Integrating High-Level Priority Decisions with Distributed Safety Filter for Multi-Satellite Collision Avoidance ⋆

Chengrui Shi * **Tao Meng** ** **Hang Zhou** *

* *School of Aeronautics and Astronautics, Zhejiang University, Hangzhou, 310027, China (e-mail: chengruishi,hangzhou@zju.edu.cn).*
** *Huanjiang Laboratory, Zhuji, 311899, China (e-mail: mengtao@zju.edu.cn)*

Abstract:
Miniaturization and dense constellation deployments exacerbate collision risks of future satellites. While numerous collision avoidance strategies have been proposed, few reconcile agent-level safety with mission-level efficiency. In this paper, we propose a distributed inter-satellite collision avoidance framework where low-level safety control is guided by high-level priority decisions. First, we formulate "safe protocol" constraints among satellites and enforce these constraints on nominal controllers through distributed safety filters, establishing collision-free coordination of the swarm. By introducing tunable priority parameters within the safety filter, collision evasion responsibilities become dynamically adjustable, enabling swarm behavior adaptation. We further demonstrate two methods to integrate with high-level decisions: using optimization to approximate global reference behaviors and using Large Language Models to accommodate to tasks, respectively. Theoretical analysis proves the safety guarantees, while numerical experiments validate the framework's efficacy.

Keywords: Multi-satellite, Collision Avoidance, Control Barrier Function

1. INTRODUCTION

Multi-satellite systems offer significant potential in various missions (Bandyopadhyay et al., 2015). Since growing number of spacecraft in orbit dramatically increases collision risks, a strategy guaranteeing real-time safety for inter-satellite collision avoidance becomes necessary. Furthermore, satellites in multi-satellite systems often serve distinct roles (e.g., leaders and followers in formations) and thus should take different collision avoidance responsibilities according to mission requirements. This further demands the compatibility of collision avoidance strategies with high-level decisions.

Extensive research has been conducted to perform safe multi-satellite missions from the control aspect, where artificial potential field (APF), velocity obstacle (VO) and control barrier function (CBF) methods are three representative approaches. APF-based methods coordinate agents through attraction and repulsion field design. For instance, Hwang et al. (2022) proposed a circulating potential field to avoid local minima, while Guan et al. (2024) introduced an APF-based method for satellite formation maintenance. Common limitations of APF-based methods include parameter sensitivity and non-optimal behaviors. VO-based methods address collision avoidance in the velocity space (Douthwaite et al., 2019). These techniques have also been applied to spacecrafts by Li et al. (2025). However, their assumption of constant velocity and velocity-controlled dynamics is often restrictive. CBFs are derived from the forward invariance condition for control-affine systems (Ames et al., 2019). By minimally modifying a nominal control to subject to CBF constraints, provable safety is ensured via safety filters. Leveraging this safety filter technique, methods have been developed for multi-agent systems (Borrmann et al., 2015) and extended to multi-satellite systems (Hibbard et al., 2022). Additional CBF-based approaches address challenges like disturbance (Cheng et al., 2020) and integrate with learning-based control (Liu and Chen, 2024).

Even though the aforementioned works provide *agent-level safety* for multi-satellite systems, few of them are compatable with *high-level decisions*: these methods typically assume homogeneous satellites and struggle to adapt swarm behavior to task-specific requirements. Consequently, swarm behavior remains inflexible despite significant mission variations. To address this problem, we propose a distributed collision avoidance framework for multi-satellite systems incorporating high-level assigned priorities. Inspired by Van Den Berg et al. (2011) for shared collision evasion responsibility and by Chen et al. (2021) for constraint decoupling, we introduce a "safety protocol" for satelite swarm safety. By integrating optimization or large language models (LLMs) to adjust priority parameters within this "protocol", we further enable dynamic adaptation of swarm behavior.

The remainder of this paper is organized as follows: Section 2 provides background on high order control barrier func-

⋆ This research was supported by Zhejiang Provincial Natural Science Foundation of China under Grant No. LR24F030001.

tions, and Section 3 foramlly formulates the problem of this paper. Section 4 introduces the distributed safety filter design, and Section 5 details the integration of safety filters with high-level decisions. Section 6 validates the proposed framework through numerical experiments.

2. HIGH ORDER CONTROL BARRIER FUNCTIONS

Consider a general continuous time control-affine system

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x})\boldsymbol{u}, \tag{1}$$

where $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^n$ is the state and $\boldsymbol{u} \in \mathcal{U} \subset \mathbb{R}^m$ is the system input. $\boldsymbol{f}$ and $\boldsymbol{g}$ are locally Lipschitz continuous functions. The high order control barrier function is defined as follows.

Definition 1. Given system (1) with relative degree r_b and a r_b-th order differentiable function $h(\boldsymbol{x})$, define a series of functions $\Psi_r, r = 0, \dots, r_b$ recursively as

$$\begin{aligned} &\Psi_0 = h(\boldsymbol{x}), \\ &\Psi_k = \dot{\Psi}_{k-1} + \alpha_k\left(\Psi_{k-1}(\boldsymbol{x})\right), k = 1, \dots, r_b, \end{aligned} \tag{2}$$

where $\alpha_k(\cdot)$ are extended class $\mathcal{K}_\infty$ functions [1]. The zero-superlevel set of these defined functions are

$$\mathfrak{S}_r = \left\{\boldsymbol{x} \in \mathbb{R}^n \mid \Psi_r(\boldsymbol{x}) \geq 0\right\}, r = 0, \dots, r_b. \tag{3}$$

h is a *High Order Control Barrier Function (HOCBF)* for system (1), if there exists extended class $\mathcal{K}_\infty$ functions $\alpha_1, \dots, \alpha_{r_b}$ such that

$$\Psi_{r_b}(\boldsymbol{x}) \geq 0 \tag{4}$$

stands for any $(\boldsymbol{x}, t) \in \mathfrak{S} \times [0, \infty]$, where $\mathfrak{S} = \bigcap_{r=0}^{r_b} \mathfrak{S}_r$.

Theorem 1. (Xiao and Belta (2022)). Following the definitions in Definition 1, once h is a HOCBF for system (1), $\mathfrak{S}$ would be a *forward invariant set* for the system, i.e., $\boldsymbol{x}(0) \in \mathfrak{S}, \boldsymbol{x}(t) \in \mathfrak{S}, \forall t > 0$.

3. PROBLEM FORMULATION

3.1 System Modelling

We consider N satellite agents with dynamics governed by Clohessy-Wiltshire equantions (Clohessy and Wiltshire, 1960). The dynamics of agent i in the reference orbit frame is given by

$$\dot{\boldsymbol{x}}_i = \begin{bmatrix} \dot{\boldsymbol{p}}_i \\ \dot{\boldsymbol{v}}_i \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_i \\ \boldsymbol{f}_{vi} \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} \\ E \end{bmatrix} \boldsymbol{u}_i, \tag{5}$$

and

$$\boldsymbol{f}_{vi} = \begin{bmatrix} -2\omega v_{yi} \\ 2\omega v_{xi} + 3\omega^2 v_{yi} \\ \omega^2 p_{zi} \end{bmatrix}, \tag{6}$$

where $\boldsymbol{p}_i = [p_{xi}\ p_{yi}\ p_{zi}]^\top \in \mathbb{R}^3, \boldsymbol{v} = [v_{xi}\ v_{yi}\ v_{zi}]^\top \in \mathbb{R}^3$ and $\boldsymbol{u}_i \in \mathbb{R}^3$ are the position, velocity and acceleration of agent i, respectively. $\boldsymbol{0}$ and E are zero matrix and identity matrix with proper size, and $\omega \in \mathbb{R}$ is the angular velocity of the reference orbit.

Assumption 1. The states $\boldsymbol{x}_j, j = 1, \dots, N$ and ω are known to each agent $i, i = 1, \dots, N$ and to the high-level decision module.

[1] A continuous function $\alpha : \mathbb{R} \to \mathbb{R}$ is an extended class $\mathcal{K}_\infty$ function if $\alpha(0) = 0$ and $\lim_{x \to \pm\infty}(x) = \pm\infty$.

This assumption is justified since agent i could estimate the states of other agents through relative position estimation techniques (Park and D'Amico, 2024), and estimate its own state through navigation techniques. The high-level decision module (e.g., a ground station) could get these information through observation (Pirovano and Armellin, 2024) or communication.

3.2 Safety Constraints

The safety constraints of satellites are to keep the safety distance between each other. Denote $r_i \in \mathbb{R}^+$ to be the safety distance of agent i, the safety constraint between agent i and agent j is then keeping the set

$$\mathfrak{S}_{0ij} = \left\{\boldsymbol{x}_i, \boldsymbol{x}_j \mid d_{ij} = \|\boldsymbol{p}_i - \boldsymbol{p}_j\| \geq R_{ij} = r_i + r_j\right\} \tag{7}$$

forward invariant. And the safety constraints of the swarm is to keep

$$\mathfrak{S}_0 = \bigcap_{i \neq j} \mathfrak{S}_{0ij},\ i, j = 1, \dots, N \tag{8}$$

forward invariant.

3.3 Main Objective

The main objective of this paper is twofold:

1) Ensuring agent-level safety: for each agent i, given the local reference control $\boldsymbol{u}_{ri}$ and observation $\boldsymbol{X} = [\boldsymbol{x}_1^\top\ \dots\ \boldsymbol{x}_N^\top]^\top$, synthesis a safeguarding policy $\boldsymbol{u}_i = \boldsymbol{\pi}_i(\boldsymbol{u}_{ri}, \boldsymbol{X})$ to keep $\mathfrak{S}_0$ forward invariant for the swarm in real time through control $\boldsymbol{U} = [\boldsymbol{u}_1^\top\ \dots\ \boldsymbol{u}_N^\top]^\top$.
2) Cooperating with high-level decisions:
   - Given the low-frequency global reference control $\boldsymbol{U}_{gr} = [\boldsymbol{u}_{gr1}^\top\ \dots\ \boldsymbol{u}_{grN}]^\top$, tune $\boldsymbol{\pi}_i, i = 1, \dots, N$ to approximate $\boldsymbol{U}_{gr}$ with $\boldsymbol{U}$.
   - Given the mission discribed with nature language, tune $\boldsymbol{\pi}_i, i = 1, \dots, N$ to adjust the collision evasion responsibily of agent i based on its mission-level importance.

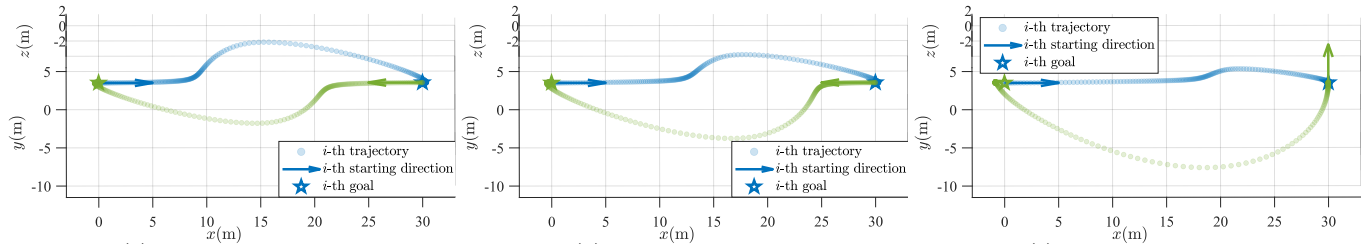
4. DISTRIBUTED SAFETY FILTER DESIGN

We first introduce the "safe protocol" constraint formulation, then give the distributed safety filter design based on the "safe protocol".

For satellite i and satellite j, the "safe protocol" set for satellite i corresponding to satellite j is a half-space described as

$$\begin{aligned} \mathcal{S}_{ij} = \Bigg\{ \boldsymbol{u}_i \mid -\hat{\boldsymbol{n}}_{ij}^\top \boldsymbol{u}_i \leq \hat{\boldsymbol{n}}_{ij}^\top \left[(\alpha_1 + \alpha_2)\boldsymbol{v}_i + \boldsymbol{f}_{vi}\right] \\ + p_{ij}\Bigg[\alpha_1\alpha_2(d_{ij} - R_{ij}) \\ + \frac{1}{d_{ij}}\left(\|\boldsymbol{v}_{ij}\|^2 - (\hat{\boldsymbol{n}}_{ij}^\top \boldsymbol{v}_{ij})^2\right)\Bigg]\Bigg\}, \end{aligned} \tag{9}$$

where $\hat{\boldsymbol{n}}_{ij} = (\boldsymbol{p}_i - \boldsymbol{p}_j)/\|\boldsymbol{p}_i - \boldsymbol{p}_j\|$ is the unit vector pointing from agent j to agent i, $\alpha_1, \alpha_2 \in \mathbb{R}^+$ are positive parameters and $\boldsymbol{v}_{ij} = \boldsymbol{v}_i - \boldsymbol{v}_j$ is the relative velocity between agent i and agent j. $p_{ij} \in \mathbb{R}$ is the priority parameter, reresenting agent i's priority over agent j.



(a) $p_{12} = 0.5, p_{21} = 0.5$ (b) $p_{12} = 0.7, p_{21} = 0.3$ (c) $p_{12} = 1.0, p_{21} = 0.0$

Fig. 1. Agents' behavior with different priority parameters.

The following result renders that two satellites are collision-free if both of the them follow the "safety protocol".

Theorem 2. For systems discribed as (5) and (6), $\mathfrak{S}_{0ij}$ is forward invariant if $\boldsymbol{u}_i \in \mathcal{S}_{ij}, \boldsymbol{u}_j \in \mathcal{S}_{ji}$ and $p_{ij} + p_{ji} \leq 1$.

Proof. The proof mainly leverages Theorem 1. Let

$$h_{ij} = d_{ij} - R_{ij} = \Psi_{0ij} \tag{10}$$

be the HOCBF candidate to keep agent i and agent j collision-free. By choosing positive proportional functions as class $\mathcal{K}_\infty$ functions, it follows the definition that

$$\begin{aligned}\Psi_{1ij} =& \dot{\Psi}_{0ij} + \alpha_1(\Psi_{0ij}) \\ =& \frac{\boldsymbol{p}_i^\top - \boldsymbol{p}_j^\top}{d_{ij}}(\boldsymbol{v}_1 - \boldsymbol{v}_2) + \alpha_1(d_{ij} - R_{ij}) \\ =& \hat{\boldsymbol{n}}_{ij}^\top \boldsymbol{v}_{ij} + \alpha_1 \cdot \Psi_{0ij}.\end{aligned} \tag{11}$$

To get the constraint on control, we further define

$$\begin{aligned}\Psi_{2ij} =& \dot{\Psi}_{1ij} + \alpha_2 \Psi_{1ij} \\ =& \frac{1}{d_{ij}^2}\left((\boldsymbol{v}_i - \boldsymbol{v}_j)^\top d_{ij} - (\hat{\boldsymbol{n}}_{ij}^\top \boldsymbol{v}_{ij})(\boldsymbol{p}_i - \boldsymbol{p}_j)^\top\right)\boldsymbol{v}_{ij} \\ &+ \hat{\boldsymbol{n}}_{ij}^\top(\boldsymbol{u}_i + \boldsymbol{f}_{vi} - \boldsymbol{u}_j - \boldsymbol{f}_{vj}) + \alpha_1 \hat{\boldsymbol{n}}_{ij}^\top \boldsymbol{v}_{ij} \\ &+ \alpha_2(\hat{\boldsymbol{n}}_{ij}^\top \boldsymbol{v}_{ij} + \alpha_1 \Psi_{0ij}) \\ =& \hat{\boldsymbol{n}}_{ij}^\top\left(\boldsymbol{u}_i - \boldsymbol{u}_j + (\alpha_1 + \alpha_2)\boldsymbol{v}_{ij} + \boldsymbol{f}_{vi} - \boldsymbol{f}_{vj}\right) \\ &+ \frac{1}{d_{ij}}\left(\|\boldsymbol{v}_{ij}\|^2 - (\hat{\boldsymbol{n}}_{ij}^\top \boldsymbol{v}_{ij})^2\right) + \alpha_1\alpha_2(d_{ij} - R_{ij})\end{aligned} \tag{12}$$

Given that $\boldsymbol{u}_i \in \mathcal{S}_{ij}$ and $\boldsymbol{u}_j \in \mathcal{S}_{ji}$, by adding up the inequalities in the definition of $\mathcal{S}_{ij}$ and $\mathcal{S}_{ji}$ and substituting $\hat{\boldsymbol{n}}_{ji} = -\hat{\boldsymbol{n}}_{ij}$ into the inequality, one can get

$$\begin{aligned}\tilde{\Psi}_{2ij} =& \hat{\boldsymbol{n}}_{ij}^\top\left(\boldsymbol{u}_i - \boldsymbol{u}_j + (\alpha_1 + \alpha_2)\boldsymbol{v}_{ij} + \boldsymbol{f}_{vi} - \boldsymbol{f}_{vj}\right) \\ &+ (p_{ij} + p_{ji})\left[\frac{1}{d_{ij}}\left(\|\boldsymbol{v}_{ij}\|^2 - (\hat{\boldsymbol{n}}_{ij}^\top \boldsymbol{v}_{ij})^2\right)\right. \\ &\left. + \alpha_1\alpha_2(d_{ij} - R_{ij})\right] \geq 0\end{aligned} \tag{13}$$

Notice that $\|\boldsymbol{v}_{ij}\|^2 - (\hat{\boldsymbol{n}}_{ij}^\top \boldsymbol{v}_{ij})^2 \geq 0$ and $d_{ij} - R_{ij} \geq 0$ hold for $(\boldsymbol{x}_i(0), \boldsymbol{x}_j(0)) \in \mathfrak{S}_{0ij}$, $\Psi_{2ij} \geq \tilde{\Psi}_{2ij} \geq 0$ then stands for any p_{ij} and p_{ji} satisfying $p_{ij} + p_{ji} \leq 1$. Consequently, according to Definition 1 and Theorem 1, h_{ij} is a valid HOCBF and thus $\mathfrak{S}_{0ij}$ is a forward invariant set for agent i and agent j.

Based on the properties of "safe protocol" set, we further design the distributed safeguarding policy in a minimum invasive way as

$$\boldsymbol{\pi}_i(\boldsymbol{u}_{ri}, \boldsymbol{X}) = \begin{cases} \underset{\boldsymbol{u} \in \cap_{j\neq i}\mathcal{S}_{ij}}{\arg\min} \|\boldsymbol{u} - \boldsymbol{u}_{ri}\|^2, & \cap_{j\neq i}\mathcal{S}_{ij} \neq \varnothing, \\ \underset{\boldsymbol{u}\in\mathbb{R}^3}{\arg\min} \max_{j\neq i} \mathrm{ESDF}(\boldsymbol{u}, \mathcal{S}_{ij}), & \cap_{j\neq i}\mathcal{S}_{ij} = \varnothing, \end{cases} \tag{14}$$

where $\mathrm{ESDF}(\boldsymbol{u}, \mathcal{S}_{ij})$ is the Euclidean signed distance function (ESDF) of $\boldsymbol{u}$ to the boundary of half-space $\mathcal{S}_{ij}$, with $\mathrm{ESDF}(\boldsymbol{u}, \mathcal{S}_{ij})$ defined to be negative if $\boldsymbol{u} \in \mathcal{S}_{ij}$.

If $\cap_{j\neq i}\mathcal{S}_{ij} \neq \varnothing$, $\boldsymbol{\pi}_i$ minimally modifies $\boldsymbol{u}_{ri}$ to follow all "safety protocols".And $\boldsymbol{\pi}_i$ is now in the form of a quadratic programming and can be solved in real time onboard (Ames et al., 2019); If $\cap_{j\neq i}\mathcal{S}_{ij} = \varnothing$, it is then impossible for agent i to follow all "safety protocols". Therefore, $\boldsymbol{\pi}_i$ synthesises a control that minimizes the maximum violation of all "safe protocols" to get the "safest possible" control. Such an optimization can also be solved in real time onboard via a linear programming (Van Den Berg et al., 2011).

Theorem 3. If $\cap_{j\neq i}\mathcal{S}_{ij} \neq \varnothing$ for all agent i, $\mathfrak{S}_0$ is forward invariant if $\boldsymbol{u}_i = \boldsymbol{\pi}_i(\boldsymbol{u}_{ri}, \boldsymbol{X})$ and $\boldsymbol{P} = \{p_{ij}\}_{(N\times N)} \in \left\{\boldsymbol{P} \in \mathbb{R}^{N\times N} \mid p_{ii} = 0, p_{ij} + p_{ji} \leq 1, i \neq j\right\}$.

Proof. This result directly follows Theorem 2. Given that $\cap_{j\neq i}\mathcal{S}_{ij} \neq \varnothing, \forall i = 1, \ldots, N$, it cleary renders that $\boldsymbol{u}_i = \boldsymbol{\pi}_i(\boldsymbol{u}_{ri}, \boldsymbol{X}) \in \cap_{j\neq i}\mathcal{S}_{ij}$ is satisfied for all agents. Since for any satellite pair composed of satellite i and satellite j, $\boldsymbol{u}_i \in \cap_{j\neq i}\mathcal{S}_{ij} \subseteq \mathcal{S}_{ij}, \boldsymbol{u}_j \in \cap_{i\neq j}\mathcal{S}_{ji} \subseteq \mathcal{S}_{ji}$ and $p_{ij} + p_{ji} \leq 1$ are satisfied, according to Theorem 2, all $\mathfrak{S}_{0ij}, i \neq j$ are forward invariant. Hence their intersection $\mathfrak{S}_0 = \bigcap_{i\neq j}\mathfrak{S}_{0ij}$, $i, j = 1, \ldots, N$ is forward invariant, rendering that the whole satellite swarm is collision-free.

Remark 1. The priority parameter p_{ij} has a clear physical interpretation. Since $\|\boldsymbol{v}_{ij}\|^2 - (\hat{\boldsymbol{n}}_{ij}^\top \boldsymbol{v}_{ij})^2 \geq 0$ holds universally and $d_{ij} - R_{ij} \geq 0$ if agent i and agent j remain collision-free, the right hand side of (9) is monotonically increasing with p_{ij}. Consequently, higher p_{ij} lowers the control modification required by agent i to satisfy $\boldsymbol{u}_i \in \mathcal{S}_{ij}$. Given the constraint that $p_{ij} + p_{ji} \leq 1$, this leads p_{ij} to a decrease, causing agent j to assume greater responsibility for avoiding collision with agent i. By tuning the priority matrix $\boldsymbol{P}$, high-level decisions can explicitly determine relative collision evasion responsibilities of satellites. Figure 1 illustrates such a behavior: two identical satellites switching positions are simulated with varying p_{ij}. By increasing p_{12} while decreasing p_{21}, satellite 1 (colored in blue) executes less evasive maneuvering, while satellite 2 (colored in red) executes more.

5. COOPERATING WITH HIGH-LEVEL DECISIONS

5.1 Cooperating with Optimization

Centralized control struggles to meet high-frequency control demands for collision avoidance due to communication delays and computational burden. Nevertheless, it can generate global reference behaviors that guide the distributed safety filter by optimizing the priority matrix. With $\boldsymbol{u}_{ri}$ and $\boldsymbol{X}$ fixed, $\boldsymbol{\pi}_i$ depends only on p_{ij} (i.e., $\boldsymbol{\pi}_i = \boldsymbol{\pi}_i(\boldsymbol{P})$). Thus, the priority matrix can be optimized to minimize the deviation of $\boldsymbol{U} = [\boldsymbol{u}_1^\top \ \dots \ \boldsymbol{u}_N^\top]^\top$ from $\boldsymbol{U}_{gr} = [\boldsymbol{u}_{gr1}^\top \ \dots \ \boldsymbol{u}_{grN}]^\top$ through

$$\begin{aligned} \boldsymbol{P}^* &= \underset{\boldsymbol{P}\in\mathbb{R}^{N\times N}}{\arg\min} \|\boldsymbol{U}_{gr} - \boldsymbol{U}(\boldsymbol{P})\|^2 \\ \text{s.t.} &\begin{cases} \boldsymbol{u}_i = \boldsymbol{\pi}_i(\boldsymbol{P}), & \forall i = 1,\dots,N \\ p_{ij} + p_{ji} \le 1, & \forall i \ne j, i, j = 1,\dots,N \\ p_{ii} = 0, & \forall i = 1,\dots,N \end{cases} \end{aligned} \tag{15}$$

By maximizing p_{ij} to satisfy $p_{ij} + p_{ji} = 1$, optimization (15) can thereby be reduced to an unconstrained problem: only the upper triangle elements $p_{ij}, i < j$ need optimization, with the lower triangle obtained through $p_{ji} = 1 - p_{ij}$. This optimization extracts the assignment preferences of centralized control into $\boldsymbol{P}^*$, thereby coordinating the satellite swarm to align with global reference behaviors.

5.2 Cooperating with Large Language Models

While collision avoidance requirements are mathematically straightforward, mission-level objectives are significantly more complex and often specified in natural language. By leveraging the reasoning ability of LLMs, the aforementioned priority parameters can be generated according to the mission requirements, thereby allocating collision evasion responsibilities.

However, directly generate a priority matrix $\boldsymbol{P}$ satisfying $p_{ij} + p_{ji} \le 1, \forall i \ne j$ through LLM is challenging due to the black-box nature of LLMs. As a result, instead of direct matrix generation, we use LLM to produce a priority array $\mathbf{p} = [p_1,\ p_2,\ \dots,\ p_N], p_i \ge 0, \forall i = 1,\dots,N$, with p_i representing the significance of satellite i in the mission. To ensure the output stability, structured output technique is adopted to force the output format on LLMs. After getting the valid output, the priority parameters are calulated through $p_{ij} = p_i/(p_i + p_j)$ to inherently satisfy the safety condition.

Prompts are also carefully engineered to enhance LLM performance. A chain of thought is given to LLMs to first classify satellites by role and mission priority levels, then assign base priorities according to levels, and finally adjust priorities for satellites within the same level. Demonstrations are also provided to leverage the LLM's in-context learning capabilities (Min et al., 2022).

6. NUMERICAL EXPERIMENTS

6.1 Cooperating with Optimization

We validate the proposed method by comparing four approaches: fixed priority, centralized control, optimized priority (proposed), and non-cooperative. For the first case, the control of each agent is calculated distributedly by (14) with $p_{ij} = 0.5, \forall i \ne j$. For the centralized control case, the control is calculated centrally through

$$\begin{aligned} \boldsymbol{U} &= \underset{\boldsymbol{U}'\in\mathbb{R}^{3N}}{\arg\min} \|\boldsymbol{U}' - \boldsymbol{U}_r\|^2 \\ \text{s.t.} &\Psi_{2ij} \ge 0, \forall i \ne j, \end{aligned} \tag{16}$$

where Ψ_{2ij} is defined in (12) and $\boldsymbol{U}_r = [\boldsymbol{u}_{r1}^\top \ \dots \ \boldsymbol{u}_{rN}^\top]^\top$ denotes the reference control. For the optimized priority case, control is calculated distributedly by (14), with $\boldsymbol{P}$ centrally optimized every 10 s using (15) and $\boldsymbol{U}_{gr}$ derives from (16). For the last case, control is calculated distributedly by each agent i through

$$\begin{aligned} \boldsymbol{u}_i &= \underset{\boldsymbol{u}_i'\in\mathbb{R}^3}{\arg\min} \|\boldsymbol{u}_i' - \boldsymbol{u}_{ri}\|^2 \\ \text{s.t.} &\Psi_{2ij} \ge 0, \forall i \ne j, \end{aligned} \tag{17}$$

with the assumption that $\boldsymbol{u}_j = \mathbf{0}, j \ne i$. All approaches above use $\boldsymbol{U}_r$ given by the same proportional derivative (PD) controller.

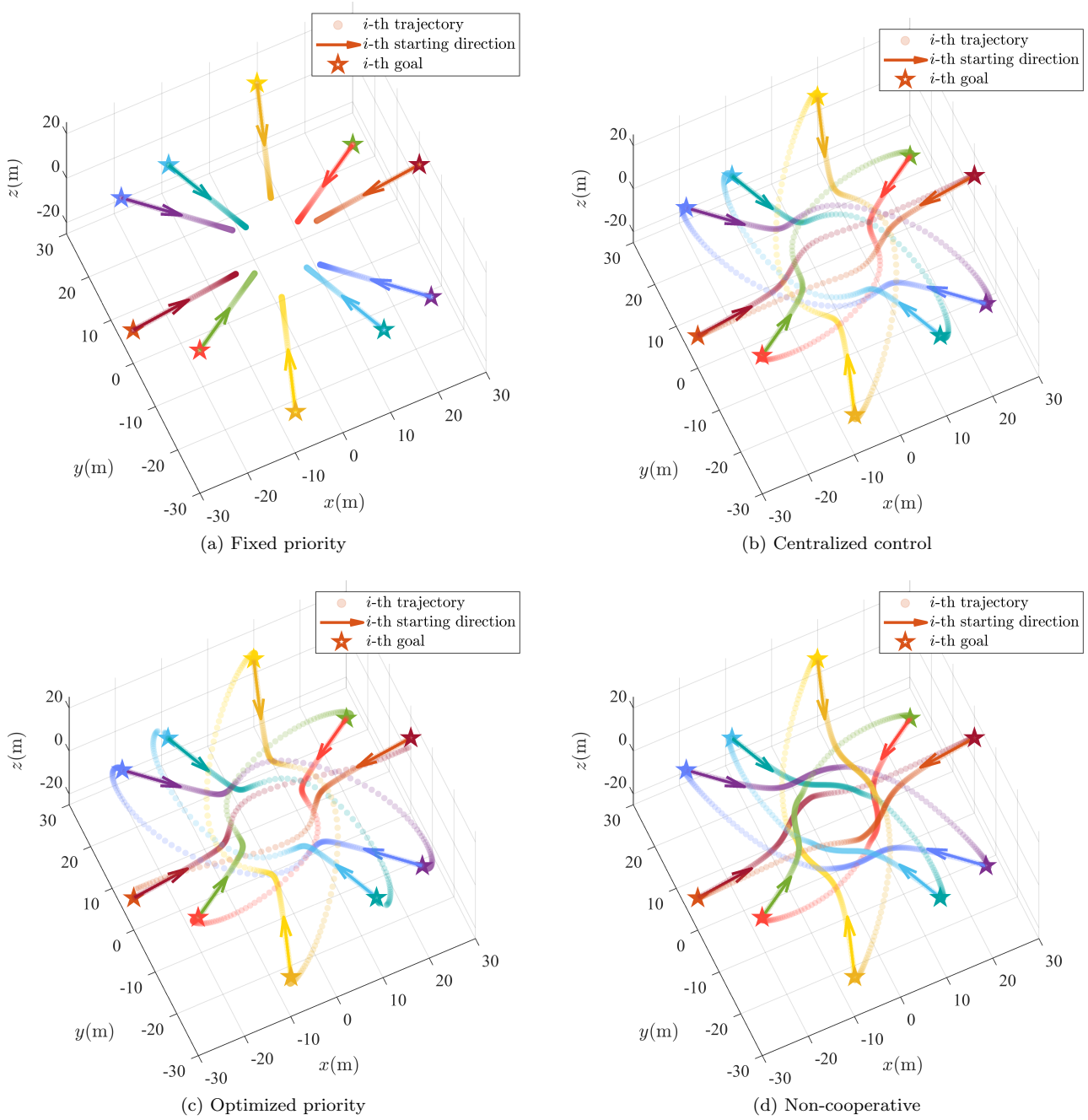
We evaluated all methods above on a position-swapping task where satellites move to their diametrically opposite positions. The trajectory of satellites is shown in Fig. 2 by plotting the position of each agent every 1.5 s. The starting directions of each satellite is illustrated by arrows, with the starting points of the arrows aligned with the initial positions of agents, and the directions of the arrows aligned with the velocity directions of each agent at $t = 0.5$ s. The minimum distance h_{ij} appear in Fig. 3.

As shown in Fig. 2a, satellites encountered a deadlock in the first case, which is related to distributed and local design of the controller. Since the centralized control case optimizes control globally, satellites were coordinated to their goals (Fig. 2b). By using the centralized control above as a low frequency global reference, the proposed method avoided the deadlock as well with distributed control (Fig. 2c). The non-cooperative case coordinated the satellites to their goals aw well (Fig. 2d). However, as shown in Fig. 3, such a method appears to be invalid since collision happened between 150 s and 200 s. This happened due to invalid $\boldsymbol{u}_j = \mathbf{0}, j \ne i$ assumptions. These results above demonstrate that our distributed safety filter can effectively leverage global reference behaviors through priority optimization.

6.2 Cooperating with Large Language Models

We further validated the effectiveness of LLM integration with our safety filter. A local `Gemma3:12b` (Team, 2024) model is deployed to generate priority assignments for a 6-satellite position-swapping task under three distinct prompts. In the first case, LLM was informed that "Satellite 1 is mission critical. All other satellites are backup satellites"; in the second, LLM was given the prompt that "Satellite 1 is mission critical. All other satellites have low fuel"; in the last case, LLM was hinted that "All satellites are the same". The trajectories of satellite are shown in Fig. 4, with T_1 representing the time required by satellite 1 to arrive its goal.

Given the prompts above, LLM outputted the priority array as $\mathbf{p} = [10, 1, 1, 1, 1, 1]$, $\mathbf{p} = [9, 7, 7, 7, 7, 7]$ and $\mathbf{p} = [5, 5, 5, 5, 5, 5]$, respectively. Consequently, the priority parameter $p_{1j}, j \ne 1$ decreased progressively across cases.



(a) Fixed priority

(b) Centralized control

(c) Optimized priority

(d) Non-cooperative

Fig. 2. Swarm behavior with different collision avoidance strategies.

The difference of p_{1j} is then reflected on the trajectory and arriving time of satellite 1. From Fig. 4a to Fig. 4c, satellite 1 experienced increasing safety filter interventions, which in turn increased the curvature of the satellite 1's trajectory and therefore lengthened T_1 from 204.5 s to 321.5 s. This demonstrates our framework's capacity to adapt collision evasion responsibilities to mission requirements through LLM-generated priorities.

REFERENCES

Ames, A.D., Coogan, S., Egerstedt, M., Notomista, G., Sreenath, K., and Tabuada, P. (2019). Control Barrier Functions: Theory and Applications. In *2019 18th European Control Conference (ECC)*, 3420–3431.

Bandyopadhyay, S., Subramanian, G.P., Foust, R., Morgan, D., Chung, S.J., and Hadaegh, F. (2015). A Review of Impending Small Satellite Formation Flying Missions. In *53rd AIAA Aerospace Sciences Meeting*. American Institute of Aeronautics and Astronautics.

Borrmann, U., Wang, L., Ames, A.D., and Egerstedt, M. (2015). Control Barrier Certificates for Safe Swarm Behavior. *IFAC-PapersOnLine*, 48(27), 68–73.

Chen, Y., Singletary, A., and Ames, A.D. (2021). Guaranteed Obstacle Avoidance for Multi-Robot Operations With Limited Actuation: A Control Barrier Function Approach. *IEEE Control Systems Letters*, 5(1), 127–132.

Cheng, R., Khojasteh, M.J., Ames, A.D., and Burdick, J.W. (2020). Safe multi-agent interaction through ro-

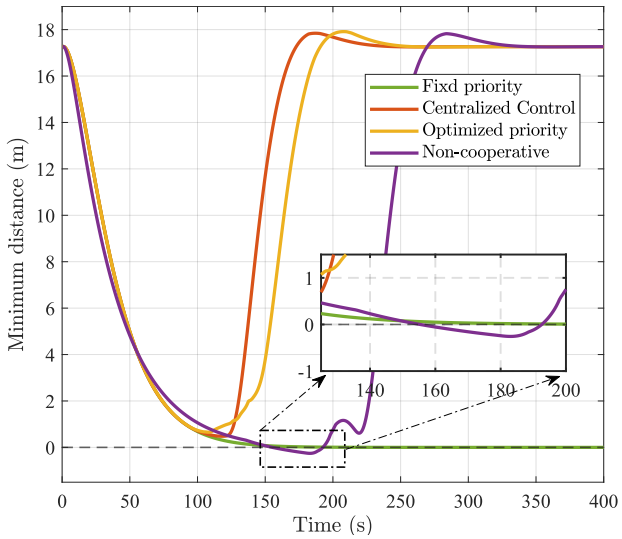
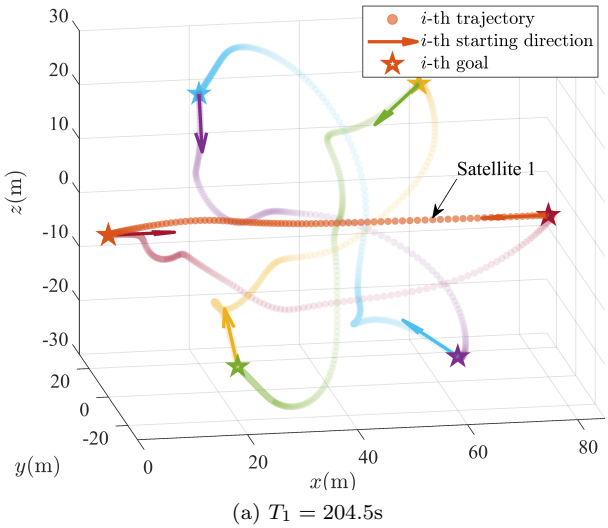
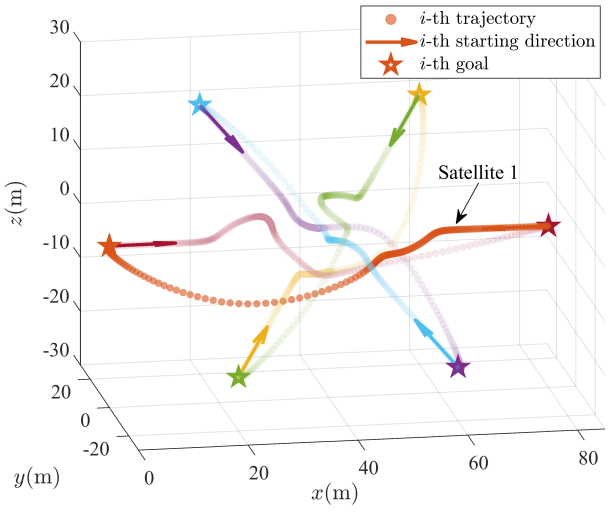


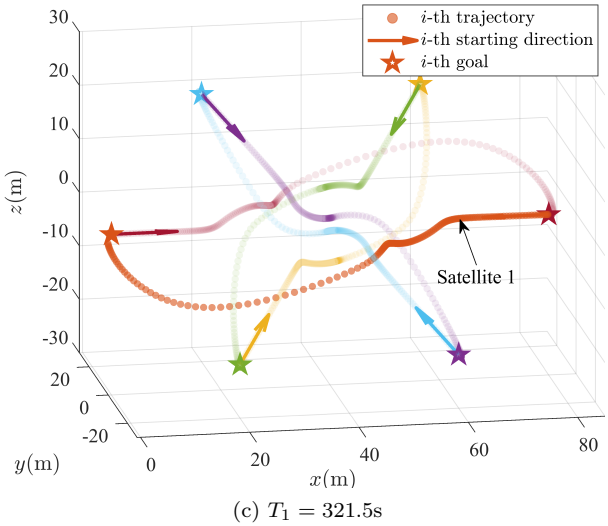
Fig. 3. Minimum distance between satellites with different collision avoidance strategies



(a) $T_1 = 204.5$s



(b) $T_1 = 295.5$s



(c) $T_1 = 321.5$s

Fig. 4. Swarm behavior with different LLM prompts.

bust control barrier functions with learned uncertainties. In *2020 59th IEEE Conference on Decision and Control (CDC)*, 777–783.

Clohessy, W.H. and Wiltshire, R.S. (1960). Terminal guidance system for satellite rendezvous. *Journal of the Aerospace Sciences*, 27(9), 653–658.

Douthwaite, J.A., Zhao, S., and Mihaylova, L.S. (2019). Velocity obstacle approaches for multi-agent collision avoidance. *Unmanned Systems*, 07(01), 55–64.

Guan, Y., Zhang, X., Chen, D., and Fan, S. (2024). Artificial potential field-based method for multi-spacecraft loose formation control. *Journal of Physics: Conference Series*, 2746(1), 012053.

Hibbard, M., Topcu, U., and Hobbs, K. (2022). Guaranteeing Safety via Active-Set Invariance Filters for Multi-Agent Space Systems with Coupled Dynamics. In *2022 American Control Conference (ACC)*, 430–436.

Hwang, J., Lee, J., and Park, C. (2022). Collision avoidance control for formation flying of multiple spacecraft using artificial potential field. *Advances in Space Research*, 69(5), 2197–2209.

Li, Z., Li, H., and Li, C. (2025). Elvo-based autonomous satellite collision avoidance with multiple debris. *Aerospace*, 12(5).

Liu, M. and Chen, Y. (2024). Safety-Guaranteed Learning-Based Flocking Control Design. *IEEE Control Systems Letters*, 8, 19–24.

Min, S., Lyu, X., Holtzman, A., Artetxe, M., Lewis, M., Hajishirzi, H., and Zettlemoyer, L. (2022). Rethinking the Role of Demonstrations: What Makes In-Context Learning Work?

Park, T.H. and D'Amico, S. (2024). Rapid abstraction of spacecraft 3d structure from single 2d image. In *AIAA SCITECH 2024 Forum*.

Pirovano, L. and Armellin, R. (2024). Detection and estimation of spacecraft maneuvers for catalog maintenance. *Acta Astronautica*, 215, 387–397.

Team, G. (2024). Gemma: Open models based on gemini research and technology.

Van Den Berg, J., Guy, S.J., Lin, M., and Manocha, D. (2011). Reciprocal n-Body Collision Avoidance. 70, 3–19. Springer Tracts in Advanced Robotics.

Xiao, W. and Belta, C. (2022). High-order control barrier functions. *IEEE Transactions on Automatic Control*, 67(7), 3655–3662.