Dwell Is Four Intervals: Queue Discipline, Slot Reservation, and Where a Truck Actually Waits

Architecture, Computational Methods, and Field Evidence from MileSoft Truck Turnaround

MileSoft Engineering Research Group

MileSoft Software Technologies

https://milesoft.net/

Corresponding authors: *Deepak Daryani* (deepakdaryani@milesoft.net), *Tanuj Daryani* (tanujdaryani@milesoft.net)

Working paper — Version 1.0 — 24 August 2026

Permanent link: https://milesoft.net/research/products/truck-turnaround

Abstract

Stand at the gate of a busy distribution centre between six and ten in the morning and the failure mode is visible from the road: fifteen trucks idling outside while two docks sit empty inside. That is not a capacity problem. It is a queue-discipline problem, and it is solvable with arithmetic that has been settled since the 1950s.

This paper presents the architecture and computational methods of MileSoft Truck Turnaround, which owns the interval from arrival at the gate to assignment at a door. We decompose dwell into four intervals — gate processing, yard queue, dock service and exit — and show that they respond to entirely different interventions, so a single dwell figure conceals which one a site should buy.

Three results follow. Yard occupancy is Little's Law applied to vehicles, which converts a dwell target into a parking requirement and makes the common decision to enlarge a yard visibly the wrong response to a queueing problem. Priority classes are not free: Cobham's non-preemptive result gives the delay a priority lane imposes on everything behind it, which is computable before the policy is adopted rather than discovered afterwards. And a reserved slot is neither free nor in use, so reservation lead time subtracts from effective yard capacity in proportion to how early the commitment is made.

A multi-client third-party logistics deployment reported average dwell falling from three hours ten minutes to one hour twenty, gate to gate. That result is attributed in its own source to this module operating with the dock module on a shared yard state, and it is reported here as their joint outcome rather than this module's alone. The product's own captures are a demonstration build with synthetic data; no figure from them is cited as evidence anywhere in this paper.

Keywords: Truck turnaround time; Yard management system; Truck dwell time; Gate queue management; ANPR gate integration; Yard slot reservation; Priority queue discipline; Carrier portal; Detention cost; Dock appointment coordination; Distribution centre yard; Gate-to-gate tracking

1 Introduction

The universal distribution-centre failure mode is visible from outside the fence. A queue of trucks idles at the gate, drivers on phones, dispatchers texting yard staff who text dock staff who reshuffle the morning's appointments by hand. Inside, two docks sit empty.

The gap between trucks waiting outside and doors idle inside is the turnaround problem, and the product's own material names its nature correctly: it is a coordination failure rather than a capacity failure. Most sites have enough doors. What they lack is a discipline governing the queue between the gate and them.

This paper takes that diagnosis and makes it quantitative. Queueing theory settled the relevant results in the 1950s, and every one of them is computable from data a gate already produces — arrival times, service times, and which vehicle was let in ahead of which.

1.1 What this paper is about, and what it is not

The turnaround problem spans two systems and it is worth drawing the line between them before any of the arithmetic.

This paper concerns the interval from a vehicle reaching the gate to it being assigned a door: identity capture, queue position, slot allocation and sequencing. What happens at the door — appointment windows, door capability, load readiness, the physical turn — belongs to the dock system and is treated in the companion paper on loading-dock management.

The two share one yard state, which is why the deployment result of Section 5.2 belongs to both. But the decisions are different: this system decides who goes next, and the dock system decides how long that takes once they get there.

> A site that improves door service and leaves gate sequencing alone will find its queue outside unchanged. A site that improves sequencing and leaves the door alone will deliver trucks to a bottleneck faster. Both are common; neither is a solution.

1.2 Contributions

1. Dwell decomposed into four intervals with different causes and different remedies (Section 4.1).
2. Yard occupancy as Little's Law, converting a dwell target into a parking requirement (Section 4.2).
3. The delay a priority class imposes on ordinary traffic, from Cobham's non-preemptive result (Section 4.3).
4. Reservation lead time expressed as a subtraction from effective yard capacity (Section 4.4).

5. An eight-dimension capability reference framework for yard and turnaround systems (Section 6.3).

2 Background and Related Work

A yard is a queueing system with an unusually clean structure: single-class or multi-class arrivals, a finite pool of servers, and a physical waiting room whose size is a hard constraint. Almost everything a turnaround system needs was derived before distribution centres existed.

2.1 The vocabulary a yard already fits

Kendall's 1953 notation classifies a queueing system by its arrival process, its service distribution and its server count. A distribution centre yard is a multi-server system with a finite waiting room, and naming it that way makes the relevant results locatable rather than reinvented.

Little's Law supplies the relation between the three quantities a yard manager cares about — how many trucks are here, how fast they arrive, how long each stays — and it holds in steady state regardless of the arrival or service distribution. Section 4.2 uses it to convert a service promise into a physical requirement.

Kingman's heavy-traffic approximation supplies the shape of the problem: waiting time rises with utilisation and with variability, and it rises without bound as utilisation approaches one. This is the formal statement of something every dock supervisor already knows — a yard running at ninety-five per cent of door capacity does not queue slightly more than one at eighty, it queues catastrophically more.

> Variability is half the equation and the half a site can usually act on. Two carriers delivering the same volume, one to a schedule and one at random, do not impose the same queue.

2.2 What a priority class actually costs

Cobham's 1954 result gives the mean waiting time of each class in a non-preemptive priority queue, where a job in service is never interrupted but the next job is chosen by class rather than by arrival.

Its consequence is the one operations teams most often discover the hard way. Adding a priority class does not create capacity; it redistributes waiting. The high class waits less, every class below waits more, and the total work in the system is unchanged. A commercial promise of priority handling to one carrier is therefore a promise to lengthen everybody else's queue by a computable amount.

The product's gate screen carries a priority level as a field on every vehicle, which means the policy exists and can be set. Section 4.3 gives the arithmetic that should accompany setting it.

2.3 Identity, custody and the clock a dispute runs on

A dwell figure is only as good as its endpoints. Both are events about a vehicle, so the yard needs vehicle identity resolved automatically enough that a timestamp is not a matter of when somebody got round to writing it down.

ISO 28000 frames the gate as an access-control and custody boundary, ISO/IEC 15459-1 covers transport-unit identity, and EPCIS with its vocabulary gives the event shape so that arrival at one site is the same recorded fact as arrival at another. Incoterms 2020 decides when custody passes, which is what a detention dispute is ultimately argued about.

This matters commercially rather than only technically. Detention charges are settled against timestamps, and a site whose arrival time is a driver's word against a gatehouse log will lose those arguments regardless of what actually happened.

3 System Overview

The system is a gate, a slot pool and a queue, with a driver-facing display and an analytics layer over the events all three produce. The captured build presents these as gate entry, yard display, parking, an administrative queue view, analytics and device controls.

> The captures in this section come from a demonstration instance. Its counters read zero, its carrier names are placeholders and its traffic chart is labelled as simulated. They are reproduced to show what the system measures, and no number visible in them is treated as evidence anywhere in this paper.

3.1 The gate, and three ways to know which truck this is

A vehicle is admitted against a record carrying its registration, driver, carrier, cargo type and a priority level. Identity can be established three ways: keyed by the gatehouse, read from a radio-frequency tag, or recognised from the plate automatically.

The three routes are not interchangeable and the difference is the subject of Section 4.1. Automatic recognition removes the manual step entirely, which the product's material puts at three to five minutes per vehicle; a tag read is fast but requires the fleet to be tagged; keyed entry always works and is the slowest.

Priority is captured at this moment, before the vehicle joins the queue, which is the correct place for it. A priority assigned after queueing has begun is a re-sequencing decision with a different and larger cost.

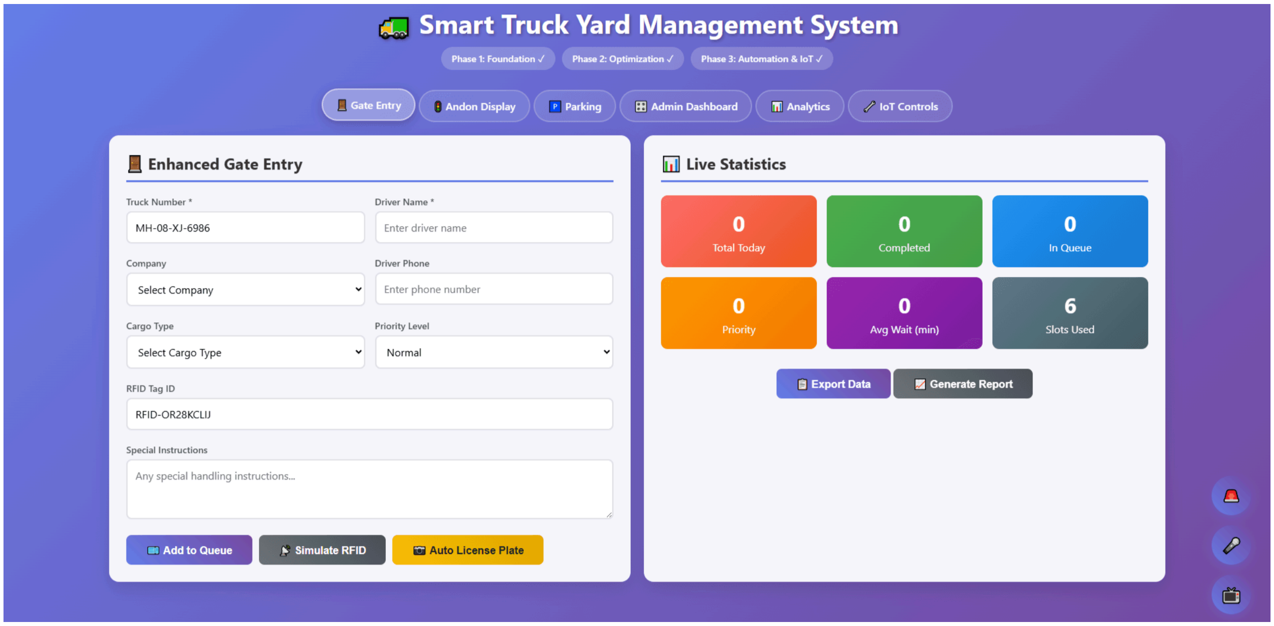

Figure 1. Product capture from a demonstration build: the gate entry screen. A vehicle record carries registration, driver, carrier, cargo type, a priority level and a radio-frequency tag identifier, with actions to add the vehicle to the queue or to acquire its identity by tag read or automatic plate recognition. The statistics panel beside it tracks arrivals today, completions, queue depth, priority count, mean wait and slots in use.

3.2 The slot pool and its three states

The yard is a fixed pool of numbered slots, each in one of three states: available, occupied, or reserved. Assignment can be manual or automatic.

The reserved state repays attention. It holds a slot against a vehicle that has not arrived, which is what allows the gate to make a promise. It is also, for as long as it lasts, capacity that neither serves a truck nor can be given to one — and Section 4.4 states the rate at which that trade runs.

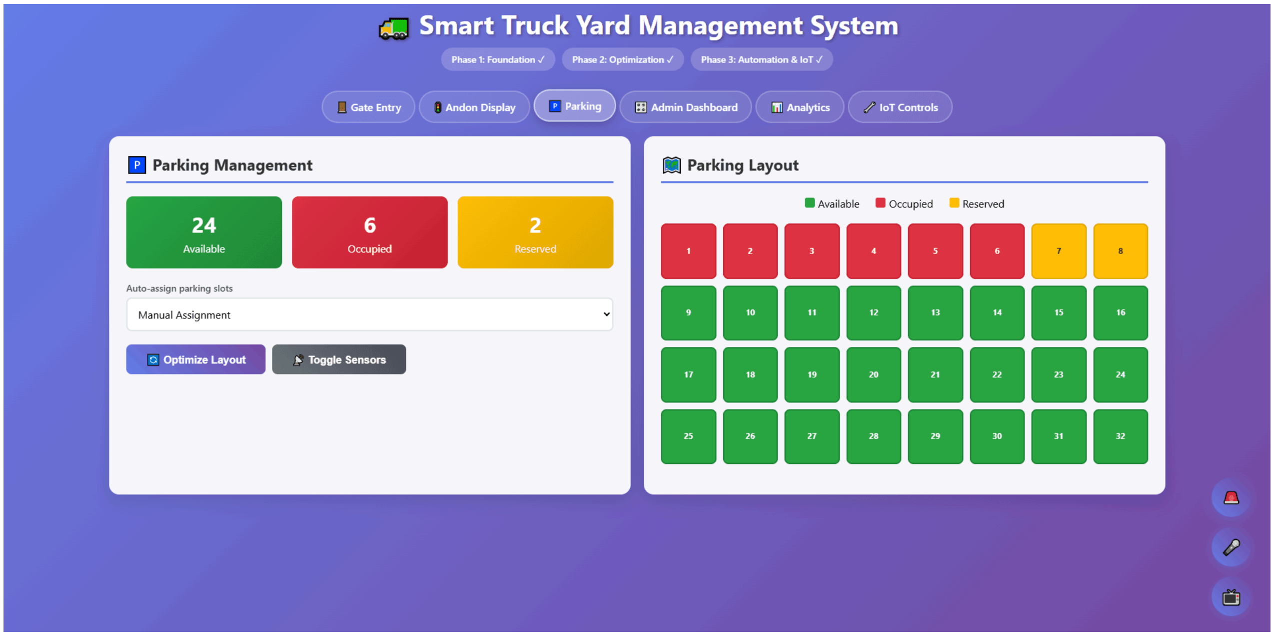

Figure 2. Product capture from a demonstration build: the yard slot pool. Slots are numbered and colour-coded across three states — available, occupied and reserved — with counters above and a selector switching between manual and automatic assignment. The reserved state is what lets the gate commit a slot before a vehicle arrives.

3.3 The display, and why the driver is part of the system

A yard display shows queue status to drivers, and a carrier-facing portal shows appointment status and queue position to dispatchers without a phone call to the gate. The product's material puts the reduction in inbound calls at sixty to eighty per cent.

The mechanism is worth stating because it is easy to dismiss as a convenience. A driver who cannot see their position generates work — a call, a walk to the gatehouse, an interruption to the person doing the sequencing. That work falls on the resource least able to absorb it, so displaying queue position does not merely inform the driver; it protects the throughput of the gate.

The same reasoning applies at shift handover. The live yard state transfers with every vehicle's position, assignment and elapsed clock intact, so the incoming supervisor inherits the queue rather than a printout of what it was twenty minutes ago.

3.4 What the system measures

The analytics layer reports turnaround by carrier, identifies the peak hour, and scores the yard's efficiency. Its structure matters more than any value it happens to display.

Measuring by carrier is what makes dwell attributable, which is the product's own framing and the correct one. A dwell figure that cannot be split between the warehouse's delay, the carrier's lateness and load readiness is a number nobody can act on and everybody can dispute.

Figure 3. Product capture from a demonstration build with synthetic data — the chart area is labelled simulated and the carrier names are placeholders. The structure is what this figure shows: peak-hour identification, mean turnaround, an efficiency score, and per-carrier turnaround with a vehicle count, which is what makes dwell attributable rather than merely aggregate.

4 Computational Methods

Four computations carry the paper. The first says what dwell is made of; the other three say what a site can do about each part.

4.1 Dwell is four intervals, not one

Gate-to-gate dwell is the interval a carrier bills against and the one a site reports. It is also the sum of four quantities with four different owners.

$$dwell = (queued - arrived) + (at\ dock - queued) + (released - at\ dock) + (departed - released), \text{ that is: } gate\ processing + yard\ queue + dock\ service + exit \qquad \text{(dwell)}$$

every endpoint is a system-generated event. t(arr) is the moment of arrival at the gate, t(q) the moment identity is resolved and the vehicle joins the queue, t(dock) the moment it is called to a door, t(rel) the moment the door releases it, and t(dep) the moment it clears the gate.

Table 1. The four intervals, who owns each, and what shortens it. No single intervention shortens more than one, which is why a combined dwell figure cannot direct an investment.

Interval	Owner	What shortens it
Gate processing	The gate	Automatic identity capture; the product's material puts this at 3-5 minutes per vehicle
Yard queue	The sequencing policy	Discipline, priority design and reservation policy — Sections 4.2 to 4.4
Dock service	The door and the load	Load readiness, door capability, equipment — the companion dock paper
Exit	The gate again	Paperwork removal, automatic departure capture

The practical consequence is that two sites reporting the same three-hour dwell can need opposite things. One with two hours in the yard queue needs sequencing; one with two hours at the door needs load readiness. Buying the wrong one is easy and, in this field, common.

The decomposition is free. Every endpoint is already an event the system writes; reporting the four intervals rather than their sum requires no new instrumentation at all.

4.2 How many trucks a yard has to hold

A site under pressure typically responds by acquiring parking. Little's Law says what that actually buys.

$$\text{trucks in the yard} = \text{arrival rate x dwell; trucks waiting} = \text{arrival rate x the yard-queue interval alone; door utilisation} = \text{arrival rate x mean service time / number of doors} \tag{little}$$

lambda is the arrival rate in vehicles per hour, D the mean dwell, s-bar the mean dock service time and m the number of usable doors. The relation holds in steady state whatever the arrival and service distributions.

Two readings follow, and the second is the one that changes decisions. Given a dwell target, the first relation converts it directly into a parking requirement — a site accepting twelve arrivals an hour at a ninety-minute dwell needs eighteen slots, and no amount of goodwill changes that. But the same relation says a yard enlarged without reducing dwell simply holds more trucks for the same time: the queue is accommodated, not shortened, and the detention clock runs identically.

Kingman's approximation supplies the reason the pressure appears when it does. Waiting time rises with utilisation and with variability, and it rises steeply as utilisation approaches one. A site at ninety-five per cent door utilisation is not slightly worse than one at eighty; it is in a different regime, and the difference between a manageable morning and a queue to the road is a few percentage points of rho.

Parking is a symptom-management purchase. It is sometimes the right one — a site must physically hold the trucks it has accepted — but it should be made knowing it moves no interval in Equation (dwell).

4.3 What a priority lane costs the trucks behind it

The gate captures a priority level on every vehicle, so the system can run a priority discipline. Cobham's non-preemptive result says what that discipline does to each class.

$$\textit{mean wait of class k = the residual service term divided by the product of one minus the cumulative load above it and one minus the cumulative load including it} \qquad \text{(priority)}$$

classes are numbered with 1 the highest priority, rho(j) is the load offered by class j, sigma(k) the cumulative load of classes 1 through k, and W(0) the mean residual service time. Service in progress is never interrupted, which is the correct model for a truck already at a door.

Three consequences are worth drawing out, because they are all counter-intuitive to somebody negotiating a carrier contract.

1. Priority creates no capacity. The work-conserving property means the load-weighted mean wait across all classes is unchanged by the discipline; priority only decides who bears it.
2. The cost falls hardest on the lowest class, and non-linearly. The denominator for the bottom class contains the total load, so as overall utilisation rises its wait diverges much faster than the mean does.
3. A priority class large enough to matter commercially is large enough to hurt. A promise extended to a carrier representing a third of the volume is not a courtesy — it is a re-design of the queue.

The useful discipline is therefore to price a priority lane before granting it. Every term in Equation (priority) is available from the gate's own history, so the delay a proposed lane would impose on ordinary traffic is a number that can be put in front of the person about to promise it.

4.4 Reservation lead time against effective capacity

A reserved slot is committed to a vehicle that has not arrived. It cannot be allocated to a truck that has. The system therefore trades certainty for capacity, and the exchange rate is the reservation lead time.

$$\textit{effective slot capacity = total slots minus the arrival rate times mean reservation lead time; the fraction of the yard actually available is one minus that ratio} \qquad \text{(reserve)}$$

L-bar is the mean interval between a slot being reserved and the vehicle occupying it. The subtracted term is itself Little's Law applied to reservations: a reservation is an object occupying a slot for L-bar with an arrival rate of lambda.

The design tension is genuine. Reserving early gives the gate a promise it can keep and gives the driver a position they can see; reserving late keeps the yard flexible. The equation says the cost of the promise is proportional to how far ahead it is made, which suggests the right policy is neither extreme but a lead time tuned against arrival-time reliability.

It also identifies the failure mode. A no-show holds its slot until the reservation expires, so the effective capacity loss from an unreliable carrier is its share of arrivals times its expiry timeout — which is why reservation policy and carrier punctuality analytics belong to the same system rather than to different teams.

> A yard reporting high availability while trucks queue at the gate is usually reserving too far ahead. The slots are not free, they are promised — and the queue outside is paying for the certainty inside.

5 Reported Outcomes and Field Evidence

Three kinds of claim appear in the product's material and they carry different weight. The captures are separated out first, because they are the weakest and the most likely to be mistaken for the strongest.

5.1 What the captures do and do not establish

The screens reproduced in Section 3 come from a demonstration instance. The queue view reads that no trucks are in the system; the gate statistics read zero on five of six counters; the traffic chart is labelled simulated; the carrier names are placeholders.

They establish that the software models a priority level at the gate, three identity capture routes, a three-state slot pool, manual and automatic assignment, and per-carrier turnaround reporting. They establish nothing about how long a real truck waits in a real yard, and nothing in this paper treats a number from them as if they did.

> This is said plainly because the alternative is easy and dishonest. A demonstration build's efficiency score is a design placeholder, and reprinting it as a result would be inventing evidence.

5.2 Figures published for this module

Table 2. Published figures for this module, with their provenance.

Figure	Value	Provenance
Manual gate processing removed	3-5 min per vehicle	Product FAQs — the step that automatic plate recognition eliminates
Inbound calls to the gate	60-80% reduction	Product FAQs — vendor-reported, no baseline volume stated
Turnaround reduction	40-60%	Product material — a range without a baseline or a site population

The first is the most defensible of the three, because it names a specific manual step and removes it. Section 4.1 places it precisely: it acts on the gate interval and on nothing else, so a site whose problem is in the yard queue should expect three to five minutes rather than a transformation.

The second has a mechanism worth taking seriously even though the figure carries no baseline. Section 3.3 argues that inbound calls consume the throughput of the resource doing the sequencing, so their removal has an effect beyond driver satisfaction. That is an argument, not a measurement, and it is offered as one.

5.3 A deployment shared with the dock module

A multi-client third-party logistics operator deployed six MileSoft warehouse modules across a distribution centre whose dock and yard scheduling previously ran on messaging apps and paper trip sheets. Carriers arrived early or late; docks sat idle while drivers waited in the gate queue.

The deployment reported average gate-to-gate dwell falling from three hours ten minutes to one hour twenty, and dock idle time falling by 38%.

> Shared attribution. Both figures are attributed in their own source to this module and Loading Dock Management operating on a single shared yard state — dock assignments and parking allocations coordinated rather than separately optimised. Neither figure is this module's alone, and the published explanation of the dock-idle result turns on re-sequencing the queue when a truck arrives early, which requires both halves.

What can be said specifically is that the mechanism named in the published account is the one Section 4.2 formalises. Re-sequencing on early arrival is a queue-discipline change; it converts an early truck from a vehicle held against a static appointment into one that fills a door that would otherwise idle. That is a reduction in the yard-queue interval and in door idle simultaneously, which is why the two results move together.

Two caveats belong with the numbers. The baseline is a site coordinating by messaging app and paper, so the improvement measures the distance from an unusually poor starting point; and no distribution is published, so a mean falling from 190 to 80 minutes says nothing about the tail, which is where detention charges are incurred.

5.4 Modelled detention and handling saving

The published return model for this module prices the saving as detention and handling time. Its assumptions are printed here so a reader can substitute their own.

Table 3. Modelled scenario — not a deployment result. Assumptions are the published defaults of the return-on-investment model for this product.

Assumption	Value
Trucks per day	80
Operating days per year	300
Time saved per truck	45 minutes
Basis for that saving	a stated 40-60% turnaround reduction
Modelled annual saving	18,000 vehicle-hours

The 45 minutes is the assumption to interrogate, and Section 4.1 says how. A saving of that size cannot come from the gate interval, which the product itself bounds at three to five minutes. It has to come from the yard queue, from dock service, or from both — and only the first belongs to this module.

> A site should decompose its own dwell before accepting this model. If two of its three hours are at the door, a yard system will not deliver 45 minutes and the honest forecast is much smaller.

6 Discussion

6.1 Coordination failure has a precise meaning

The product's material calls the turnaround problem a coordination failure rather than a capacity failure. The equations in Section 4 let that be said exactly rather than rhetorically.

A capacity failure is a site where door utilisation rho is at or above one: more work arrives than can be served, and no discipline helps. A coordination failure is a site where rho is comfortably below one and the yard-queue interval is large anyway — which happens when arrivals are variable, when appointments are static, or when the queue is sequenced by a person who cannot see the doors.

The distinction is diagnosable in an afternoon. Compute rho from arrivals and service times; decompose dwell by Equation (dwell). A site with rho at 0.7 and two hours in the yard queue has a discipline problem, and buying doors will not fix it.

> Kingman's approximation is what makes the diagnosis subtle. A site at rho = 0.9 with high arrival variability queues like a site at rho = 0.98 with none — so utilisation alone does not settle it, and variability has to be measured too.

6.2 Attribution is what makes dwell a usable metric

The product's material makes attribution a first-class concern — whether a wait was the warehouse's fault, the carrier's, or load readiness. That instinct is right and Section 4.1 explains why in structural terms.

Dwell is billed. A carrier invoices detention beyond an agreed threshold, and the argument that follows is about which side caused the overrun. Without the four intervals, that argument is settled by relative bargaining power rather than by evidence, and the party with the better records wins irrespective of the facts.

This gives the decomposition a commercial value distinct from its operational one. A site that can show a carrier that ninety minutes of a two-hour dwell was its own late arrival has converted a recurring dispute into a conversation about arrival windows — which is where the underlying variability, and therefore the queue, actually lives.

6.3 A capability reference framework for yard and turnaround systems

Table 4. Capability reference framework. Each dimension separates a turnaround system from a digital gate log, and each is answerable by demonstration on a live gate rather than by reading a specification.

Dimension	Question the system must answer by demonstration
D1 Four intervals	Report last month's dwell split into gate, yard queue, dock service and exit. Can it?
D2 Generated endpoints	Is any of arrival, queue entry, dock call or departure typed by a person?
D3 Attribution	For one overrun, can the system say whose delay it was without a manual investigation?
D4 Priority priced	Before granting a carrier a priority lane, can the system state the delay it adds to everyone else?
D5 Re-sequencing	A truck arrives ninety minutes early to an idle door. Does the queue re-sequence, or does the appointment hold?
D6 Reservation lead time	How far ahead is a slot committed, and what fraction of the yard is held in reservation right now?
D7 Distribution not mean	Does dwell reporting show a percentile, or only an average?
D8 Shared yard state	Do the gate and the dock read and write the same yard state, or exchange it?

D5 is the dimension the published deployment result actually turns on. A system that holds an early truck against a static appointment while a door idles is not coordinating; it is scheduling, and the two are different jobs.

6.4 Generalisability

The interval decomposition generalises to any served-arrival process with a waiting area and a commercial clock: port terminals, vehicle workshops, hospital admissions, contact centres. Cobham's result generalises to every priority scheme anywhere, and the observation that priority creates no capacity is the one most often forgotten in all of them.

What does not generalise is the deployment figure. Three hours ten to one hour twenty is measured against a site coordinating by messaging app, and the size of the improvement is a property of that baseline rather than of the software. A site already running a digital yard should expect a far smaller change, and should size its expectation from its own interval decomposition.

7 Threats to Validity and Limitations

1. The captures are a demonstration build. Counters read zero, carrier names are placeholders and the traffic chart is labelled simulated; nothing about real yard performance can be read from them, and this

paper reads nothing from them.

2. The dwell result is shared and unbounded above. Three hours ten to one hour twenty is attributed jointly with the dock module, and no distribution is published — so nothing is known about the tail, which is where detention charges are actually incurred.
3. The baseline was unusually poor. The prior state was messaging apps and paper trip sheets, so the improvement measures a distance from that rather than the value of the software against a competent digital alternative.
4. The 40-60% turnaround range has no population. No number of sites, no arrival rates and no definition of what turnaround included are published alongside it.
5. The 60-80% call reduction has no baseline. A percentage reduction in calls with no starting volume cannot be converted into any operational quantity.
6. Little's Law assumes steady state. A yard's arrival rate is strongly non-stationary — the six-to-ten peak is the whole problem — so Equation (little) applies within a period of stable arrivals, not across a day.
7. Cobham's result assumes a single-server model and known distributions. A multi-door yard with heterogeneous doors departs from those assumptions, so Equation (priority) gives the direction and order of magnitude of a priority cost rather than an exact figure.
8. The reservation model treats lead time as exogenous. In practice it interacts with carrier punctuality, and a site with unreliable arrivals will find that shortening lead time increases no-show exposure rather than freeing capacity cleanly.

> The sixth limitation is the one that most affects practice. Every result here is a steady-state result and a yard's defining feature is a morning peak — so the arithmetic should be applied within the peak, using the peak's arrival rate, not the day's average.

8 Future Work

- Publishing the four intervals as standard. Every endpoint is already captured, so reporting the decomposition of Equation (dwell) rather than its sum is a reporting change rather than an engineering one, and it is the single most useful thing this system could add.
- Percentile dwell rather than mean. Detention is charged on the tail, so a ninetieth-percentile figure describes the commercial exposure that a mean conceals.
- Priority cost as a pre-commitment calculation. Every term in Equation (priority) comes from the gate's own history; surfacing the imposed delay before a lane is granted would put the trade in front of the person making it.
- Reservation lead time tuned against measured punctuality. The exchange rate in Equation (reserve) is known; the missing input is each carrier's arrival-time distribution, which the system already has the events to compute.

- Non-stationary analysis over the peak. Applying the steady-state results within the morning peak, using peak arrival rates, would make the sizing advice in Section 4.2 usable rather than merely correct.

9 Conclusion

Fifteen trucks outside and two empty doors inside is not a shortage of doors. It is a queue without a discipline, and the arithmetic that fixes it has been available since the 1950s.

Three results carry the argument. Dwell is four intervals with four owners, and reporting their sum makes it impossible for a site to tell whether it should buy identity capture, sequencing, or load readiness — a decomposition that costs nothing, because every endpoint is already an event. Yard occupancy is arrival rate times dwell, which converts a service promise into a parking requirement and shows that enlarging a yard accommodates a queue rather than shortening it. And a priority lane creates no capacity: it moves waiting from one carrier onto everyone behind, by an amount computable before the promise is made.

The reservation result is the quietest and the most practical. A slot committed early is capacity withheld, so a yard reporting healthy availability while trucks queue at the gate is usually promising too far ahead — and the queue outside is paying for the certainty inside.

The framework of Section 6.3 is offered as the durable contribution, and its fifth dimension is the one to test first: send a truck ninety minutes early to an idle door, and see whether the queue moves.

Appendix A Nomenclature

Table 5. Symbols and abbreviations used in this paper.

Symbol / term	Meaning
D	Gate-to-gate dwell — the interval a carrier bills against
t(arr), t(q)	Arrival at the gate, and entry to the queue once identity is resolved
t(dock), t(rel)	Call to a door, and release from it
t(dep)	Departure through the gate
lambda	Arrival rate, in vehicles per hour
N	Vehicles in the yard, by Little's Law
s-bar	Mean dock service time
m, m(eff)	Total slots or doors, and the effective number after reservations
rho	Utilisation — offered load divided by capacity
W(k)	Mean wait of priority class k
W(0)	Mean residual service time — the delay from work already in progress
sigma(k)	Cumulative load of priority classes 1 through k
L-bar	Mean reservation lead time — reservation to occupancy
eta	Fraction of the yard actually available after reservations
ANPR	Automatic number plate recognition
TAT	Turnaround time, used interchangeably with dwell in commercial contexts

Appendix B Worked Numerical Examples

Appendix B.1 Two sites, one dwell figure, opposite purchases

Two distribution centres each report a mean dwell of 190 minutes. Site A decomposes as 9 minutes at the gate, 128 in the yard queue, 46 at the door and 7 on exit. Site B decomposes as 11, 24, 149 and 6.

Applying Equation (dwell) gives those pairs directly; their sums are identical, which is the only figure a conventional report would show.

The purchases are opposite. Site A holds two thirds of its dwell in the yard queue: it needs sequencing, priority design and reservation policy, and automatic plate recognition would recover at most five of its 190 minutes. Site B holds four fifths at the door: it needs load readiness and door capability, and a yard system would leave 149 of its minutes untouched.

> Site B buying a turnaround system on the strength of a 40-60% published range would recover perhaps twelve minutes and conclude the category does not work. Its own data said so beforehand, for free.

Appendix B.2 What a dwell target costs in tarmac

A site accepts 12 arrivals an hour through a morning peak and wants to hold mean dwell at 90 minutes. It has 6 doors and mean dock service is 25 minutes.

Applying Equation (little): N = 12 x 1.5 = 18 vehicles in the yard at any moment, so 18 slots are the minimum the target implies. Door utilisation is rho = 12 x (25/60) / 6 = 0.83.

Now suppose the site instead buys 12 more slots and changes nothing else. Arrival rate and service are unchanged, so dwell is unchanged, and by the same equation the yard simply holds 18 trucks in a 44-slot yard rather than 18 in a 32-slot one. The queue is more comfortable and exactly as long.

Finally, suppose demand rises to 14 arrivals an hour. Utilisation goes to 0.97, and by Kingman's approximation the waiting term scales roughly as rho / (1 — rho): from 0.83/0.17 = 4.9 to 0.97/0.03 = 32.3, a factor of about 6.6 on the queueing component. A 17% rise in arrivals produced a several-fold rise in waiting — which is why the morning peak feels like a cliff rather than a slope.

Appendix B.3 Pricing a priority lane before promising it

A yard runs at total utilisation 0.80 across a single ordinary class, with a mean residual service term W(0) of 8 minutes. Applying Equation (priority) with one class: W = 8 / (1 x 0.20) = 40 minutes of mean wait.

A carrier representing 30% of the volume now asks for a priority lane. Splitting the load into a priority class at rho(1) = 0.24 and an ordinary class taking the remaining 0.56: the priority class waits 8 / (1 x 0.76) = 10.5 minutes, and the ordinary class waits 8 / (0.76 x 0.20) = 52.6 minutes.

The priority carrier gains 29.5 minutes. Everyone else loses 12.6. Weighted by volume the total is unchanged, as the work-conserving property requires — 0.30 x 10.5 + 0.70 x 52.6 = 40.0 — which is the arithmetic proof that no capacity was created.

> Twelve and a half minutes added to seventy per cent of arrivals is a real cost with real detention exposure attached. It is also a number that could have been put in front of the person signing the contract, in about two minutes.

Appendix B.4 The capacity a promise consumes

A 32-slot yard reserves a slot at appointment confirmation, on average 55 minutes before the vehicle arrives, against 12 arrivals an hour.

Applying Equation (reserve): the reservation term is 12 x (55/60) = 11 slots held at any moment. Effective capacity is 32 — 11 = 21 slots, so eta = 0.66 — a third of the yard is committed to trucks that are not there.

Shortening the lead time to 20 minutes reduces the held slots to 4 and raises effective capacity to 28, a 33% increase in usable yard with no construction. The price is that the promise is made later, so a driver learns their slot closer to arrival and a no-show is discovered later in the sequence.

The trade should be made per carrier rather than globally, because it depends on arrival reliability. A punctual carrier can be reserved late with little risk; an unreliable one either needs a long lead time or should not hold a reservation at all — and the arrival history to make that distinction is already in the system.

Data provenance

Every quantitative claim in this paper resolves to one of the entries below. Figures marked as modelled estimates are not deployment measurements; their assumptions are stated.

Metric	Reported value	Provenance
Average truck dwell time	1 h 20 min (from 3 h 10 min)	**Operator-reported — Tier 1 Manufacturers (3PL, Madhya Pradesh, India)** /case-studies/warehouse — Gate-to-gate, measured across all carriers at the distribution centre.
Dock idle time	38% reduction	**Operator-reported — Tier 1 Manufacturers (3PL, Madhya Pradesh, India)** /case-studies/warehouse
Manual gate processing removed per vehicle by plate recognition	3-5 minutes	**Product specification** /products/truck-turnaround
Inbound phone calls to the gate avoided by a carrier-facing portal	60-80% reduction	**Product specification** /products/truck-turnaround — Vendor-reported on the product FAQs; no baseline call volume or sample is stated.
Vehicle identity capture routes offered at the gate	keyed plate, RFID tag, automatic plate recognition	**Product specification** /products/truck-turnaround — Read from the gate entry screen captured on the product page.
Yard slot states tracked in the parking layout	available, occupied, reserved	**Product specification** /products/truck-turnaround — Read from the parking screen captured on the product page; the capture is a demonstration build, so its slot counts are not operational data.
Truck turnaround time reduction	40-60%	**Product specification** /products/truck-turnaround

Metric	Reported value	Provenance
Detention and handling time saved per truck	45 min across 80 trucks/day, 300 days (modelled)	Modelled estimate 80 trucks per day across 300 operating days; 45 minutes saved per truck, from a stated 40-60% turnaround reduction; Model and defaults published in src/data/roiModels.ts

References

[1] Little, J. D. C. (1961). A proof for the queuing formula: L = lambda W. Operations Research, 9(3), 383-387.

[2] D. G. Kendall (1953). Stochastic Processes Occurring in the Theory of Queues and their Analysis by the Method of the Imbedded Markov Chain. The Annals of Mathematical Statistics, vol. 24, no. 3, pp. 338-354; the source of the A/B/c notation used to classify queueing systems. https://doi.org/10.1214/aoms/1177728975

[3] A. Cobham (1954). Priority Assignment in Waiting Line Problems. Journal of the Operations Research Society of America, vol. 2, no. 1, pp. 70-76; gives the waiting time of each class in a non-preemptive priority queue. https://doi.org/10.1287/opre.2.1.70

[4] J. F. C. Kingman (1961). The Single Server Queue in Heavy Traffic. Mathematical Proceedings of the Cambridge Philosophical Society, vol. 57, no. 4, pp. 902-904; the first statement of the approximation relating waiting time to utilisation and variability. https://doi.org/10.1017/S0305004100036094

[5] de Koster, R., Le-Duc, T., and Roodbergen, K. J. (2007). Design and control of warehouse order picking: A literature review. European Journal of Operational Research, 182(2), 481-501.

[6] Gu, J., Goetschalckx, M., and McGinnis, L. F. (2007). Research on warehouse operation: A comprehensive review. European Journal of Operational Research, 177(1), 1-21.

[7] International Organization for Standardization / International Electrotechnical Commission (2024). ISO/IEC 19987: Information technology - EPC Information Services (EPCIS) Standard, version 2.0. ISO/IEC, Geneva. https://www.iso.org/standard/85557.html

[8] International Organization for Standardization / International Electrotechnical Commission (2024). ISO/IEC 19988:2024 - Information technology - GS1 Core Business Vocabulary (CBV). ISO/IEC, Geneva, edition 3, March 2024; publishes GS1 CBV 2.0, ratified June 2022, which fixes the vocabularies for the EPCIS business step, disposition and business transaction type fields. https://www.iso.org/standard/85558.html

[9] GS1 (2024). Serial Shipping Container Code (SSCC). GS1 identification key series; an 18-digit key for a logistic unit, carried under application identifier 00. https://www.gs1.org/standards/id-keys/sscc

[10] GS1 (2024). GS1 General Specifications. GS1 AISBL; defines the GTIN, SSCC and GLN identification keys. https://ref.gs1.org/standards/genspecs/

[11] International Organization for Standardization / International Electrotechnical Commission (2014). ISO/IEC 15459-1: Information technology - Automatic identification and data capture techniques - Unique identification - Part 1: Individual transport units. ISO/IEC, Geneva. https://www.iso.org/standard/54779.html

[12] International Organization for Standardization (2022). ISO 28000: Security and resilience - Security management systems - Requirements. ISO, Geneva; published 15 March 2022. https://www.iso.org/standard/79612.html

[13] International Organization for Standardization (2014). ISO 22400-1: Automation systems and integration - Key performance indicators (KPIs) for manufacturing operations management - Part 1: Overview, concepts and terminology. ISO, Geneva. https://www.iso.org/standard/56847.html

[14] International Organization for Standardization (2014). ISO 22400-2: Automation systems and integration - Key performance indicators (KPIs) for manufacturing operations management - Part 2: Definitions and descriptions. ISO, Geneva; defines OEE as availability x effectiveness x quality ratio. https://www.iso.org/standard/54497.html

[15] International Electrotechnical Commission / International Society of Automation (2025). Enterprise-control system integration - Part 1: Models and terminology. IEC 62264-1; ANSI/ISA-95.00.01-2025 (IEC 62264-1 Mod).

[16] International Society of Automation (2013). Enterprise-control system integration - Part 3: Activity models of manufacturing operations management. ANSI/ISA-95.00.03-2013 (IEC 62264-3 Modified).

[17] International Chamber of Commerce (2019). Incoterms 2020: ICC Rules for the Use of Domestic and International Trade Terms. ICC Publication No. 723E, Paris; launched September 2019 and in force from 1 January 2020, defining the eleven trade terms that allocate cost and risk between seller and buyer.

[18] European Committee for Standardization (2009). EN 1398: Dock levellers - Safety requirements. CEN, Brussels.

[19] Clarke, G., and Wright, J. W. (1964). Scheduling of vehicles from a central depot to a number of delivery points. Operations Research, 12(4), 568-581.

[20] Kagermann, H., Wahlster, W., and Helbig, J. (2013). Securing the future of German manufacturing industry: Recommendations for implementing the strategic initiative INDUSTRIE 4.0. Final report of the Industrie 4.0 Working Group, acatech - National Academy of Science and Engineering. https://www.acatech.de/wp-content/uploads/2018/03/Final_report__Industrie_4.0_accessible.pdf