UDK 614.842.8:004

IMPROVING THE OPERATIONAL MANAGEMENT OF FIRE AND RESCUE UNITS THROUGH DIGITAL TECHNOLOGIES

Ermirzaev Ozodbek Mahmudjon ugli

(Independent Researcher, Center for Professional Development of Employees of the Ministry of Poverty Reduction and Employment)

Abstract. Fire and rescue operations require the coordination of heterogeneous resources under changing incident conditions and incomplete information. This study develops an auditable decision support framework linking validated observations, capability-aware allocation, commander review, and operational feedback. A constrained formulation combines priority-weighted delay, capability shortfall, and coverage loss; its timing component is evaluated by exhaustive assignment enumeration. A deterministic example involving two incidents and four units reduces weighted mean operational package-ready time from 11.07 to 9.57 min, while delaying the higher-priority incident by one minute. A computational experiment then evaluates the same 10,000 synthetic travel-time scenarios at four forecast-error levels. With exact travel information, joint allocation reduces the mean paired timing endpoint by 76.02 s, with a 95% Monte Carlo interval of 74.75–77.28 s. At a forecast-error coefficient of variation of 0.40, the mean benefit declines to 50.81 s, and 5.36% of scenarios have a worse aggregate outcome than the greedy comparator. A forecast-based protection constraint reduces incident-level deterioration but does not guarantee realised protection when forecasts are inaccurate. These findings establish the behaviour and limitations of the allocation mechanism under explicit synthetic assumptions. They do not demonstrate field effectiveness or casualty reduction. The contribution is a transparent computational method for jointly evaluating timing gains, incident-level trade-offs, and information quality before operational trials.

Keywords: fire and rescue services; operational management; decision support; capability-aware allocation; incident command; information latency; digital technologies

Introduction. Operational management in the fire and rescue service is a coordination problem in which time, resource capability, and incident development interact. A vehicle may arrive rapidly but lack the equipment or personnel required for a critical task. Conversely, an apparently efficient allocation to one incident may remove the only resource capable of responding promptly to another. Digital technologies become operationally meaningful when their information changes a feasible decision and the consequences of that decision are evaluated through clearly defined outcomes.

Field experiments reported by Averill et al. [1] investigated crew size, apparatus arrival, and timed fireground tasks. Their work supports assessing deployment through task readiness as well as arrival time. It does not establish a transferable effect size for digital dispatch systems. Accordingly, the present study uses operational package readiness as its analytical endpoint and reserves claims about suppression, rescue success, and human outcomes for subsequent empirical evaluation.

The scientific question is whether capability-aware assignment can improve aggregate operational readiness while making incident-level trade-offs and forecast uncertainty explicit. The study evaluates a defined allocation mechanism rather than an operationally deployed system. Its evidence consists of a deterministic counterexample and a reproducible computational experiment. All travel-time inputs are synthetic; no incident registry, expert panel, or field trial is represented as having participated.

The objectives are to formalise the information-to-decision process, specify a capability-aware allocation model, evaluate it against an explicit greedy comparator under travel-time uncertainty, and identify when additional decision latency offsets its timing benefit. The contribution is the integrated treatment of package readiness, forecast-based protection, and realised outcome evaluation. Digital integration and mathematical resource recommendation are established research directions [2–5]; the present study provides a bounded computational analysis of their timing and uncertainty implications.

Related work and conceptual foundation. The NIST smart firefighting roadmap [2] treats sensor, computing, and building-system integration as a basis for supporting fire-service operations. Holmberg et al. [3] examine the delivery of building information to responders and the associated communication and information-organisation challenges. Together, these sources motivate an architecture in which observation, processing, and operational delivery are distinct responsibilities. A dashboard alone does not determine which unit should be deployed or how uncertain information should influence that decision.

Communication research adds a human-factors dimension. Choong et al. [4] identify first-responder communication needs through user-centred interviews, providing a basis for considering operational context, trust, and interface requirements. In the proposed framework, a recommendation therefore includes the observations used, their age, binding constraints, and the alternative affected by the allocation. Commander review is part of the decision process and contributes to total latency.

Mercadé et al. [5] present IRIS-AI, which predicts required resource characteristics and combines recommendations with a mixed-integer allocation component. This is direct evidence that intelligent resource recommendation for fire departments is an established research direction. The present model does not reproduce its predictive architecture. Instead, demand is an explicit input, and the analysis concentrates on multi-incident timing, service constraints, and the additional latency that can offset allocation benefits.

Recent developments extend these foundations. GUIDE2FR integrates building monitoring and digital-twin information for emergency scenarios [6], while Yang et al. [7] combine firefighting resource scheduling with a virtual forest environment. Their operational settings and endpoints differ from those used here, preventing direct numerical comparison. Research on communication technology adoption also identifies trust and responder autonomy as relevant design considerations [8]. The resulting research gap is specific: the timing value of a joint assignment should be evaluated together with forecast error, individual-incident deterioration, and incremental information-to-action delay. This targeted synthesis does not establish that no previous system addresses these issues.

Table 1.

Research foundations and their role in the proposed framework

Source	Main contribution	Use in this study
[1] Fireground experiments	Deployment and timed operational tasks	Distinguish arrival from task readiness
[2] Smart firefighting roadmap	Cyber-physical information integration	Separate acquisition, analysis, and delivery
[3] Building intelligence	Building data for responders	Represent provenance and data exchange
[4] Responder communication study	Operational needs and trust	Include understandable commander review
[5] IRIS-AI	Resource characteristics and allocation	Position capability-aware recommendations
[6] GUIDE2FR	Building monitoring and digital twin	Separate monitoring from allocation evidence
[7] Forest firefighting simulation	Scheduling and virtual operational scenarios	Locate the model in simulation research
[8] Communication adoption	Trust and responder autonomy	Retain review and accountable overrides

Materials and methods. The study combines conceptual design, mathematical formulation, deterministic assignment enumeration, and Monte Carlo evaluation with controlled forecast error. The analytical unit is one allocation decision for two simultaneous incidents. The computational evaluation tests the timing mechanism across 10,000 unique synthetic travel matrices, reused across four information conditions. The system architecture is proposed for future implementation; the study is not a field trial, calibrated fire-spread simulation, or deployed software evaluation. Figure 1 specifies the information flow and human decision boundary.

Each observation is represented by an incident or resource identifier, a measured or reported value, generation time, receipt time, source, and quality status. Generation time permits assessment

of freshness; receipt time identifies network delay. Uncertain availability is kept distinct from confirmed unavailability. A stale vehicle position is not silently accepted as current, and a missing hazard report is not interpreted as absence of hazard.

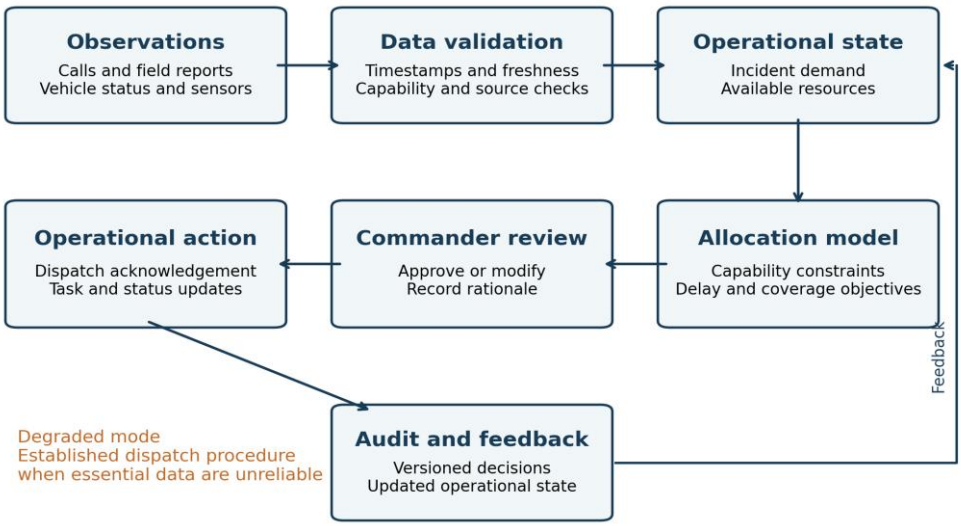


Figure 1. Proposed information-to-action architecture. The feedback loop updates operational state; essential data failures trigger the established dispatch procedure. The architecture is a design proposal, not a tested deployment.

Building sensors, mobile reports, vehicle location services, and geographic information systems are optional input channels. Their inclusion depends on actual availability and interoperability. A spatial display is not termed a digital twin unless it provides a documented, synchronised representation with a defined update and validation process. The framework can operate with dispatch and resource-status records before advanced sensing is introduced.

Operational timing and priority representation. For a predefined operational action, time from call receipt to intervention comprises mutually exclusive intervals: call processing, turnout, travel, and on-scene setup. When several capabilities are required before that action may begin, readiness depends on the last required resource rather than the earliest arriving vehicle. In the analytical example, all selected units share processing and turnout times and the package has one common setup interval.

$$T_i = p_i + q_i + \max_{j \in S_i} \tau_{ij} + s_i \quad (1)$$

Here T_i is package-ready time in minutes, p_i is call-processing time, q_i is common turnout time, τ_{ij} is travel time of unit j to incident i, s_i is setup time, and S_i is the selected operational package. Equation (1) is restricted to the stated common-timing assumptions. With staggered dispatch, readiness must instead be derived from each unit's absolute arrival and setup timestamps. Some tasks can begin before the complete package arrives; the endpoint therefore concerns the defined joint action, not every possible fireground activity.

$$P_i(t) = w_L L_i(t) + w_H H_i(t) + w_V V_i(t) + w_E E_i(t) \quad (2)$$

The normalised indicators represent life threat, hazard severity, occupant vulnerability, and escalation potential. Non-negative weights sum to one. P_i is a priority index, not a probability of injury. Operational use requires documented indicator definitions and expert-supported weight selection. The numerical example specifies P directly to avoid implying that the proposed weights or indicators have been empirically validated. Missing information is flagged for review rather than assigned an automatically low priority.

Capability-aware allocation formulation. Let x_{ij} equal one when resource j is assigned to incident i and zero otherwise. Let a_j indicate confirmed availability and c_{jk} denote usable capacity for capability k. Capabilities may represent trained personnel, pumping capacity, aerial access, or specialist equipment, but their units must be defined separately. Resources already committed to non-interruptible tasks are excluded from reassignment.

$$\sum_i x_{ij} \le a_j, \quad x_{ij} \in \{0, 1\} \quad (3)$$

$$\sum_j c_{jk} x_{ij} + u_{ik} \geq d_{ik}, \quad u_{ik} \geq 0 \tag{4}$$

Demand d_{ik} is the required capability and u_{ik} records shortfall in the same unit. Mandatory capabilities use $u_{ik} = 0$. An infeasible mandatory requirement must trigger escalation or mutual-aid review; the optimiser must not label a resource-deficient assignment operationally adequate. For non-mandatory demand, shortfall may remain in the objective, but its penalty must be accompanied by a visible operational warning.

$$\widetilde{D} = \frac{\sum_i P_i T_i}{T_{\text{ref}} \sum_i P_i} \tag{5}$$

$$\widetilde{U} = \sum_{i,k} \frac{v_{ik} u_{ik}}{d_{\text{ref},k}}, \quad \widetilde{C} = \frac{\sum_g b_g r_g}{\sum_g b_g} \tag{6}$$

Tref is a fixed positive reference time; $d_{ref,k}$ is a positive reference capacity for each capability, and the non-negative v_{ik} weights are fixed before comparison. For coverage loss, rg identifies a zone that was adequately covered before allocation but loses the predefined capability-specific coverage level afterwards. bg weights the zone by a fixed demand or population measure. Initially uncovered zones are recorded separately. Reference quantities and thresholds remain identical across candidate methods. The coverage component is specified for future evaluation and is not activated in the numerical experiments.

$$\min_x J = \alpha \widetilde{D} + \beta \widetilde{U} + \gamma \widetilde{C}, \quad \alpha, \beta, \gamma \geq 0 \tag{7}$$

The normalisation prevents adding minutes directly to personnel counts or population totals. At least one objective weight is positive. Because a weighted sum can accept deterioration at an individual incident, protected incidents may also receive a bound $T_i \leq B_i$. For implementation, introduce a package travel variable z_i and require $z_i \geq \tau_{ij} - M_{ij}(1 - x_{ij})$, with valid upper bounds M_{ij} and $z_i \geq 0$; use $T_i = p_i + q_i + z_i + s_i$. Positive timing weights make z_i equal the largest selected travel time at optimum. Mandatory capability constraints prevent empty packages. Enumeration evaluates the maximum directly in the present small problem. Uncertain forecasts require a separate interpretation of bounds: satisfying a predicted bound does not establish a guarantee about realised travel times.

Both numerical evaluations examine the timing subproblem only: $\alpha = 1$, $\beta = \gamma = 0$, and all mandatory demand is satisfied. They therefore test package timing and protection behaviour without claiming evidence about coverage preservation or shortage management. Operational adoption would require these omitted components to be evaluated with actual attendance and geographical data.

Decision procedure and degraded operation. At each material update, the procedure validates input records, freezes existing operational commitments, updates incident demand, generates feasible packages, evaluates the objective and protected bounds, and submits the preferred plan for review. A resource command is effective only after acknowledgement. Unacknowledged commands remain explicitly unresolved. Updates can be triggered by a new incident, confirmed unit availability, a changed access route, or a revised capability requirement.

Repeated reassignment is controlled by retaining committed packages and requiring a minimum documented benefit before redirecting uncommitted units. The intended implementation should log the input version, candidate plans, selected objective value, constraint status, reviewer decision, and dispatch acknowledgement. If required data fail freshness or consistency checks, the system returns control to the established dispatch process instead of generating an unsupported precision estimate.

Fully specified analytical scenario. The constructed scenario contains incidents A and B with priority indices 0.9 and 0.5. A requires one engine and one aerial unit for its predefined operation; B requires one engine. E1 and E2 provide the engine capability, while L1 and L2 provide aerial access. Each unit can serve only one incident, all units are initially available, and all personnel and equipment necessary for their stated capability are assumed present. These binary capabilities simplify actual attendance requirements and are not recommended staffing levels.

Table 2. Constructed travel-time matrix and resource capabilities

Unit	Capability	Travel to A min	Travel to B min
E1	Engine	4	3
E2	Engine	7	9
L1	Aerial	6	8
L2	Aerial	9	4

Call processing is 1 min, turnout 1 min, and common setup 2 min for both incidents and methods. Thus, package-ready time equals travel readiness plus 4 min. Only minimal packages are considered: one engine at each incident and one aerial at A, leaving the other aerial in reserve. Extra non-required deployments are excluded. There are exactly four eligible assignments, permitting complete enumeration without a heuristic search or solver optimality assumption.

The comparator is a priority-first nearest-capable greedy rule: process A before B, assign the nearest available resource for each required capability, and then assign remaining compatible resources to B. This gives E1 and L1 to A and E2 to B. It is a precisely defined analytical comparator; it is not described as the current practice of any fire service. The joint method evaluates all four eligible plans using the same input information.

Analytical sensitivity and scope of inference. The deterministic analysis varies the incident-priority ratio and additional decision latency. The computational analysis holds priorities and capability demand fixed and varies uncertainty in travel forecasts. Incremental latency is imposed on the joint method relative to the comparator; common latency cancels from their absolute difference. Statistical intervals in the computational evaluation quantify Monte Carlo sampling precision under the chosen generator and are not confidence intervals for an actual fire-service population.

Synthetic scenario generation and forecast uncertainty. The nominal travel matrix in Table 2 defines arithmetic means for a constructed probability model. Each realised travel time is independently lognormal with coefficient of variation 0.25. This positive distribution introduces variability without allowing negative travel times. Independence between routes is a deliberate simplifying assumption; the generator does not represent calibrated traffic conditions or shared congestion.

$$\tau_{ij} = m_{ij}\exp(\sigma Z_{ij} - \sigma^2/2), \quad \sigma = \sqrt{\ln(1+c^2)} \qquad (10)$$

Here m_{ij} is the nominal mean, c = 0.25, and Z_{ij} is standard normal. Generate 10,000 matrices in row order E1, E2, L1, L2 and column order A, B. Forecasts equal realised travel times in the exact-information condition. For error coefficients of variation e = 0.10, 0.25, and 0.40, multiply each realised time by an independent mean-one lognormal factor. Thus forecast error is conditionally unbiased in the arithmetic mean, but individual forecasts can overestimate or underestimate the realised value.

$$\hat{\tau}_{ij} = \tau_{ij}\exp(s_e Y_{ij} - s_e^2/2), \quad s_e = \sqrt{\ln(1+e^2)} \qquad (11)$$

All 10,000 realised matrices are reused across information conditions; only the forecasts change. There are 40,000 condition evaluations, not 40,000 independent realised scenarios. NumPy 2.3.5 and default_rng with seed 20261014 are used. First draw the complete realised array of shape (10000, 4, 2); then draw error arrays of the same shape sequentially for e = 0.10, 0.25, and 0.40. No error draws occur at e = 0. This draw order and the supplied parameters determine the numerical replication.

Comparators and realised performance measurement. All three methods receive the same forecast matrix. The greedy method assigns A the engine with the shortest predicted travel time and the aerial unit with the shortest predicted travel time, then assigns the remaining engine to B. Joint allocation minimises the predicted weighted mean package-ready time over the four minimal packages. The protected variant adds a predicted readiness bound at A equal to the predicted greedy readiness at A for that scenario. The greedy package remains feasible under this bound, ensuring that the protected candidate set is nonempty.

Ties are resolved by the candidate order in Table 3; comparisons use a numerical tolerance of 10^{-9} min. Each selected package is subsequently evaluated using the realised travel matrix. The primary computational endpoint is the paired difference in realised weighted mean package-ready time, expressed in seconds; positive values favour the joint method. Secondary outcomes include the

proportion improving or worsening the aggregate endpoint, the proportion delaying A, and the protected variant's mean benefit and A-delay proportion. These measures concern defined package readiness, not task completion or casualties.

$$g_r = 60\frac{\sum_i P_i(T_{ir}^G - T_{ir}^J)}{\sum_i P_i} \quad (12)$$

Monte Carlo precision and verification. For each condition, the interval for the mean paired benefit is the sample mean plus or minus 1.96 times the sample standard deviation divided by the square root of 10,000. This is an approximate 95% Monte Carlo interval conditional on the synthetic model. It excludes uncertainty in the chosen mean matrix, distribution, and capability assumptions. No field-effect hypothesis test is performed. Scenario dependence across conditions prohibits treating their combined records as independent observations**.**

Implementation checks verify the four nominal objective values, exclusivity of engine assignments, satisfaction of mandatory capabilities, and feasibility of the protected forecast constraint. With exact information, joint and protected timing benefits cannot be negative relative to the included greedy candidate; protected allocations cannot delay A. These properties are checked for every generated scenario. No end-to-end runtime or communication-system performance measurement is claimed.

Analytical and computational results

Table 3.

Exhaustive evaluation of minimal feasible plans in the constructed example

Engine A	Aerial A	Engine B	T A min	T B min	Σ P T min
E1	L1	E2	10	13	15.5
E1	L2	E2	13	13	18.2
E2	L1	E1	11	7	13.4
E2	L2	E1	13	7	15.2

The greedy plan has package-ready times of 10 min at A and 13 min at B. Its priority-weighted time sum is 0.9 × 10 + 0.5 × 13 = 15.5 min. The joint optimum assigns E2 and L1 to A and E1 to B, producing 11 min at A and 7 min at B. The weighted sum is 13.4 min, the smallest of the four eligible values in Table 3. Figure 2 makes the engine exchange and retained aerial reserve explicit.

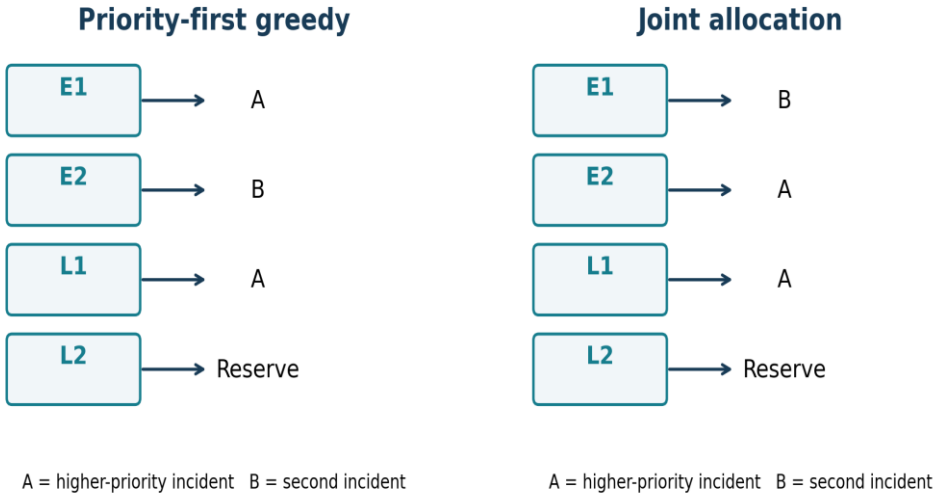


Figure 2. Resource assignment infographic for the constructed example. E1 and E2 exchange incident destinations; L1 serves A and L2 remains in reserve. No geographical coverage outcome is inferred from the reserve count.

Dividing each weighted sum by the total priority 1.4 gives weighted mean package-ready times of 11.071 and 9.571 min. Their difference is 1.500 min, or 90 s. The relative reduction is (15.5 − 13.4)/15.5 × 100 = 13.55%. All values follow directly from the stated assumptions. They are not observed response-time reductions and do not quantify improvements in dispatch computation speed.

Table 4.

Timing benefits and incident-level trade-offs in the analytical example

Outcome	Greedy	Joint	Joint minus greedy
Package-ready time at A min	10	11	+1
Package-ready time at B min	13	7	−6
Priority-weighted mean min	11.071	9.571	−1.500
Required capability shortfall	0	0	0
Unassigned aerial units	1	1	0

Incident-level safeguards and priority sensitivity. The higher-priority incident becomes ready one minute later under the joint optimum. The weighted benefit arises because a six-minute improvement at B outweighs that deterioration at the chosen priorities. A stricter protected bound B A = 10 min makes the joint optimum inadmissible; the greedy plan becomes the lowest-time feasible assignment in the restricted candidate set. This is a substantive change in the feasible decision, not an implementation error.

$$\Delta W = P_A(11 - 10) + P_B(7 - 13) = P_A - 6P_B \quad (8)$$

The exchange improves the weighted objective when P A / P B < 6, ties at a ratio of 6, and worsens it above 6. The ratio in the example is 1.8. Changing priorities can therefore reverse the recommendation even with identical travel times. The result also shows why a priority index cannot substitute for a hard operational requirement: a weighted objective expresses a trade-off, whereas a protected bound prohibits it.

$$G(\delta) = 90 - \delta \quad \text{seconds} \quad (9)$$

G is the reduction in weighted mean package-ready time after adding the same incremental latency δ to both joint-method incidents. At δ = 30 s, the residual advantage is 60 s; at 60 s, it is 30 s; at 90 s, it disappears. At 120 s, the joint method is 30 s worse on this endpoint. This linear result follows from the assumed uniform additive delay and should not be extrapolated to asynchronous acknowledgements or latency that changes resource availability.

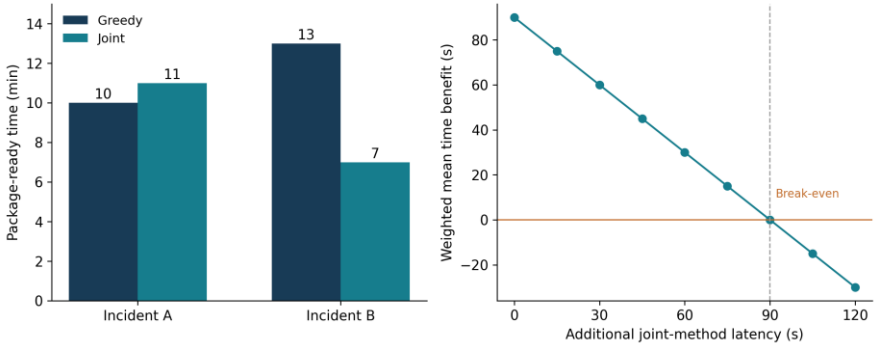


Figure 3. Deterministic timing and latency analysis. Left: incident-specific package-ready times. Right: weighted mean time benefit after adding latency to the joint method alone. The curves are calculations from the constructed scenario, not measurements.

The exact-information experiment yields mean weighted package-ready times of 10.919 min for the greedy method and 9.652 min for joint allocation. The mean paired benefit is 76.02 s, with a Monte Carlo interval of 74.75–77.28 s. Joint allocation improves the aggregate endpoint in 77.61% of scenarios and ties in the remainder; it never worsens that endpoint under exact information because the greedy plan is included in its feasible set. However, A is delayed in 50.99% of scenarios, showing that aggregate optimality does not imply incident-level protection.

Table 5.

Joint allocation performance over 10,000 synthetic scenarios per forecast condition

Error CV	Mean gain s	95% MC interval s	Improved %	Worsened %	A delayed %
0.00	76.02	74.75 to 77.28	77.61	0.00	50.99
0.10	72.69	71.38 to 74.01	70.90	3.04	46.99
0.25	61.08	59.66 to 62.49	57.72	4.93	37.40
0.40	50.81	49.37 to 52.25	46.68	5.36	28.95

Increasing forecast error reduces the observed mean benefit to 72.69, 61.08, and 50.81 s at error CVs of 0.10, 0.25, and 0.40, respectively. At the largest tested error level, 5.36% of realised scenarios

have a worse aggregate endpoint despite the recommendation being optimal for the forecast. The 95th percentile of the weighted scenario endpoint rises from 11.141 to 12.447 min for joint allocation, compared with 12.987 to 13.680 min for the greedy method. These percentiles describe weighted scenario means, not the distribution of individual incident response times.

Table 6.

Protected allocation outcomes and joint-method mean latency thresholds

Error CV	Protected mean gain s	Protected A delayed %	Joint mean break-even s
0.00	34.20	0.00	76.02
0.10	33.35	5.48	72.69
0.25	27.30	8.72	61.08
0.40	22.86	8.37	50.81

With exact information, the protection constraint reduces mean benefit from 76.02 to 34.20 s while eliminating realised deterioration at A. With forecast error, the constraint remains satisfied in the forecast space but A is delayed in 5.48%, 8.72%, and 8.37% of realised scenarios. The non-monotonic final proportion is reported without inferring a general monotonic error-response relationship. The result establishes the distinction between deterministic forecast feasibility and actual protection.

Under the uniform additive latency assumption, the computational mean benefit decreases by one second for each second of incremental delay. Therefore, the scenario-average break-even threshold is 76.02 s with exact information and 50.81 s at error CV 0.40. The nominal deterministic threshold remains 90 s. These are different quantities: one concerns a single specified case and the other the mean of a synthetic scenario distribution. Even below a positive mean threshold, individual scenarios may be worse.

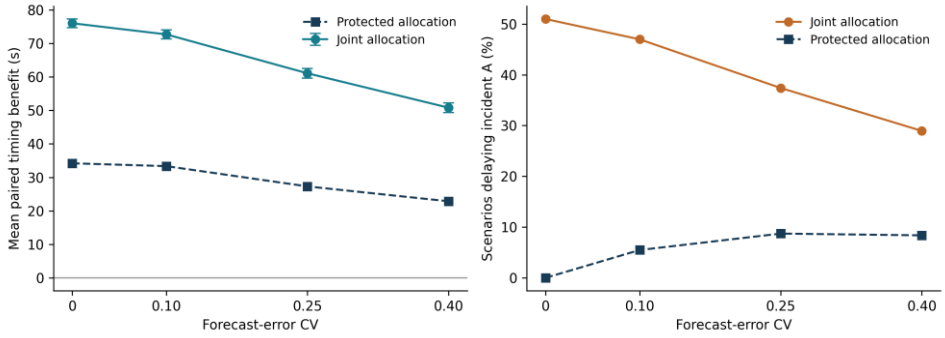


Figure 4. Computational sensitivity to forecast error. Left: mean paired timing benefit, with approximate 95% Monte Carlo intervals for joint allocation. Right: realised incident A delay proportions. All values are generated from the specified synthetic model.

Discussion. The numerical mechanism is the opportunity cost of assigning a flexible resource. E1 is attractive at A because its travel time is shorter than that of E2, but its advantage is partly masked by the arrival of the required aerial unit. At B, the choice of engine determines readiness directly. Reserving E1 for B therefore has a larger marginal effect than assigning it to A. A comparator that considers one incident at a time cannot evaluate this cross-incident opportunity cost within the defined scenario.

The deterministic and computational analyses support a bounded conclusion: simultaneous capability-aware assignment can improve aggregate package timing relative to the defined greedy rule under the tested assumptions. Forecast error reduces this benefit and permits worse realised outcomes even when the forecast objective improves. The higher-priority incident can deteriorate unless constraints are imposed, and forecast-based constraints do not provide realised guarantees under uncertain information. Optimisation quality therefore depends on the operational feasibility criteria and forecast validity as well as the computed objective.

The architectural contribution is to place information quality, allocation feasibility, and human approval in a traceable sequence. Smart firefighting and building intelligence research [2,3] motivates integration of observations, while IRIS-AI [5] provides an established resource-recommendation approach. GUIDE2FR [6] and forest-fire simulation research [7] demonstrate adjacent monitoring and visualisation developments. This study contributes a fully specified small allocation benchmark linking package readiness, forecast protection, and realised timing. It does not claim superior predictive accuracy or operational effectiveness relative to those systems.

The timing value of a digital recommendation depends on its incremental information-to-action cost. Across the computational conditions, the mean allocation benefit ranges from 50.81 to 76.02 s before additional latency, making system, reviewer, and acknowledgement delay material. Deployment should measure the complete chain rather than algorithm execution alone. Trust and responder autonomy also matter for adoption [8]; a recommendation should explain which incident improves, which may deteriorate, and what information could change the decision.

A staged study should first replay historical incidents using only information available at each original decision timestamp. Future observations must not enter earlier recommendations. The local dispatch policy should be reconstructed from documented rules and decision logs, and compared with both the explicit nearest-capable rule and the proposed model. Resource readiness, geography, weather, access restrictions, and concurrent demand should be held identical within each replay comparison.

Controlled exercises should then assess the full human-in-the-loop process, including interpretation and acknowledgement. Scenario order should be counterbalanced to reduce learning effects; repeated observations within crews and stations require clustered analysis. The primary endpoint should be prespecified, with incident-specific protected-bound violations and capability shortfalls reported alongside it. Sample size should follow an operationally meaningful minimum difference and variance estimated from a pilot, rather than an arbitrary number of trials.

Table 7.

Proposed validation outcomes and evidence requirements

Outcome	Measurement	Required evidence
Primary package-ready time	Call receipt to defined joint action	Synchronised task and resource timestamps
Decision latency	Validated inputs to approved recommendation	System and reviewer logs
Capability adequacy	Required versus acknowledged deployed capacity	Verified crew and equipment records
Protected-bound violations	Count and magnitude of exceedances	Prespecified incident-specific limits
Coverage loss	Capability-specific zones below threshold	Validated road network and demand weights
Robustness	Matched performance under data degradation	Prespecified latency and dropout scenarios

For empirical evaluation, paired effect estimates and 95% confidence intervals should be reported with upper-tail timing and protected-bound violations. Bootstrap resampling should respect incident or exercise clusters. These population-inference intervals differ from the Monte Carlo precision intervals reported here. If predictive modules are added, all records from one incident must remain in the same training or test partition; incident-level separation is also described in IRIS-AI [5]. Ablation comparisons should remove one component at a time, and runtime evidence should state hardware, workload, optimisation limits, and returned solution status.

Both numerical evaluations retain two simultaneous incidents, four units, and binary capability requirements. Travel-time variability and forecast errors are independently generated rather than calibrated to a road network, operational records, or sensor performance. Shared congestion, correlated errors, dynamic fire growth, task precedence, water supply, fatigue, route closure, and mutual-aid mobilisation are omitted. The chosen priority indices have not undergone expert elicitation. The full coverage and shortage objective is specified but not evaluated.

No casualty, firefighter exposure, suppression effectiveness, or property-loss outcome is modelled. The nominal 13.55% reduction and Monte Carlo time benefits cannot be converted into lives saved or damage avoided. Monte Carlo intervals quantify numerical precision under the synthetic generator and cannot establish representativeness. The exact-information condition is an ideal reference, while the protected forecast constraint offers no probabilistic guarantee. Stronger protection would require calibrated uncertainty bounds or chance constraints and separate validation of their coverage.

Future work should also test authentication, access control, continuity during network interruption, and intelligibility of recommendations. These are proposed implementation

requirements rather than certified system properties. An operational trial should begin in advisory or shadow mode, with final allocation authority retained by the responsible dispatcher or incident commander and all overrides recorded for analysis.

Conclusions. This study develops an auditable decision support framework for capability-aware resource allocation and evaluates its timing component through deterministic enumeration and a synthetic computational experiment. The nominal example reduces weighted mean package-ready time from 11.07 to 9.57 min but delays the higher-priority incident by one minute. Across 10,000 synthetic scenarios, the exact-information mean paired benefit is 76.02 s; it declines to 50.81 s at forecast-error CV 0.40. These findings demonstrate model behaviour under explicit assumptions, not field effectiveness.

Forecast-based protection eliminates incident A deterioration only in the exact-information condition; under forecast error, realised deterioration remains possible. Additional information-to-action latency can also erase the mean benefit. Consequently, practical evaluation must report aggregate readiness, incident-level outcomes, and information quality together. Historical replay, controlled exercises, and field observation are required before making claims about actual fire-service efficiency or safety.

References:

[1] Averill, J. D., Moore-Merrell, L., Barowy, A. M., Santos, R., Peacock, R. D., Notarianni, K., and Wissoker, D. (2010). Report on Residential Fireground Field Experiments. NIST Technical Note 1661. National Institute of Standards and Technology. https://www.nist.gov/publications/report-residential-fireground-field-experiments-0

[2] National Institute of Standards and Technology. (2015). Research Roadmap for Smart Fire Fighting. NIST Special Publication 1191. https://doi.org/10.6028/NIST.SP.1191

[3] Holmberg, D. G., Raymond, M. A., and Averill, J. D. (2013). Delivering Building Intelligence to First Responders. NIST Technical Note 1648. https://doi.org/10.6028/NIST.TN.1648

[4] Choong, Y. Y., Dawkins, S., Furman, S., Greene, K. K., Spickard Prettyman, S., and Theofanos, M. F. (2018). Voices of First Responders: Identifying Public Safety Communication Problems. Findings from User-Centered Interviews, Phase 1, Volume 1. NISTIR 8216. https://doi.org/10.6028/NIST.IR.8216

[5] Mercadé, S., Tarzan, M., Alzueta, D., Trejo, K., Escrig, J., Serra, S., and Touma, R. (2025). IRIS-AI: An AI-Based Resource Recommender for Emergency Decision Support Systems of Fire Departments. International Journal of Computational Intelligence Systems, 18, 280. https://doi.org/10.1007/s44196-025-00994-2

[6] Delgado Álvaro, M., Novak, R., Fernández Barbosa, P. R., Chicano Capelo, I., Gallego, M., and Rodriguez-Sanchez, M. C. (2025). GUIDE2FR: A smart monitoring platform with a digital twin of a firefighter training tower for emergency scenarios. Internet of Things, 34, 101768. https://doi.org/10.1016/j.iot.2025.101768

[7] Yang, S., Huai, Y., Nie, X., Meng, Q., and Zhang, R. (2024). Visualization of Real-Time Forest Firefighting Inference and Fire Resource Allocation Simulation Technology. Forests, 15(12), 2114. https://doi.org/10.3390/f15122114

[8] Choong, Y. Y., Buchanan, K., Dawkins, S., and Spickard Prettyman, S. (2022). Voices of First Responders: How to Facilitate Adoption and Usage of Communication Technology. An Integrated Analysis of Qualitative and Quantitative Findings. NISTIR 8443. https://doi.org/10.6028/NIST.IR.8443