ACCUSTOM Formulation as a Dynamic Pickup & Delivery Problem (DPDP)

Given: VOXEL-BASED COORDINATE SYSTEM ($x_{max}, y_{max}, z_{max} \rightarrow \Delta x, \Delta y, \Delta z \cdot n_x, n_y, n_z$)

① $p_g(t) = p_g, \forall t, \forall g \in \mathcal{G}, \mathcal{G} = \{1, 2, 3, \ldots, G\}$;

Stationary GNs, GN locations known after random deployment, Total G number of GNs.

② $p_u(0) = [x_u, y_u, z_u] \in \mathcal{F}$ (DEPOT), $\forall u \in \mathcal{U}, \mathcal{U} = \{1, 2, 3, \ldots, U\}$;

All UAVs are at the depot (takeoff/landing/charging zone) at mission start ($t=0$), Total U number of rotary-wing UAVs

③ A GN $g \in \mathcal{G}$ has A_g antennas in a rectangular planar array, driven by A_g Tx/Rx chains via the TDD protocol.

④ A UAV $u \in \mathcal{U}$ has A_u antennas in a rectangular planar array, driven by A_u Tx/Rx chains via the TDD protocol.

⑤ $p_{g_1} \neq p_{g_2}, \forall g_1, g_2 \in \mathcal{G}, g_1 \neq g_2$: GNs cannot overlap. (i.e., they cannot belong to the same voxel)

⑥ $p_{u_1}(t) \neq p_{u_2}(t), \forall u_1, u_2 \in \mathcal{U}, \forall t, u_1 \neq u_2$: UAV collision-avoidance. (i.e., UAVs cannot belong to the same voxel simultaneously)

⑦ $p_u(t) \notin \mathcal{Z}$ (NFZs & OBSTACLES), $\forall u \in \mathcal{U}, \forall t$: UAVs can never belong to the voxels corresponding to NFZs and unavoidable obstacles.

Also, $\mathcal{F} \cap \mathcal{Z} = \phi$.

⑧ Each UAV has an on-board energy source (i.e., a battery) with max capacity E_{max}.

$E_u(0) = E_{max}, \forall u \in \mathcal{U}$: UAVs are assumed to be fully-charged at mission start.

⑨ Equipped with our radio map generation algorithm, we assume that for any given GN Rx ANT (p_m) → UAV Tx ANT (p_n) position pair (p_m, p_n), the channel gain h_{mn} is known.

$$h_{mn} \approx \hat{h}_{mn} = \sum_{r=1}^{R} \left(\beta_r - 10\alpha_r \log_{10} \underbrace{d_{mn}}_{\|p_m - p_n\|_2} + \xi_r + \lambda_r\right) \hat{z}_r(p_m, p_n).$$ (See radio map modeling...)

(10) Traffic offload requests generated by GNs according to a Poisson process with rate Λ requests per unit time.

REQUEST from GN $g \in \mathcal{G}$: $\langle$ TRAFFIC CLASS, $\underset{\text{priority value}}{\chi_g}$, $\underset{\text{max allowable delay}}{d_{g,max}}$, $\underset{\text{payload size}}{\nu_g}$, $\underset{\text{disc. factor}}{\gamma_g}$, $\left\{\begin{matrix}\text{PAYLOAD}\\ \text{FRAMES}\end{matrix}\right\}$ $\rangle$

(BW = W)

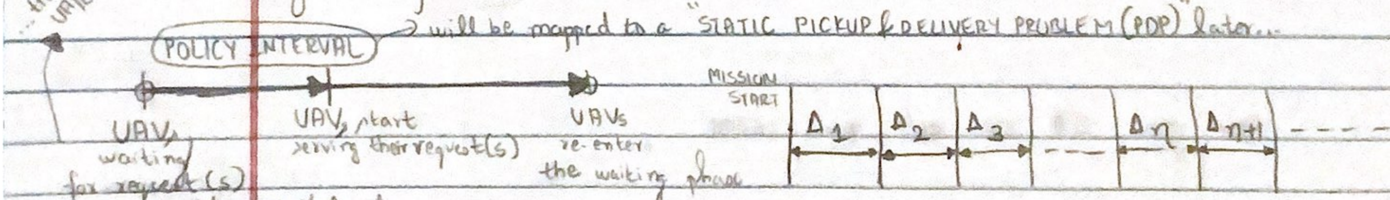
(11) The spectrum allocated to this application (ACCUSTOM) is discretized into U channels, each channel having a preset bandwidth of B. Let the band-edges of this spectrum (each band-edge having a preset bandwidth of B_c, $B_c \ll B$) be designated as control channels for UAV⇔UAV coordination and for control messages between the centralized operations hub and the UAVs in the fleet. Thus, $W = B \cdot U + 2B_c$.

(12) Since the control traffic (UAV⇔UAV and Hub⇔UAV) involves short control frames relative to the large payload frames in the data traffic, we can safely ignore the resultant latencies in our communication model and the subsequent DPDP formulation.

Multi-UAV Orchestration Policy (formally defined later)

(13) Let ζ (zeta) be the policy that defines the GN request scheduling & UAV association, the corresponding UAV positioning along with its constituent 3D trajectory, and the activation/re-charging/deactivation strategy for the UAVs with long-term dynamics in mind.

Then, the expected long-term average reward per UAV in the fleet is defined as follows:



Let Δ_η be the duration of the η^{th} policy interval.

Let $E_{u,\eta}$ be the mobility energy consumption of the UAV u in this policy interval.

Breakdown For UAV $u \in \mathcal{U}$: $\Delta_{u,\eta} = \delta^{(wait)}_{u,\eta} + \delta^{(service)}_{u,\eta} + \delta^{(charge)}_{u,\eta}$ (≥ 0, ≥ 0, ≥ 0)

policy interval duration for a UAV $u \in \mathcal{U}$ $= \underbrace{\delta^{(wait)}_{u,\eta}}_{\text{waiting time component (if any)}} + \underbrace{\sum_{g \in \mathcal{G}_\eta} \psi_{gu} \delta_{gu}}_{\text{service time component (if any)}} + \Xi_{u,\eta} \tau_{u,chrg,\eta}$ } charging time component (if any)

ψ_{gu}: UAV u scheduled to serve GN request g in this policy interval?

$\mathcal{G}_\eta$: GN requests in this η^{th} policy interval

AVERAGE REWARD PER UAV UNDER POLICY ζ: $\bar{\Omega}_\zeta$ (\Omega) (\zeta) $= \lim_{\eta \to \infty}$ (\eta) $\mathbb{E}_\zeta \left[\frac{1}{U(\eta-1)} \sum_{\tau=1}^{\eta-1} \sum_{g \in \mathcal{G}_\eta} \sum_{u=1}^{U} \psi_{gu} \Omega_{gu} \right]$, where

(\tau \to (\iota ota)) 'g' scheduled with 'u'? $\psi_{gu} \in \{0,1\}$

$\Omega_{gu} \triangleq \mathcal{X}_g \gamma_g^{(\delta_{gu} - \delta_{g,max})}$ is the reward obtained by UAV $u \in \mathcal{U}$ for serving GN request g $\langle$CLASS, $\mathcal{X}_g, \delta_{g,max}, \vartheta_g, \gamma_g \rangle$ in δ_{gu} time. (HEADER)

$\delta_{gu} \triangleq \dfrac{\vartheta_g}{\int_{\tau = t_{u,init,\eta}}^{t_{u,term,\eta}} C_{gu}(p_u(\tau), p_g, \mathcal{B}^*, \Phi_{gu}(\tau))}$

δ_{gu} depends on A2G channel conditions, interference from other GNs, the payload size ϑ_g, and the beam-forming design

C_{gu}: Achievable data rate given positions, channel conditions, and beamforming design

AVERAGE ENERGY CONSUMPTION PER UAV UNDER POLICY ζ: $\bar{E}_\zeta = \lim_{\eta \to \infty} \mathbb{E}_\zeta \left(\frac{1}{U(\eta-1)} \sum_{\tau=1}^{\eta-1} \sum_{u=1}^{U} \left[E_{u,\eta} - \Xi_{u,\eta} \tau_{u,chrg,\eta} \omega_u \right] \right)$

PROJECTION $\rightarrow [E_{u,min}, E_{u,max}]$

$\Xi_{u,\eta} \in \{0,1\}$ charge or not? ; $\tau_{u,chrg,\eta}$ charge time ; ω_u UAV $u \in \mathcal{U}$ charge rate energy delivered to battery per unit time

$$E_{u,\eta} = E_{3D}\left(\left[p_{u,\eta}(\tau), \vec{v}_{u,\eta}(\tau) \right]_{\tau = t_{u,init,\eta}}^{\tau = t_{u,term,\eta}} \right)$$

$$= \underbrace{E_{horz}\left(\left\{ \left[p_{u,\eta}(\tau), v_{u,horz,\eta}(\tau) \right]_{\tau = t_{u,init,\eta}}^{\tau = t_{u,term,\eta}} \right\} \right)}_{Q_{u,horz,\eta} \; (\backslash mathcal\{Q\})} + \underbrace{E_{vert}\left(\left\{ \left[p_{u,\eta}(\tau), v_{u,vert,\eta}(\tau) \right]_{\tau = t_{u,init,\eta}}^{\tau = t_{u,term,\eta}} \right\} \right)}_{Q_{u,vert,\eta} \; (\backslash mathcal\{Q\})}.$$

can be a specialized case with the UAV just hovering)...

- All the UAV information parameters $\left\{\mathcal{U}, \{A_u\}_{u=1}^{U}, \{E_{u,min}, E_{u,avg}, E_{u,max}\}_{u=1}^{U}, \{\omega_u\}_{u=1}^{U}, \{v_{u,max}, a_{u,max}\}_{u=1}^{U}\right\}$;

All the GN information parameters $\left\{\mathcal{G}, \{A_g\}_{g=1}^{G}, \{p_g\}_{g=1}^{G}\right\}$;

$\mathcal{B}^* = \{(\alpha_r^*, \beta_r^*, \sigma_{SF,r}^*, K_{SF,r}^*, \pi_r^*) : r = 1,2,3, R\}$.

OBJECTIVE (P0)

$$\underset{\zeta}{\text{maximize}} \quad \bar{\Omega}_{\zeta}\Big(\underset{(\mathcal{I})}{\mathcal{I}_U}, \underset{(\mathcal{I})}{\mathcal{I}_G}, \underset{(\mathcal{B})}{\mathcal{B}^*}\Big)$$

Avg. per UAV energy consumption constraint (C1):

$$\text{such that} \quad \bar{E}_{\zeta}\left(\mathcal{I}_U, \mathcal{I}_G, \mathcal{B}^*\right) \leq \underbrace{\frac{1}{U}\sum_{u=1}^{U} E_{u,avg}}_{\bar{E}_{avg}};$$

Grid tessellation: $0 \leq x \leq x_{max},\ 0 \leq y \leq y_{max},\ 0 \leq z \leq z_{max},\ n_x = \frac{x_{max}}{\Delta x},\ n_y = \frac{y_{max}}{\Delta y},\ n_z = \frac{z_{max}}{\Delta z},\ \Delta x > 0, \Delta y > 0, \Delta z > 0;$

Fixed deployment parameters: $|\mathcal{U}| = U$ (fixed); $|\mathcal{G}| = G$ (fixed); $\mathcal{B}^*$ radio map (pre-determined);

UAV deployment constraints (C2):

$$p_u(0) \in \mathcal{F},\ \forall u \in \mathcal{U},\quad p_{u_1}(t) \neq p_{u_2}(t),\ \forall u_1, u_2 \in \mathcal{U},\ u_1 \neq u_2,\ \forall t \in [t_{u,init,\eta}, t_{u,term,\eta}),\ \forall \eta \in \{1,2,3,\ldots\};$$

$$p_u(t) \notin \mathcal{Z},\ \forall u \in \mathcal{U},\ \forall t \in [t_{u,init,\eta}, t_{u,term,\eta}),\ \forall \eta \in \{1,2,3,\ldots\};$$

Fixed and stationary GN deployment:

$$p_g(t) = p_g \text{ (fixed)},\ \forall g \in \mathcal{G},\ \forall t \in [t_{u,init,\eta}, t_{u,term,\eta}),\ \forall \eta \in \{1,2,3,\ldots\};$$

UAV velocity & acceleration constraints (C3):

$$0 \leq v_{u,horz,\eta}(t) \leq v_{u,max},\quad 0 \leq v_{u,vert,\eta}(t) \leq v_{u,max},$$

$$0 \leq a_{u,horz,\eta}(t) \leq a_{u,max},\quad 0 \leq a_{u,vert,\eta}(t) \leq a_{u,max},\ \forall u \in \mathcal{U},$$

$$a_{u,horz,\eta}(t) \triangleq \dot{v}_{u,horz,\eta}(t),\quad a_{u,vert,\eta}(t) \triangleq \dot{v}_{u,vert,\eta}(t),\ \forall t, v_{u,horz,\eta}(t), v_{u,vert,\eta}(t) \in \mathcal{Q}_{u,\eta},\ \forall \eta \in \{1,2,3,\ldots\};$$

GN scheduling/association constraints (C4):

$$\psi_{gu} \in \{\underset{\text{not scheduled}}{0}, \underset{\text{scheduled}}{1}\},\quad \sum_{u=1}^{U} \psi_{gu} = 1,\ \forall u \in \mathcal{U},\ \forall g \in \mathcal{G}_\eta,\ \mathcal{G}_\eta \subseteq \mathcal{G},\ \forall \eta \in \{1,2,3,\ldots\};$$

GN-UAV Full service constraint (C5):

$$\int_{t=t_{u,init,\eta}}^{t_{u,term,\eta}} C_{gu}(p_u(t), p_g, \mathcal{B}^*, \Phi_{gu}(t))\,dt \geq \vartheta_g,\ \forall u \in \mathcal{U},\ \forall g \in \mathcal{G}_{u,\eta},\ \forall \eta \in \{1,2,3,\ldots\};$$

Simultaneous GN service by UAVs (C6):

$$0 \leq \sum_{g \in \mathcal{G}_\eta} \psi_{gu} \leq |\mathcal{G}_\eta|,\ \forall u \in \mathcal{U},\ \mathcal{G}_\eta \subseteq \mathcal{G},\ \forall \eta \in \{1,2,3,\ldots\};$$

Charging & Battery constraints (C7):

$$\Xi_{u,\eta} \in \{\underset{\text{no charging}}{0}, \underset{\text{charging}}{1}\},\quad 0 < \tau_{u,chrg,\eta} < \Delta_\eta,\quad \omega_{u,\eta} = \omega_u \text{ (fixed charge rate per UAV)},\ \forall u \in \mathcal{U},\ \forall t \in [t_{u,init,\eta}, t_{u,term,\eta}),$$

$$E_{u,min} \leq \underset{\text{(remaining battery)}}{E_{u,rem}(t)} \leq E_{u,max},\quad E_{u,min} \leq E_{u,\eta} - \Xi_{u,\eta}\tau_{u,chrg,\eta}\omega_u \leq E_{u,max},\ \forall \eta \in \{1,2,3,\ldots\}.$$

DPDP Construction:

Summarised Optimization Problem

$$\underset{\xi}{\text{maximize}} \left[\bar{\Pi}_{\xi}\left(\mathcal{I}_U, \mathcal{I}_{G}, \mathcal{B}^*\right) + \varepsilon\left(\bar{E}_{\xi}\left(\mathcal{I}_U, \mathcal{I}_G, \mathcal{B}^*\right) - \bar{E}_{avg}\right)\right]$$

subject to C2, C3, C4, C5, and C6.

$\varepsilon \geq 0$ dual variable

POLICY: ξ

$$\left\{ \underbrace{\{v_{gu}\}_{g\in\mathcal{G}_\eta, u\in\mathcal{U}}}_{\text{SCHEDULE / GN ASSOCIATION ACTIONS}}, \underbrace{\{\Xi_{u,\eta}, \tau_{u,charging,\eta}\}_{u\in\mathcal{U}}}_{\text{CHARGING ACTIONS}}, \underbrace{\{Q_{u,\eta}\}_{u\in\mathcal{U}}}_{\text{UAV POSITIONING \& TRAJECTORY ACTIONS}}, \underbrace{\{\Phi_{gu}(t)\}_{g\in\mathcal{G}_u, u\in\mathcal{U}}^{t=t_{u,term,\eta}}_{t=t_{u,init,\eta}}}_{\text{BEAM-FORMING DESIGN}} \right\}_{\eta=1}^{\infty}$$

"move-and-receive" along the trajectory $Q_{u,\eta}, \forall u\in\mathcal{U}, \forall\eta$

HRL formulation to solve this DPDP

- A policy interval can be modeled as:

LOWER AGENT

Static PDP (MIP problem)

solved via Cross-Layer mTSP

One batch of Cached requests released by the UPPER AGENT (UA) in the HRL framework

Next batch of Cached requests released by the UPPER AGENT (UA) in the HRL framework

η, Δ_η policy interval

- UPPER AGENT (UA) Model

MDP State: S_{UA}

$$\left[N_{cache,\eta}, N_{UAVs,\eta}, \mathcal{I}_{U,\eta}, \mathcal{I}_{G,\eta} \right] \longrightarrow \text{indexed by } \eta.$$

(/mathcal{I}) (/mathcal{I})

$N_{cache,\eta}$: number of GN requests cached in the buffer

$N_{UAVs,\eta}$: number of available UAVs

$|\mathcal{I}_{U,\eta}| = N_{UAVs,\eta}$ info about available UAVs

$|\mathcal{I}_{G,\eta}| = N_{cache,\eta}$ info about the GN requests

$$\mathcal{J}_{U,\eta} \triangleq \left\{ \left(p_u(t_{u,init,\eta}),\ E_{u,rem}(t_{u,init,\eta}) \right) : u = 1,2,3,\ldots,N_{UAVs,\eta} \right\}.$$

$$\mathcal{J}_{G,\eta} \triangleq \left\{ \left(p_g,\ (\chi_g, \delta_{g,max}, \vartheta_g, \gamma_g) \right) : g \in \mathcal{G}_\eta,\ |\mathcal{G}_\eta| = N_{cache,\eta} \right\}.$$

ξ_{UA} ACTION: (a_{UA}) Binary action: Release (1) or Continue Caching (0) "do not release".

REWARD: (r_{UA}) Solve the resultant static PDP given the policy interval (η) if released else the reward is zero.

If released:

$$\bar{\Omega}_{\xi_{LA}}\left(\mathcal{X}_U, \mathcal{X}_G, \mathcal{B}^*, \mathcal{J}_{U,\eta}, \mathcal{J}_{G,\eta}\right)$$

Lower Agent Policy (cross-layer mTSP)

$$+$$

$$\varepsilon\left(E_{\xi_{LA}}\left(\mathcal{X}_U, \mathcal{X}_G, \mathcal{B}^*, \mathcal{J}_{U,\eta}, \mathcal{J}_{G,\eta}\right)\right).$$

UA SOLUTION PROCESS: Deep Q-Network (DQN) to solve the Q-learning (TD) problem, i.e.,

$$L_{UA}(\Theta_{UA,\ell}) = \mathbb{E}_{(s_{UA}, a_{UA}, r_{UA}, s'_{UA}) \sim \mathcal{U}nif(\mathcal{E})}\left[\left(r_{UA} + \gamma_{UA} \max_{a'_{UA}} Q(s'_{UA}, a'_{UA}; \Theta_\ell^-) - Q(s_{UA}, a_{UA}; \Theta_\ell)\right)^2\right]$$

DQN network parameters in training iteration ℓ

replay buffer (\mathcal{E})

UA discount factor

target network parameters

Q-network parameters

Upon releasing the request cache to the Lower Agent (LA):

Use our [Cross-Layer mTSP + LCSO] ICC-2024 approach to solve the resultant static pickup and delivery problem (mTSP/VRP) and hence get the reward for the policy interval η.