.

1 Introduction

A junction controller must serve competing approaches without wasting too much time on phase changes. Nearby agents can help by reporting demand. Missing reports, however, do not identify an empty approach: they may instead indicate a poor communication channel. The controller needs a timing rule that remains well defined in both cases.

Broadcast-Gain (BG) uses two-byte reports to score the competing approaches. It discounts old reports, requires stronger evidence when few reports are available, and lengthens its minimum phase under low support. These choices are intended to avoid unnecessary switching. The difficulty lies in their interaction with the fallback. An empty cache invokes a four-cycle timer, while a nonempty cache can keep the phase active for eight cycles. A report may therefore disable the shorter deadline without providing enough evidence to authorize a switch.

The simplest example has one agent waiting on the unserved axis and no traffic on the served axis. With all reports erased, the controller switches after four cycles. With reliable delivery from that same agent, it waits eight. The message is timely and identifies the correct demand; the delay is caused by the control rule, not by a misleading report. We derive the expected switch time under independent losses and show that it increases with delivery probability in this example.

We then separate two questions. Does making the fallback unconditional improve the workload results? And how much of that improvement comes from the strictly low-support state in the counterexample? The first comparison preserves the original four-cycle fallback duration and changes only its applicability. A second intervention changes only the low-support branch. Phase-duration histograms and switch-reason counts distinguish these effects from simply selecting a shorter timer.

The experiments use one junction with fixed routes. Greedy movement is the primary control because legal movement has no energy or risk penalty. Learned movement and message-content controls are secondary checks.

The deadline correction improves the main high-load comparison, but the best duration depends on clearance cost and demand imbalance. The claim is therefore about keeping a service deadline active under partial observation, not about one timer being best for every workload.

Maximum green times and phase recall are established signal-control mechanisms [1]. The distinction here is where the deadline is enforced: it runs even without a detected opposing call, and surviving reports cannot cancel it. Pressure control allocates service from queue measurements [2]; our packet-pressure comparison instead uses the same incomplete local reports as BG. The analysis concerns this observation and fallback logic, rather than network stability or optimal signal timing.

2 Junction model and Broadcast-Gain

2.1 Movement, reports, and service

Four 48-cell directed lanes feed a single gate. Opposing lanes form an axis, and the controller alternates between axes $S \in \{-1, +1\}$. A switch closes the gate for c environment steps before the new axis opens. A *phase duration* includes that clearance interval. The main workload uses $c = 2$; sensitivity tests vary it. Reports are attempted every C steps, with the reporting calendar fixed in physical time.

Agents choose `MOVE` or `WAIT`. Movement into an occupied cell, or through a red or clearing gate, is masked. Occupancy is checked before simultaneous movement, so an agent cannot enter a cell vacated in the same step. When both lanes of the served axis attempt a crossing, an interlock admits the agent with greater accumulated waiting, breaking ties by identifier. At most one agent crosses per step. This interlock is identical for all methods, and its tie-breaking information is not supplied to the scheduler.

Greedy movement chooses `MOVE` whenever legal. The learned alternative uses a frozen proximal policy optimization (PPO) actor; deterministic PPO selects its larger legal logit instead of sampling. A reliable local interface publishes phase and clearance status. Only the agent reports are lossy. There is one authoritative controller, so agents do not have to reconstruct the phase from different received packets.

An agent within eight cells of the gate reports its axis s_i , distance d_i , and a waiting-progress residual:

$$x_i = \text{clip}\left(\frac{w_i}{64} - \bar{v}_i, -1, 1\right), \quad z_i = \text{round}(127x_i). \quad (1)$$

Here w_i includes waiting in the approach's input buffer, and $\bar{v}_i$ is the fraction of the preceding C steps with accepted forward progress. One signed byte stores z_i and a second byte stores axis and distance. Sender identity and sequence number are transport fields outside this payload. Reports are not acknowledged or retransmitted.

The controller keeps the newest report per sender, removing it at age $2C$ or when the sender is observed crossing. For report age a_j , define freshness $\eta_j = e^{-a_j/C}$ and decoded residual $\hat{x}_j = z_j/127$. The scores and support are

$$Q_s = \sum_{j:s_j=s} \eta_j e^{-d_j/8} \frac{1 + \hat{x}_j}{2}, \quad n_{\text{eff}} = \sum_j \eta_j, \quad u = \min(1, n_{\text{eff}}/8). \quad (2)$$

Distances are not extrapolated. Support measures the amount and age of received information, not the fraction of a queue observed. The controller does not know how many reports are missing.

2.2 The original switching rule

Let h be the number of complete reporting cycles since the last switch. The unserved axis's normalized advantage and BG's minimum hold are

$$a_s = \frac{Q_{-s} - Q_s}{Q_{+1} + Q_{-1} + 1}, \quad (3)$$

$$m(u) = \max\{2 + \lceil 2(1 - u) \rceil, \lceil c/C \rceil + 1\}. \quad (4)$$

With a nonempty cache, BG waits until $h \geq m(u)$, then switches if

$$ua_S > 0.12 + 0.20(1 - u) \quad \text{or} \quad h \geq 8. \quad (5)$$

With an empty cache, it switches at $h \geq 4$ instead. Ties hold the phase until a duration limit applies. Cache updates precede the decision; a switch resets phase age and begins clearance immediately. BG considers switches only at report boundaries.

BG scheduling denotes this rule with greedy movement and no logit adjustment. *Full BG* also applies a report-conditioned move-logit increase of at most 0.4 to the frozen PPO actor near an open gate. The appendix specifies that adjustment and the learned-policy controls. Neither is needed for the failure analysis.

3 Partial information can postpone the fallback

3.1 The single-reporter failure

The scores are nonnegative, so $a_S < 1$. An evidence-triggered switch therefore requires

$$u > 0.12 + 0.20(1 - u) \iff u > \frac{4}{15} \iff n_{\text{eff}} > \frac{32}{15}. \quad (6)$$

This condition is necessary, not sufficient. One or two effective reports cannot trigger the evidence branch, even when every report favors the unserved axis.

Consider a phase starting at time zero with an empty served axis and one stationary agent at the opposite gate. No other agents arrive. If all reports are erased, the cache stays empty and BG switches after four cycles. If the agent supplies a fresh report every cycle, $u = 1/8$. Its reports keep the cache nonempty but never pass the evidence test, so BG waits eight cycles. At $C = 6$, the switch moves from step 24 to step 48. After the two-step clearance, the waiting axis first opens at step 26 or step 50, respectively.

Let each report be delivered independently with probability $q = 1 - p$. With one sender, the cache is empty at a boundary exactly when the current and preceding attempts were both lost. Before cycle eight, the next switch is the first such boundary at or after cycle four. If $f_n(q)$ is the probability that n attempts contain no adjacent losses, then

$$f_0 = f_1 = 1, \quad f_n = qf_{n-1} + q(1 - q)f_{n-2}, \quad \mathbb{E}[\tau] = C \left(4 + \sum_{n=2}^5 f_n(q) \right), \quad (7)$$

where τ is the next switch time. At $q = 0.30$ and $C = 6$, the expectation is 31.38 steps. It rises from 24 under complete loss to 48 under perfect delivery (Figure 1). Turning a loss into a delivery can only remove an empty-cache opportunity in this example. Coupling the delivery indicators with common uniforms therefore proves that the expected switch time is nondecreasing in q . Appendix B gives the recurrence and an exhaustive enumeration check.

The example does not show that information is intrinsically harmful. The controller could ignore the extra report and retain the four-cycle schedule. Instead, its branch logic suppresses that schedule as soon as the cache becomes nonempty.

3.2 A deadline that reports cannot postpone

When a phase starts at t_0 , assign it a deadline $t_0 + H$, with integer $H > c$ measured in environment steps. Reports may trigger an earlier switch, but cannot extend the deadline. Let $E(t)$ denote BG's eligible evidence event: t is a report boundary, $u > 0$, $t - t_0 \geq Cm(u)$, and the inequality in (5) holds. The corrected controller, denoted BG-D(H), switches at

$$\tau = \min\{t_0 + H, \inf\{t > t_0 : E(t)\}\}. \quad (8)$$

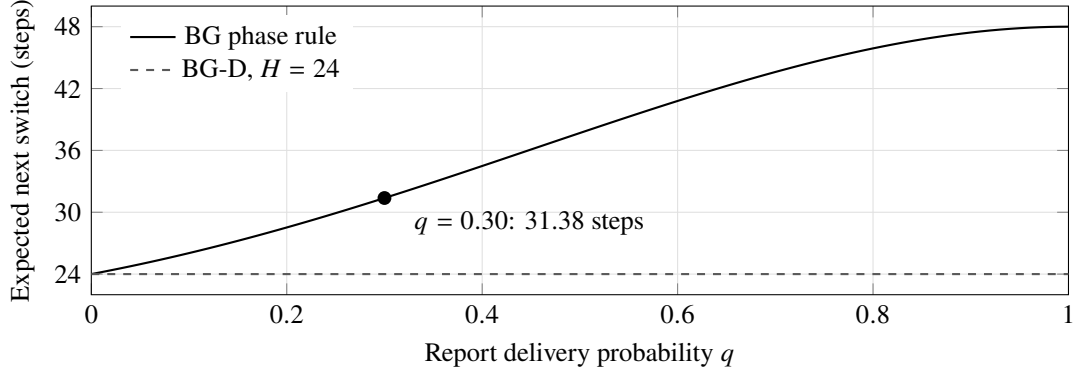


Figure 1. Exact next-switch time for the single-reporter example. The served axis is empty, one agent waits at the opposite gate, and $C = 6$. More reliable delivery keeps a weak cache alive and delays the original rule’s next switch. An unconditional deadline at 24 steps gives the same switch time for every delivery pattern. Clearance follows the switch in both cases.

The deadline replaces both the empty-cache fallback and the eight-cycle maximum. It is checked every environment step, outside the support and minimum-hold tests (Algorithm 1). Evidence thresholds, reports, and cache handling are otherwise unchanged.

Algorithm 1 BG with an unconditional physical-time deadline

Require: Phase S , start t_0 , duration $H > c$, report interval C

- 1: At step t , expire old reports and remove senders observed crossing.
 - 2: Set $E \leftarrow \text{false}$.
 - 3: **if** t is a report boundary **then**
 - 4: Insert received reports; compute $Q_s, u, a_s, m(u)$.
 - 5: $E \leftarrow [u > 0] \wedge [t - t_0 \geq Cm(u)] \wedge [ua_s > 0.12 + 0.20(1 - u)]$.
 - 6: **end if**
 - 7: **if** $t \geq t_0 + H$ or E **then**
 - 8: $S \leftarrow -S$; $t_0 \leftarrow t$; close the gate for the next c steps.
 - 9: **end if**
 - 10: Publish phase and clearance status; apply the common movement interlock.
-

Proposition 1 (Bounded interruption of service). *For every report sequence, $BG-D(H)$ switches no later than $t_0 + H$. Each axis’s uninterrupted interval without green is at most $H + c$, and each phase provides at least one non-clearance step.*

Proof. The deadline condition forces a switch at $t_0 + H$ unless evidence has already caused one. An axis that loses green at t_0 is selected again by $t_0 + H$ and opens after another c steps. A deadline-triggered phase lasts $H > c$ steps, while the minimum in $E(t)$ also exceeds clearance. Thus the return to green supplies at least one open step. $\square$

For the single reporter, $H = 24$ preserves the original no-report switch time and removes the extra delay for every delivery pattern. More generally, reports cannot make the next switch later than the same phase’s no-report deadline. This property does not order arbitrary closed-loop trajectories by their amount of information, nor does it bound every agent’s waiting behind a queue.

Ordinary actuated maximum-green timing can depend on the presence of a conflicting call; phase recall ensures that a phase is requested even without detection [1]. BG-D instead makes the deadline unconditional within the stated two-axis model. The bound depends on that placement, not on a new queue estimator.

3.3 Unequal loss also biases the scores

Keeping a deadline does not remove observation bias. For a stationary sender with an unchanged score contribution over two report cycles, the retained report is current with probability q , one cycle old with probability $(1 - q)q$, and otherwise absent. Its expected freshness weight is

$$\mathbb{E}[\eta \mathbf{1}\{\text{cached}\}] = q[1 + (1 - q)e^{-1}]. \quad (9)$$

This is 0.488 at $q = 0.40$ and 0.259 at $q = 0.20$. Identical stationary queues can therefore contribute different expected scores. A report-independent deadline limits how long biased scores can withhold a phase; it does not make the scores unbiased. The axis-dependent-loss test below measures the remaining difference in service.

4 Evaluation

4.1 Workloads, comparisons, and measurements

The main grid crosses populations $N \in \{100, 120, 140\}$, independent erasure probabilities $p \in \{0.60, 0.65, 0.70\}$, and intervals $C \in \{3, 5, 6\}$. These 27 conditions share one layout. After crossing, an agent takes an uncongested return journey lasting an integer number of steps drawn uniformly from 64 to 96, then joins an approach’s first-in, first-out buffer. Every 256 steps, the probability of assignment to the horizontal axis alternates between 0.75 and 0.25. Directions within an axis are equally likely. Population is fixed, so later arrival times depend on earlier service.

A fixed timer alternates every H physical steps without queue reports. Separate development workloads select $H = 12$. A boundary-constrained timer uses $C\lceil 12/C \rceil$, separating the duration choice from permission to switch between reports. Packet-pressure uses BG’s cache and advantage but fixed minimum hold, threshold, and maximum hold: $(m, \theta, M) = (3, 0.24, 6)$ cycles. It retains the four-cycle empty-cache fallback. Its search includes longer holds and larger thresholds than a simple deletion of BG’s support terms would allow.

We evaluate two deadline settings. BG-D($4C$) preserves each condition’s original empty-cache fallback duration and tests making it unconditional. BG-D(12) uses a physical duration selected independently of C . The first is a matched intervention on fallback logic, not a recommendation to tie deadlines to the reporting clock. A *weak-cache patch* provides a narrower diagnostic: the four-cycle fallback applies whenever $u \leq 4/15$, rather than only at $u = 0$. Above that threshold the original rule is unchanged.

For a visit ready at b_i and completed at e_i , waiting is

$$W_i = e_i - b_i - 48. \quad (10)$$

This subtracts free-flow traversal while retaining buffer and on-grid delay. Each realization has 10,000 warm-up steps, a 100,000-step scoring window, and 20,000 follow-up steps with traffic continuing unchanged. Every visit becoming ready in the scoring window enters the waiting calculation. Unfinished visits receive infinite waiting rather than being dropped; none remains unfinished in the reported outcomes. Throughput counts scoring-window crossings per 1,000 steps.

Each condition has 20 blocks of four realizations. For learned policies, each block uses a different training seed. We compute a p95 in each realization, average within blocks, then across blocks. Thus p95 waiting denotes a mean of realization-level quantiles, not a pooled percentile. Methods share initial states and indexed random inputs for routes, returns, erasures, and action sampling. Indexing preserves pairing after trajectories diverge without imposing identical arrival times.

Intervals are direct paired 95% percentile bootstrap intervals from 10,000 resamples of the 20 blocks, retaining each block’s workloads and conditions. Individual visits are not treated as independent experiments [3]. Grid averages weight conditions equally. Selection uses separate seeds and workloads at $(N, p, C) = (110, 0.50, 4)$; parameters remain fixed for all test conditions. Appendix A gives the step order, selection grids, and diagnostic definitions.

4.2 The corrected controller in the workload comparisons

At $N = 140$, $p = 0.70$, and $C = 6$, making the 24-step fallback unconditional lowers p95 from 173.4 to 164.9 steps (Table 1). The paired change is -8.5 steps, with interval $[-9.8, -7.2]$. Throughput increases by 3.9 crossings per 1,000, with interval $[3.3, 4.5]$. No report format or evidence parameter changes in this comparison.

Table 1. Scheduling at $N = 140$, $p = 0.70$, $C = 6$, and $c = 2$, with greedy movement. Differences are relative to the 12-step fixed timer. Throughput is crossings per 1,000 steps. BG-D(12) is exactly the 12-step timer in this condition, so it shares the reference row.

Scheduler	p95 wait	Throughput	Difference [95% interval]
Fixed 12 / BG-D(12)	156.8	683.6	Reference
Fixed 24	184.6	653.8	+27.8 [26.5, 29.1]
Packet-pressure	171.2	682.1	+14.4 [12.4, 16.3]
Original BG	173.4	678.9	+16.6 [15.5, 17.7]
BG, weak-cache patch	172.7	679.3	+15.9 [14.9, 16.9]
BG-D(24)	164.9	682.8	+8.1 [6.7, 9.4]

The 12-step timer still has lower waiting. BG-D(24) adds 8.1 steps relative to it, with interval $[6.7, 9.4]$, and loses 0.8 crossings per 1,000, with interval $[-1.5, -0.1]$. BG-D(12) reproduces the timer exactly at $C = 6$: its minimum evidence-triggered hold is at least 12 steps, so evidence cannot preempt the deadline. This equality is a property of the rules, not a separate performance estimate.

Across the grid, BG-D(4C) lowers waiting relative to original BG by 2.8 steps on average, with interval $[2.0, 3.6]$ for the reduction (Table 2). Its advantage over a 12-step timer is not consistent across reporting intervals. BG-D(12) instead lowers grid-average waiting by 0.7 steps relative to that timer, with interval $[0.5, 0.9]$ for the reduction. Its small difference comes from earlier evidence-triggered switches at $C = 3$ and $C = 5$; at $C = 6$ the methods coincide.

Table 2. Main-grid comparison with greedy movement. Each row averages nine population-loss conditions. “Boundary” denotes the timer with duration $C \lceil 12/C \rceil$. BG-D(4C) preserves the original no-report duration; BG-D(12) keeps its deadline fixed in physical time.

C	Fixed 12	Boundary	Pressure	BG	BG-D(4C)	BG-D(12)
<i>p95 waiting, steps</i>						
3	128.3	128.3	127.4	127.1	126.6	126.6
5	128.3	133.1	133.8	134.1	130.7	128.0
6	128.3	128.3	138.5	139.9	135.4	128.3
<i>Throughput, crossings per 1,000 steps</i>						
3	608.5	608.5	609.6	610.0	610.2	610.2
5	608.5	606.6	607.7	607.2	608.3	608.7
6	608.5	608.5	605.2	603.3	605.7	608.5

The pressure comparison is also sensitive to tuning. A narrow search over short holds and low thresholds gives stress-case p95 176.8, which would favor original BG. The expanded search gives 171.2. Across the grid, original BG minus the expanded-grid pressure rule is only +0.4 waiting steps, with interval $[-0.4, 1.2]$. A competitive simple scheduler is therefore needed before attributing a gain to support adaptation.

4.3 Which part of the fallback causes the delay?

The phase traces connect the controller change to its operation. In the stress workload, 46.0% of original BG phases continue beyond the four-cycle fallback time, and 10.0% reach the eight-cycle maximum (Figure 2). BG-D(24)

removes this long-duration group. Its mean phase length falls from 28.4 to 21.5 steps; 69.0% of its switches occur at the deadline and the rest are triggered by evidence.

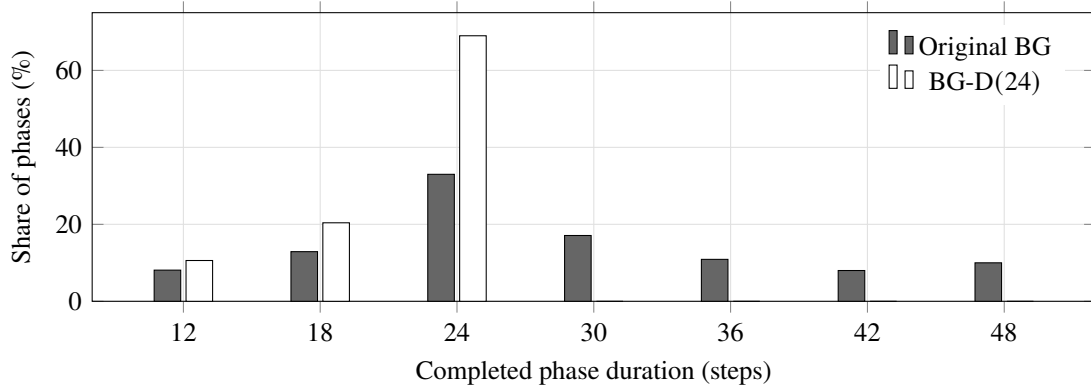


Figure 2. Phase durations in the stress workload, including clearance. The original controller can continue beyond its 24-step no-report fallback whenever the cache is nonempty. The deadline intervention removes those extensions while retaining earlier evidence-triggered switches. Each histogram is normalized within realizations before block averaging.

Strictly low support is uncommon in this crowded workload. Only 0.9% of report boundaries have a nonempty cache with $n_{\text{eff}} \leq 32/15$ (Table 3). Among suppressed four-cycle fallbacks, 1.6% lie in that range. Most extensions occur with more reports, but an advantage below the switching threshold. Thus the single-reporter example is an exact demonstration of the failure, not the dominant traffic state.

Table 3. Controller diagnostics in the stress workload. Switch reasons partition all switches; support classes partition report boundaries. A duration above 24 steps and unused green are separate measures, not additional switch reasons. Percentages use the denominators specified in Appendix A.4.

Diagnostic (%)	Original BG	BG-D(24)
<i>Reason for switching</i>		
Evidence	89.9	31.0
Empty-cache fallback	0.1	0.0
Eight-cycle maximum	10.0	0.0
Unconditional deadline	0.0	69.0
<i>Support at report boundaries</i>		
Empty: $n_{\text{eff}} = 0$	0.1	0.1
Weak: $0 < n_{\text{eff}} \leq 32/15$	0.9	1.2
Intermediate: $32/15 < n_{\text{eff}} < 8$	35.8	38.0
Support at its cap: $n_{\text{eff}} \geq 8$	63.2	60.7
Phases longer than 24 steps	46.0	0.0
Unused green with opposing gate demand	8.7	5.0

The weak-cache patch changes only states in which the evidence test cannot pass. It reduces p95 by 0.7 steps, with interval $[0.2, 1.2]$, compared with the 8.5-step reduction from the unconditional deadline. The larger intervention also shortens phases held on inconclusive, better-supported reports. Because each intervention changes subsequent traffic, these reductions do not partition individual waiting times. The larger reduction requires a deadline that also applies when there are enough reports to pass the support test.

Unused green falls from 8.7% to 5.0% of non-clearance steps. This measure counts a waiting agent at the unserved gate when neither served lane has an agent at its gate. It is obtained from the traffic trace, not supplied as an extra observation to the corrected controller.

5 Sensitivity to channel quality, demand, and clearance

5.1 Different loss rates on the two axes

The asymmetric-channel test assigns erasure probabilities 0.60 and 0.80 to the axes and repeats each realization with those labels exchanged. Routing is unchanged; the lower-loss axis accounts for between 49.9% and 50.2% of new visits across the compared controllers. Original BG gives that axis 58.0% of non-clearance service time and a 37.3-step advantage in p95 waiting (Table 4).

Table 4. Axis-dependent erasures at $N = 140$, $C = 6$, and $c = 2$, with greedy movement. “Green share” is the fraction of non-clearance time assigned to the lower-loss axis. The waiting gap is higher-loss minus lower-loss p95. Loss labels are exchanged within every paired realization.

Scheduler	Lower-loss p95	Higher-loss p95	Green share	Gap [95% interval]
Fixed 12 / BG-D(12)	156.8	156.8	50.0%	0.0 (exact)
Original BG	166.1	203.4	58.0%	37.3 [34.7, 39.9]
BG-D(24)	160.8	178.6	53.6%	17.8 [15.3, 20.5]

BG-D(24) reduces the gap to 17.8 steps and the green share to 53.6%. Higher-loss waiting falls by 24.8 steps, with interval [23.5, 26.1] for the reduction. The remaining gap is consistent with biased evidence still affecting switches before the deadline. The 12-step timer has exactly equal axis means after label exchange because erasures do not affect its trajectory. At this reporting interval BG-D(12) has the same property because it is the same scheduler.

5.2 When a shorter deadline is not the better choice

Table 5 changes one factor at a time from the stress workload. It includes stationary balanced demand, a persistent 90:10 axis split, faster alternation, clearance durations of one and six steps, and populations of 80 and 180. The controller parameters are not retuned. The 12-step and 24-step timers expose the cost of a fixed duration under these changes.

Table 5. One-factor workload sensitivity at $p = 0.70$ and $C = 6$. All methods use greedy movement. Each pair of columns gives p95 waiting in steps and throughput per 1,000 steps (rate). The reference has $N = 140$, $c = 2$, and a 75:25 axis preference alternating every 256 steps. All other factors remain at those values.

Change	Fixed 12		Fixed 24		BG		BG-D(24)	
	p95	Rate	p95	Rate	p95	Rate	p95	Rate
Reference	156.8	683.6	184.6	653.8	173.4	678.9	164.9	682.8
Steady 50:50	143.5	691.2	164.7	666.3	149.8	697.0	141.9	699.5
Steady 90:10	307.2	443.8	283.0	485.4	214.5	632.0	263.7	546.2
64-step alternation	171.6	679.8	203.7	641.1	190.2	661.5	178.4	675.3
Clearance $c = 1$	139.7	714.2	167.8	685.9	154.6	711.8	146.8	715.7
Clearance $c = 6$	280.4	477.2	181.6	655.1	202.3	663.8	192.7	641.6
Population $N = 80$	77.6	476.8	91.4	454.8	96.3	454.6	82.9	470.8
Population $N = 180$	211.6	757.5	244.8	731.4	237.9	749.8	222.4	756.2

Two changes reverse the main timing comparison. With six-step clearance, a 12-step phase spends half its time closed, imposing a throughput ceiling of 500 crossings per 1,000 steps. The 24-step timer reaches p95 181.6, compared with 280.4 for the 12-step timer. BG-D(24) reaches 192.7, adding 11.1 steps relative to the longer timer, with interval [9.0, 13.3]. Its evidence-triggered switches can incur clearance more frequently than a fixed 24-step phase.

Under persistent 90:10 demand, long phases on the busier axis are useful. Original BG gives lower waiting and higher throughput than either equal-split timer. Capping it at 24 steps increases p95 by 49.2 steps, with interval [45.6, 53.6], while reducing throughput. The cap still enforces its service bound, but gives up capacity that could serve the dominant queue. Deadline placement and deadline duration are different design choices: the first determines whether the fallback survives partial observations; the second sets a tradeoff between service frequency and sustained capacity.

The 12-step timer remains favorable in the faster-alternation, light-population, and heavy-population comparisons. Stationary balanced demand gives a small mean advantage to BG-D(24) instead. These tests vary the arrival pattern, switching cost, and congestion on the same junction. They do not cover a road network, downstream spillback, transport delay, or failure of the status interface.

6 Discussion and conclusion

The partial-information failure comes from a conditional deadline. Original BG treats a nonempty cache as sufficient reason to leave its fallback schedule, even when the received information cannot justify a switch. The single-reporter calculation makes this defect explicit: more reliable delivery increases the expected delay before service returns.

An unconditional deadline removes that behavior and bounds each axis’s uninterrupted loss of green by $H + c$. The workload intervention improves the main comparison without changing reports or evidence thresholds. Its phase traces also sharpen the explanation: strictly low-support caches are rare under high load, and most extended phases have more reports but inconclusive evidence. The general issue is allowing those observations to suppress the fallback.

The deadline does not choose its own duration. Short phases are effective when switching is inexpensive and demand alternates, while longer or unequal allocations are valuable under costly clearance and sustained imbalance. The useful design rule is to set the service deadline explicitly, keep it independent of report reception, and evaluate its duration against strong timing controls. Message precision and learned movement should then be tested for what they add to that scheduler.

References

- [1] Federal Highway Administration. *Traffic Signal Timing Manual*. Report FHWA-HOP-08-024, 2008, Chapter 5. Basic signal timing procedure and controller parameters.
- [2] P. Varaiya. Max pressure control of a network of signalized intersections. *Transportation Research Part C*, 36:177–195, 2013. doi:10.1016/j.trc.2013.08.014.
- [3] R. Agarwal, M. Schwarzer, P. S. Castro, A. Courville, and M. G. Bellemare. Deep reinforcement learning at the edge of the statistical precipice. In *Advances in Neural Information Processing Systems*, volume 34, 2021. arXiv:2108.13264.
- [4] H. Wei, C. Chen, G. Zheng, K. Wu, V. Gayah, K. Xu, and Z. Li. PressLight: Learning max pressure control to coordinate traffic signals in arterial network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019. doi:10.1145/3292500.3330949.
- [5] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. 2017. arXiv:1707.06347.
- [6] J. Schulman, P. Moritz, S. Levine, M. I. Jordan, and P. Abbeel. High-dimensional continuous control using generalized advantage estimation. In *International Conference on Learning Representations*, 2016. arXiv:1506.02438.

A Implementation and selection details

A.1 Step order, reports, and measurements

Return journeys finish at the start of a step. A ready agent joins its assigned input buffer; the buffer head enters at distance 48 when that cell is empty. Insertion consumes no movement step. Reports describe the resulting pre-movement state and, when delivered, arrive before the phase decision with no further transport delay. Expiration and removal of observed crossings precede report insertion. Phase changes precede movement, and the status interface is read every step. For the original cycle-based rules, the first scheduled decision follows one complete reporting cycle. For BG-D, a deadline is checked at every environment step, including steps without reports. A phase starting at t_0 clears during $t_0, \dots, t_0 + c - 1$ and first opens at $t_0 + c$.

Accepted progress before time t is $P_i(t)$, so online waiting is $w_i(t) = t - b_i - P_i(t)$. Completion requires 48 accepted moves. Progress history resets at each visit and is padded with zeros until C steps are available. Agents begin each realization in return transit with remaining durations drawn uniformly from zero to 96; initial phases are balanced across the four workloads. The empirical p95 is the smallest observed value whose empirical cumulative probability reaches 0.95.

Byte 0 stores $z_i \in \{-127, \dots, 127\}$; -128 is unused. Byte 1 is $128\mathbf{1}\{s_i = +1\} + d_i$, with $1 \leq d_i \leq 8$ in the reporting region. Each eligible agent attempts one report per boundary, without acknowledgment or retransmission. Offered uplink payload is $16/C$ bits per eligible agent per step, excluding identity, sequence, framing, and channel-access overhead. Fixed timers need no uplink. The common status interface carries phase, clearance, and an eight-bit gain field in two bytes per step; that field is zero when unused.

Return and route uniforms are indexed by agent and visit; erasure and action uniforms by agent and environment step. Arrays are shared across paired methods and, within each population, across p and C . These indices determine latent draws, not identical closed-loop traffic arrivals. For greedy policies, the 20 blocks contain workload variation only.

A.2 Controller selection

Test seeds are 0 through 19 and development seeds 100 through 103. All additional workload and channel tests retain the selected controllers; they are not used to revise the parameters. Each candidate uses four workloads per development seed. Selection minimizes mean p95 subject to throughput no lower than the $4C$ -step reference under the same movement policy. Ties prefer fewer switches, then the lexicographically smaller parameter tuple. Selected settings remain fixed across the test grid.

The physical-time candidates are $H \in \{6, 9, 12, 15, 18, 24, 30, 36, 48\}$, selecting 12 separately for greedy and stochastic PPO. BG-D repeats the same H search with its inherited BG evidence rule and selects 12 under greedy movement. BG-D($4C$) and the weak-cache patch are diagnostic interventions: neither receives additional tuning. The training reference selects four cycles from $\{2, 4, 6, 8\}$ at development time. BG searches minimum-hold base $m_0 \in \{1, 2\}$, stretch $\lambda_m \in \{0, 2\}$, base threshold $\theta_0 \in \{0.08, 0.12\}$, and support penalty $\theta_1 \in \{0, 0.20\}$. The minimum is $\max\{m_0 + \lceil \lambda_m(1 - u) \rceil, \lceil c/C \rceil + 1\}$ and the evidence test is $ua_S > \theta_0 + \theta_1(1 - u)$. Both movement policies select $(2, 2, 0.12, 0.20)$; the eight-cycle maximum and four-cycle fallback are fixed.

Packet-pressure searches $m \in \{2, 3, 4\}$, $\theta \in \{0.04, 0.08, 0.12, 0.24, 0.44, 0.64\}$, and $M \in \{4, 6, 8\}$, with $m \leq M$. It selects $(3, 0.24, 6)$. Its minimum also respects $\lceil c/C \rceil + 1$. The narrow comparison searches $m \in \{1, 2\}$ and $\theta \in \{0.08, 0.12\}$ with $M = 8$, selecting $(2, 0.12, 8)$.

Each report-content control repeats the BG search and replaces the denominator offset 1 in (3) by a tuned $\kappa \in \{0.25, 0.5, 1, 2, 4\}$. The selected phase parameters remain $(2, 2, 0.12, 0.20)$. The residual and distance-weighted presence select $\kappa = 1$; unweighted presence selects 2. Presence has constant score factor one, so no scalar quantization is required. The retained dummy byte makes its erasure model identical in packet size to the residual version.

A.3 Learned movement and the optional gain

The actor and a separate value network each have two 64-unit tanh layers. The actor observes an approach-oriented 5×5 occupancy patch, distance divided by 48, a four-component approach identifier, served-axis and clearance indicators, waiting clipped at 256 and divided by 256, and a last-progress indicator. Off-lane patch cells are occupied. Buffer and return-transit agents do not act. The actor has no queue-report input.

Training uses PPO [5] with generalized advantage estimation [6]. Eight environments supply 128-step rollouts, for 1,024 joint steps per update and 400 updates per seed. Each update uses four epochs and minibatches of 4,096 on-grid transitions, retaining a smaller final batch. Environments reset after 4,096 joint steps and sample (N, C) uniformly from the nine training settings, with a $4C$ -step phase. GAE ends at crossing and bootstraps at time-limit truncation. Feature and value normalization are frozen after training.

The common reward is $r_t = X_t - 0.01B_t$, where X_t counts crossings and B_t counts buffer or approach agents making no accepted progress. Adam’s learning rate is 3×10^{-4} ; discount and GAE parameters are 0.99 and 0.95. PPO clipping, entropy coefficient, value-loss coefficient, and gradient-norm limit are 0.20, 0.01, 0.50, and 0.50. Training uses neither BG nor a movement gain.

Full BG computes $g = uQ_S/(Q_{+1} + Q_{-1} + 1)$ after phase selection and holds $\tilde{g} = \text{round}(255g)/255$ until the next boundary. The move-logit shift is

$$\delta_i = 0.4\tilde{g} e^{-(d_i-1)/3} \mathbf{1}\{s_i = S, 1 \leq d_i \leq 4, \text{ gate not clearing}\}. \quad (11)$$

The wait logit is unchanged, and the same action mask applies. The fixed-bias control replaces $0.4\tilde{g}$ with $b = 0.20$, selected from $\{0.05, 0.10, 0.20, 0.30, 0.40\}$. Deterministic PPO chooses the larger legal logit, breaking a tie toward movement.

A.4 Phase diagnostics and channel comparisons

Phase-duration histograms include phases starting in the scoring window and follow each to its terminating switch. Each realization’s bin counts are normalized before averaging within and across blocks. A phase’s duration includes its initial clearance. The support fractions in Table 3 use report boundaries after cache update and before phase selection, including boundaries at which the phase is held.

Switch reasons are mutually exclusive. For original BG, an empty-cache switch is labeled fallback; otherwise the eight-cycle maximum takes precedence over the evidence test. For BG-D, the physical deadline takes precedence over evidence. A suppressed fallback is recorded when the original phase reaches age $4C$ and is held. Its cache is necessarily nonempty. The weak subset additionally satisfies $n_{\text{eff}} \leq 32/15$. The patch changes the empty-cache test to $u \leq 4/15$ at $h \geq 4$ and leaves the other branches unchanged.

Unused green is the fraction of non-clearance steps with at least one agent at distance one on the unserved axis and none at distance one on the served axis. These counts use the underlying traffic state for diagnosis only; they are not inputs to any scheduler. Per-axis waiting uses the same visit-cohort and realization-level quantile definitions as aggregate waiting.

For the asymmetric channel, each realization is evaluated twice with the erasure probabilities 0.60 and 0.80 exchanged between axes. The two label assignments are averaged within the realization before block resampling. Routing and return inputs are unchanged by the exchange. Report-independent controllers have identical trajectories under the two assignments, so their lower-loss and higher-loss axis means are exactly equal after averaging. The sensitivity tests in Table 5 change one factor at a time from the stress condition and use the same 20 blocks and four realizations per block.

B Exact single-reporter calculation

Write $D_h = 1$ when the report at cycle h is delivered, with independent probability q . Before the first switch the sole agent remains at the opposite gate. The cache at cycle h is nonempty exactly when $D_h = 1$ or $D_{h-1} = 1$. Its support never exceeds $1/8$, so the imbalance branch is inactive for every residual value.

Let $T = \tau/C$. Then T is the first $h \in \{4, 5, 6, 7\}$ with $D_{h-1} = D_h = 0$, or 8 if there is no such h . For $k = 4, 5, 6, 7$, the event $T > k$ requires that the string $D_3, \dots, D_k$ contain no adjacent losses. Hence $\Pr(T > k) = f_{k-2}(q)$ and the tail-sum identity gives $\mathbb{E}[T] = 4 + f_2 + f_3 + f_4 + f_5$.

For $n \geq 2$, a loss-free-adjacency string either begins with a delivery, followed by a valid string of length $n - 1$, or begins with a loss and a delivery, followed by a valid string of length $n - 2$. These disjoint cases give the recurrence in (7). Turning any loss in a fixed string into a delivery can only remove an empty-cache opportunity, so T is nondecreasing under that change. Coupling delivery indicators with common uniforms proves that its expectation is nondecreasing in q . The curve in Figure 1 evaluates the exact recurrence; enumerating all 2^8 eight-cycle delivery strings gives the same values.

C Message-content and learned-movement controls

These controls separate the scheduler from the report residual and the optional movement gain. They use the original BG scheduling family, not BG-D, and remain secondary to the deadline comparison.

The residual in (1) saturates at one when $w_i \geq 64(1 + \bar{v}_i)$. A stationary agent saturates after 64 waiting steps, and saturation occurs by 128 even at maximal recent progress. Saturated reports contribute only a freshness- and distance-weighted count. The fractions of attempted reports at saturation are 57.6%, 66.9%, and 74.8% at populations 100, 120, and 140, averaged over the nine loss-interval settings at each population. These fractions are measured before quantization, regardless of delivery.

Distance-weighted presence sets every residual to one; unweighted presence also removes the distance factor. Each control receives its own development search and retains a dummy byte to match payload size. In the stress workload, distance-weighted presence changes p95 by +0.5 steps (Table 6); the grid-wide difference is +0.3, with interval $[-0.1, 0.7]$. The waiting-progress byte adds little to this scheduling comparison.

Table 6. Secondary controls at $N = 140$, $p = 0.70$, $C = 6$, with the original BG scheduling family. The first panel uses greedy movement and compares report contents; the second compares movement rules. Each panel has its own reference. Throughput is per 1,000 steps. Changing movement or report contents also changes later closed-loop reports and phase choices.

Control	p95 wait	Throughput	Difference [95% interval]
<i>Report contents, with separate development tuning</i>			
Waiting-progress residual	173.4	678.9	Reference
Distance-weighted presence	173.9	678.5	+0.5 $[-0.2, 1.2]$
Unweighted presence	175.2	675.4	+1.8 $[0.5, 3.1]$
<i>Movement, with the residual-based scheduler</i>			
Stochastic PPO, no gain	183.1	658.8	Reference
Stochastic PPO, report gain	180.0	664.5	-3.1 $[-4.2, -2.1]$
Stochastic PPO, fixed bias	179.3	665.2	-3.8 $[-5.2, -2.5]$
Deterministic PPO, no gain	174.1	677.6	-9.0 $[-13.4, -4.4]$
Greedy movement	173.4	678.9	-9.7 $[-14.4, -4.9]$

For stochastic PPO, a fixed near-gate bias performs similarly to the report-conditioned gain. Their direct report-gain-minus-fixed-bias difference is +0.7 waiting steps, with interval $[-0.3, 1.7]$. Removing action sampling has a larger effect: deterministic PPO reaches 174.1 steps, close to greedy movement's 173.4. These are controls

for avoidable hesitation in a fixed-route task. Learning which signal phase to select, as in PressLight [4], is a different decision problem.

The stochastic actor also illustrates the timing-reference issue (Figure 3). Full BG improves stress-case p95 from 197.9 to 180.0 relative to the 24-step training schedule, but the same frozen actor with a 12-step timer reaches 168.4. The BG-minus-timer difference is +11.6 steps, with interval [9.4, 13.7]. Throughputs are 632.5 for the 24-step timer, 664.3 for the 12-step timer, and 664.5 for full BG. The full-BG-minus-12-step throughput difference is +0.2, with interval $[-1.6, 2.0]$. Increasing the reporting interval also lengthens the $4C$ reference, so its widening gap from BG does not isolate a communication benefit.

Under the axis-dependent erasure test, full BG with stochastic PPO gives lower-loss and higher-loss p95 values of 179.3 and 216.7 steps. The lower-loss axis accounts for 50.2% of new visits and 58.2% of non-clearance service time. The greedy comparisons in Table 4 remove the movement gain from that channel-bias test.

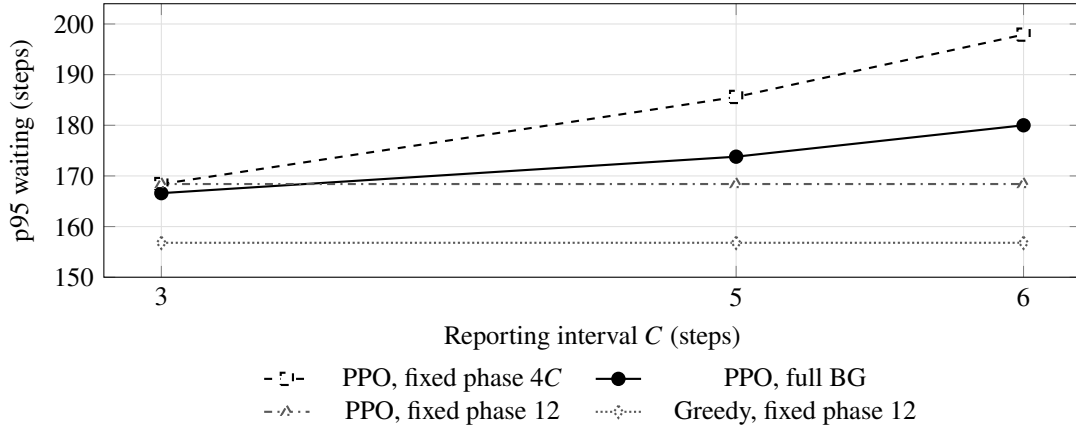


Figure 3. The reference timer changes the apparent value of communication. At $N = 140$ and $p = 0.70$, full BG’s advantage over the $4C$ -step reference grows as that reference becomes slower. The report-independent timers stay at the same operating points. Points are block means; Appendix C gives the corresponding stress-case controls.