Comparative Study for Coordinating Multiple Unmanned HAPS for Communications Area Coverage

Ogbonnaya Anicho, Philip B Charlesworth, Gurvinder S Baicher, Atulya Nagar, and Neil Buckley
Department of Mathematics and Computer Science
Liverpool Hope University
Email: anichoo@hope.ac.uk

***Abstract*—This work compares the application of Reinforcement Learning (RL) and Swarm Intelligence (SI) based methods for resolving the problem of coordinating multiple High Altitude Platform Stations (HAPS) for communications area coverage. Swarm coordination techniques are essential for developing autonomous capabilities for multiple HAPS/UAS control and management. This paper examines the performance of artificial intelligence (AI) capabilities of RL and SI for autonomous swarm coordination. In this work, it was observed that the RL approach showed superior overall peak user coverage with unpredictable coverage dips; while the SI based approach demonstrated lower coverage peaks but better coverage stability and faster convergence rates.**

I. Introduction

HAPS is defined by the International Telecommunications Union (ITU) as "a station located on an object at an altitude of 20 to 50 Km and at a specified, nominal, fixed point relative to the earth" [1]. At this stratospheric altitude, wind profile is described as mild and suitable for hosting platforms with minimal station keeping requirements [2]. HAPS can be used to provide persistent communications coverage to mobile and fixed users; leveraging its unique technical strengths which combines those of terrestrial and satellite communication systems [2]–[4]. The capacity to offer large footprints with signal latency similar to terrestrial systems further places it as a dominant aerial infrastructure. As an aerial platform, it can be easily recovered and redeployed to meet various operational scenarios, an additional capability that neither satellite nor terrestrial systems can offer effectively [5]. However, HAPS are actually distinct from low altitude platforms (LAPS) [6], which typically operate within the troposphere, with lower endurance capabilities and footprints.

The state of the art in operating unmanned HAPS systems require between two (2) to four (4) ground-based crew members overseeing various aspects of mission planning, flight control, sensor operation and data assessment;also known as *many-to-one* ratio [7], [8]. The current capability implies that deploying multiple HAPS platforms as a network may be technically and economically challenging. Operational complexity and cost will likely scale accordingly in scenarios where the platforms work together as a swarm. However, deploying multiple HAPS can extend area coverage capacity using a network of HAPS. The challenge of flipping the *many-to-one* ratio to *one-to-many* ratio is at the core of the multiple HAPS coordination problem. To solve the operating ratio problem will involve designing HAPS platforms with some level of autonomy. Autonomy will eliminate the need for direct human intervention on many operational levels and elevate HAPS platforms/systems to higher layers in the decision making logic hierarchy. Another challenge lies in defining, designing and integrating autonomy solutions and concepts relevant to each use-case or problem.

This work, therefore, focuses on analysing the application of Reinforcement learning (RL) and Swarm Intelligence (SI) in the multiple HAPS coordination problem for communications area coverage. Both algorithms were applied within the same problem scenario and evaluated for performance using metrics like user coverage and algorithm convergence. The performance of the two algorithms were analysed within the context of this work, where multiple HAPS platforms (swarm) are deployed to provide area coverage to mobile users. Self-organising capability is identified as the key autonomy index for any swarm based unmanned aerial systems (UAS) implementation; figure 1 shows a conceptual multiple HAPS network.

In this paper, section I gives an overview of HAPS and the multiple HAPS coordination problem. Section II introduces the concept of autonomy; while section III provides some background on the RL and SI algorithms. Section IV, describes the modeling and simulation methodology applied in this work. In section V, simulation results and analysis are presented. Finally, section VI draws conclusions on the work and considers future work.

II. Fully or Semi-Autonomous and Cooperative HAPS Swarm

This work investigates the implementation of semi or fully autonomous high altitude platform swarm with self-organising capabilities for communications area coverage. The autonomous capability of the HAPS is defined within the context of decision making or self governance within the specific problem area [5]. However, levels of autonomy exist and may depend on design, functions and specifics of the mission [9]. In the application of HAPS or UAS platforms,

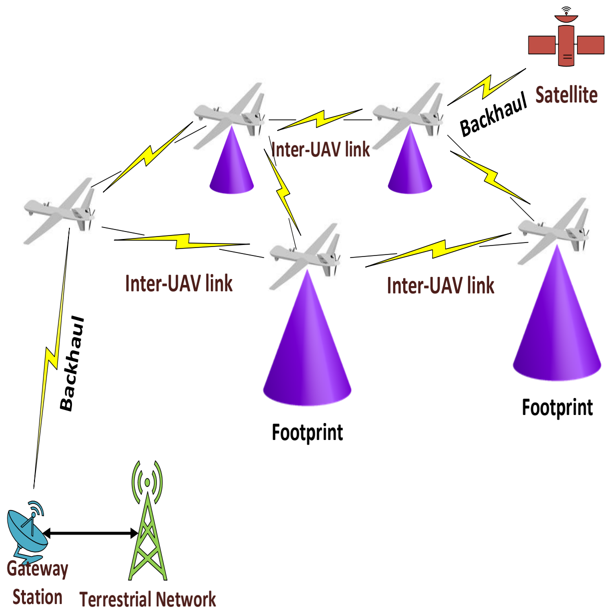

Fig. 1. Conceptual Multiple HAPS Network

autonomy can be a spectrum of capabilities ranging from zero autonomy to full autonomy. For instance, the Pilot Authority and Control of Tasks (PACT) assigns levels of authority, from level 0 (full human pilot authority) to level 5 (full UAV autonomy) [10]. The Autonomous Levels For Unmanned Systems (ALFUS) is more general but very useful model that describes levels of autonomy in unmanned systems [8]. However, the future is that aerial vehicles will have fully autonomous algorithms managing high level mission objectives like maintaining network connectivity, data rate and communications coverage [11]. The question of autonomy is not a linear problem or easily calibrated against any particular autonomy spectrum model. The position this paper assumes is that the definition and application of autonomy will be problem and application specific. For instance, a solar-powered fixed-wing unmanned HAPS deployed for communications area coverage will demand a different set of autonomy requirements in comparison to a quadcopter used for parcel delivery. For this reason, the scope and dimension of autonomy defined for this work is specific to communications area coverage, where the swarm of HAPS self-organises to maximise area coverage for a set of mobile users.

III. Reinforcement Learning & Swarm Intelligence in Unmanned Aerial Systems

The application of reinforcement learning and swarm intelligence to various problem domains is covered in literature. However, this work focuses on the application of these techniques in UAS/HAPS related areas and specifically for autonomous coordination of UAV swarm for communications area coverage. It is important to isolate the specific problem area as these techniques vary significantly with application scenario.

A. Application of Reinforcement Learning in Unmanned Aerial Systems & HAPS

Reinforcement learning (RL) also known as adaptive (or approximate) dynamic programming (ADP) is now a popular technique in solving complex sequential decision-making problems [12]. RL is a paradigm of learning whereby the agent (HAPS in this case) learns through exploring or interacting with its environment. These interactions involve the agent taking actions that trigger transitions from one state to another with associated rewards or punishments. The details and mathematical abstractions for these relationships are covered in the literature. However, this paper will address applications of RL in the literature that are relevant to the problem domain.

A paper by Pham et al [13], proposed a distributed Multi-Agent Reinforcement Learning (MARL) algorithm to tackle the problem of UAV team cooperation for full coverage of an unknown field of interest. This approach demonstrated that teams of UAVs can succeed in the mission without the need of a mathematical model, however, the work was not application specific and further concluded that the stochastic aspect of the problem was not addressed. The stochastic nature of an environment is a critical aspect for consideration and one which this paper highlights (e.g. the mobility of the users/subscribers). In this work the behaviour of the RL algorithm in the presence of stochastic user mobility is considered a key performance indicator. Adaptive state focus Q-learning was applied to solve the problem of learning convergence [14]. To solve the lack of convergence, the learner dynamically expands its state space by incorporating more state information; which is essentially state information of other agents. This approach assumes that the other agents must be accessible and have useful or better state information; this assumption may be problematic and may not resolve the slow convergence issue. An area coverage control in conjunction with reinforcement learning was applied to asymptotically converge UAV agents in optimal configurations but does not consider the impact of intermittent communication network [15]. Cooperative UAS swarm must be able to function if there is a loss of connectivity; this is another performance indicator that should be considered within this work. The work by Hung & Givigi [16], considered the application of reinforcement learning (using Q-learning) to the flocking problem but in the context of followers learning a control policy in the leader-follower topology. This differs from the equal hierarchy topology explored in this work, and moreover does not address the area communications coverage specific scenario. Nguyen et al [17], applied Apprenticeship Bootstrapping via Inverse Reinforcement Learning using Deep Q-learning (ABS via IRL-DQN) to a UAV and UGV (Unmanned Ground Vehicle) coordination task. The UAV was required to maintain about three UGVs in its camera field of view (FoV), however, it differs from the communications area coverage problem considered in this paper. It is not the aim of this paper to provide an exhaustive list of all RL based UAV applications but to identify implementations that reinforces the context of the work.

However, in this paper, the Q-learning approach was adopted; the central idea in the Q-learning algorithm is to store the state-action pair value Q(s, a) called Q-values of each iteration as the agents interact with the environment (Q stands for "Quality"). At the beginning of the simulation the Q-values are initialised to zero and stored in a table or an array. The agent visits some state *s*, and takes action *a*, and then transits into another state. The immediate reward gained from this action is stored and the Q-value updated using the following mathematical relationship [12], [18];

$$Q(s,a) \approx (1-\alpha)Q(s,a) + \alpha\left[r + \gamma\, max_{t+1} Q(s_{t+1}, a_{t+1})\right] \tag{1}$$

Where *r* denotes the reward at time t, $0 < \alpha < 1$ is a given learning rate and γ is discount factor. The expression is used to update the Q-table until the values converge to a near-optimal solution. In the simulation carried out, the HAPS are defined as agents and user mobility modeled as part of the environment and 'states' are pre-selected as fixed coordinates (beacons). The agent can execute two action set: Relocate from or Remain within a 'state' as user density changes due to user mobility. Reward (or penalty) signals are fed back to the HAPS to reinforce actions that influence goals (e.g. maximise user coverage) positively or otherwise.

B. Application of Swarm Intelligence in Unmanned Aerial Systems

Generally, swarm intelligence is inspired by biological systems and their collective behaviour as an organised group [19]. A swarm in this context are simple agents interacting locally within themselves and their environment without any central control leading to an emergent global behaviour [20]. It is important to highlight the decentralised control architecture of swarm creatures which is highly desirable in practical scenarios. In literature different applications of this technique to various problems are available but this work addresses only UAV related applications.

A swarm intelligence approach based on Particle Swarm Optimisation (PSO) was explored by [21], to coordinate multiple UAVs but applied within the target tracking and energy consumption problem context. The application of swarm intelligence for real time UAV coordination for search operations was proposed by Varela et al [22], and specifically applied in an environmental monitoring and pollution source detection problem scenario. An analysis of swarm intelligence based coordination under three specific scenarios was covered by Monteiro et al [23]. These scenarios assessed several features including the self-segregation and self-aggregation and a metric to analyse cohesion of the swarm. The application of swarm intelligence for communications area coverage using fixed-wing unmanned multiple HAPS platforms has unique application requirements. Though in literature different application of swarm intelligence in UAV coordination have been cited but the area coverage scenario is largely unavailable. Due to the advances in machine learning and artificial intelligence, swarm intelligence as originally conceived is not popularly applied.

A swarm intelligence based algorithm (a variant of the bee algorithm) is developed for this work and leverages the strengths of swarm self-organising capabilities. The fundamental concepts to achieving swarm intelligence are self-organisation and division of labour [19], [24]; both influenced the logic behind the algorithm. The participating HAPS in the swarm exchange essential data as they explore the environment akin to foraging. By exchanging critical data and using swarm techniques the HAPS provide persistent and resilient coverage over the area of interest. Figure 2, shows the flow chart for the applied swarm intelligence algorithm, its key aspects, and application to the multiple HAPS coordination problem.

IV. MODELING AND SIMULATION BACKGROUND

In this work, a swarm of multiple solar-powered fixed-wing unmanned HAPS with communications payload, providing area coverage over a specified geographical region was modeled and simulated. In modeling this system some aspects of the system was abstracted to simplify the model without losing relevant system attributes.The following conceptual models (see list below) were implemented in software and constitutes a key aspect of the modeling and simulation process or methodology for this research work. These models are simplified enough to meet the specific scope and interest of the research using standard mathematical and physics models of aerodynamics and communications without compromising theoretical or practical considerations [5].

- HAPS Flight Dynamics Model (FDM).
- HAPS Propulsion Model.
- HAPS Navigation Model.
- HAPS to Ground Link Model.
- Inter-HAPS Link Model.
- User/Subscriber Mobility Model.

Some models, for instance the HAPS FDM were implemented mostly from first principles for higher accuracy, and fidelity of model validation and verification procedures. MATLAB in-built functions were used where appropriate. This practice, in addition to modular codes provided another layer of validation and easy verification through debugging and antibugging. The need for model fitness for purpose and fidelity also informed the decision not to use various open source software/models e.g. JSBSim flight dynamics model.

The parameters in table I, describes the HAPS system communications and link budget parameters which ultimately defines the profile of the service segment e.g. HAPS communications payload power and link data rates. The

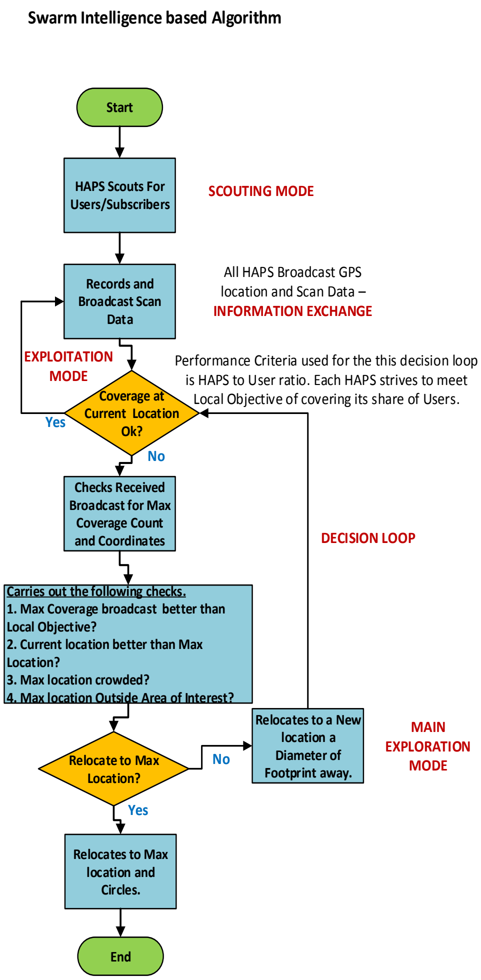

Fig. 2. Swarm Intelligence Algorithm Flow Chart

link budget is based on a payload power of 80 Watts, with the simulated HAPS network supporting about 500 subscribers/users spread over a large area (typical coverage density profile for HAPS). In such thinly populated scenarios, terrestrial networks would not be economical and satellites may be too expensive and ineffective.

V. Results and Analysis

The simulation was run with four (4) HAPS covering an area of about 102,100km^2, with 500 users distributed over the area of interest. The initial distribution of the HAPS and the ground users is shown in figure 3. The ground users move randomly with a mobility model that is not predictable by the HAPS. However, in order to provide a method to validate the algorithms and simulation method; certain benchmarks were introduced. One of such benchmarks was to keep the HAPS platforms fixed (circling around a fixed coordinate), and note the performance. The outcome of this

TABLE I
HAPS System Communications and Link Budget Parameters

S/N	Item	Specification	Justification
1	Half Power Beam Width (HPBW)	145 degrees	Specific to Model
2	Normalised Signal to Noise Ratio (Eb/No)	10 dB	Assumed for Link
3	EIRP	Depends on Slant Range	Power to support 1 subscriber at edge of cover
4	Data Rate	10 Kbit/s	Desired Link Data Rate
5	HAPS Transmitter Antenna Efficiency	0.75	Assumed for Model
6	Ground Receiver Antenna Gain	1	Assumed for Model
7	Signal Frequency	7 GHz	Assumed for Model
8	System Noise Temperature	350K	Standard

experiment will provide a means to measure the performance of the coordination algorithms, and compare outcomes. The following coverage measurement indices are applied;

- Local Coverage: Measures individual HAPS Coverage.
- Global Coverage: Measures Total Coverage (All HAPS combined).

Also note that in computing coverage, no user can be covered by more than one HAPS at a time. This is also applicable to the HAPS; this way duplication of coverage is avoided. It is assumed in this simulation that hard hand-off is implemented in the system (as each HAPS completely releases a user, before the next HAPS attaches it).

HAPS 3, is deliberately located initially where the user density is zero, providing another validation and evaluation condition. In any swarm coordination scenario, it is desirable to see 'starved' or dislocated agents find solution or 'forage' within the foraging space. The performance of this disadvantaged HAPS(agent) will test the self-organisation and coordination robustness of the algorithms in this specific problem environment. The simulation was run for 6 hours, which is reasonable in this problem domain as convergence rate is required to be fast, else users may not be connected and revenue will suffer or in emergency or life-critical missions, lives may be at stake. Previous runs of this simulation also demonstrates that a 6 hour window provides a reasonable time to test convergence, isolate any potential user density dispersion issues and computational overheads. However, extended runs are also carried out to test other aspects or parameters as the research warrants.

A. Performance Analysis without any Coordination Algorithms

This scenario represents the "do nothing" solution, where all the HAPS maintain a fixed circling formation without any coordination algorithms. The performance of the simulated HAPS network is bench-marked with this experiment. Any claim to improved performance due to the coordination

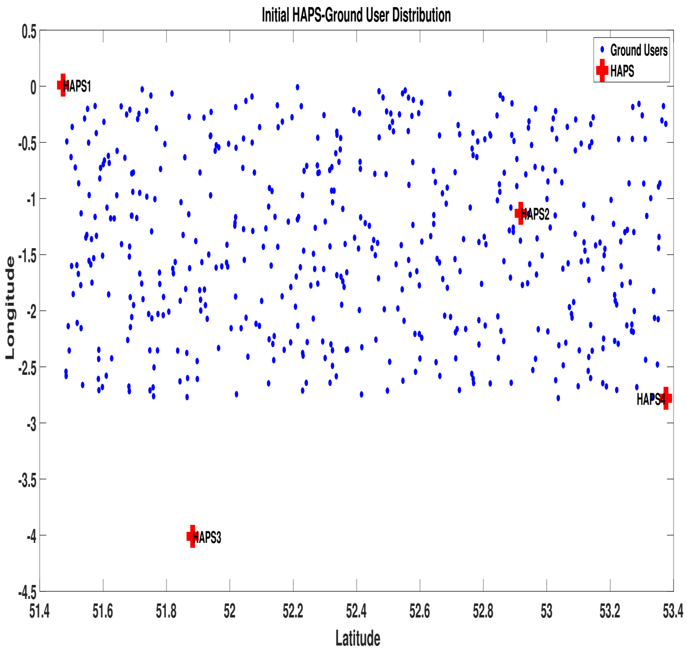

Fig. 3. Initial HAPS versus Ground Users Distribution

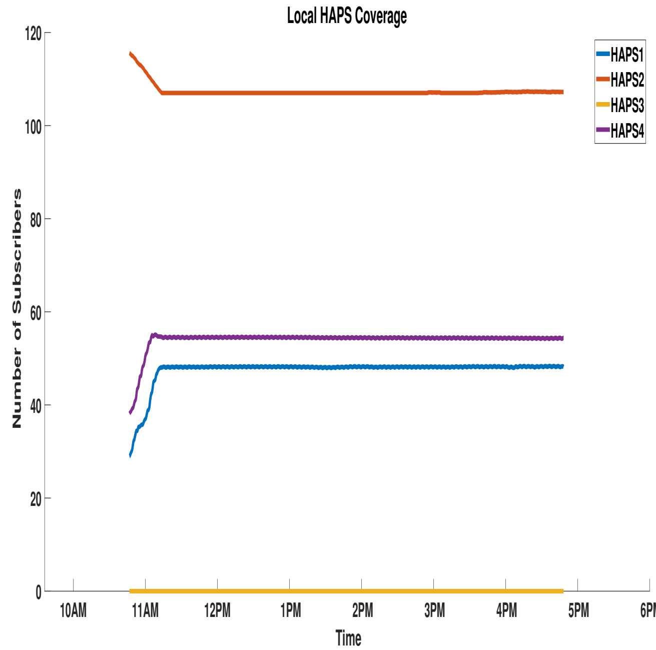

Fig. 4. Local HAPS Coverage without Coordination Algorithm

algorithms must have better global coverage outcomes than the 'do nothing' case. As shown in figure 4, HAPS 1, 2 and 4 maintained their local performance of 45, 110 and 58 users respectively throughout the duration of the experiment. While HAPS 3, as expected had zero coverage throughout the same period. The global coverage performance, (see figure 5, rose from 183 and converged at about 210 users and remained within this threshold throughout the simulation. This outcome provides a performance benchmark as stated earlier, and also shows the peak performance of the system under this circumstance with no upward trends in view.

B. Performance Analysis - Reinforcement Learning Algorithm

In the application of RL for HAPS coordination, it is observed from figure 6, that the algorithm records remarkable peak coverage results, for instance HAPS 3, achieved coverage of over 100 users. Individual HAPS at various times during the simulation recorded good local coverage, reaching almost 250 users. However, due to the exploration-exploitation dilemma, local coverage drops as the HAPS explores other 'states' with probability ϵ. This trend is noticed across all HAPS, as their coverage randomly drops, due to exploration decisions. In contrast the SI method showed stable but lower peak coverages. However, the global coverage performance (see figure 7) of the RL algorithm peaked at above 400 users and clearly demonstrates superior global coverage performance. It is possible that the RL algorithm will achieve some convergence after a long run, however, the nature of the technique inherently explores its environment with some probability. This is a strength and not a weakness and guarantees that the HAPS is never completely biased to exploiting a local optima. This highlights the impact of the learning algorithm approach to HAPS coordination problem. From the results RL coverage performance is better than the SI approach but with the risk of unpredictable dip in performance since navigation is driven solely by exploration decisions. The challenge with the RL based approach lies with balancing the exploration-exploitation trade-off to encourage more exploitation when certain peaks are achieved. But since user mobility implies

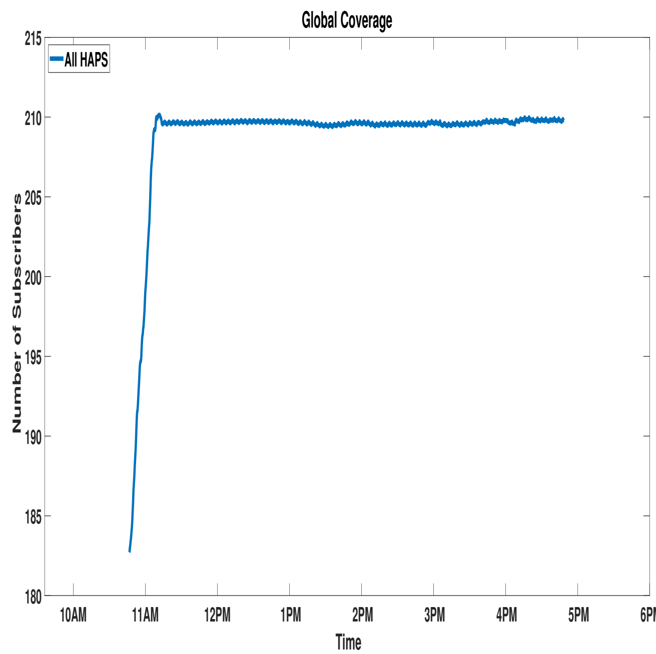

Fig. 5. Global HAPS Coverage without Coordination Algorithm

that 'states' will be stochastic (specifically non-stationary stochastic), inducing more uncertainty, the RL approach demonstrates more resilience to this noise and unpredictable model of mobile user environment. Additionally unlike the SI based method, the RL approach does not depend on a feedback loop or broadcast from the swarm. This quality adds to its resilience to adversarial attacks or normal cross-agent communication faults.

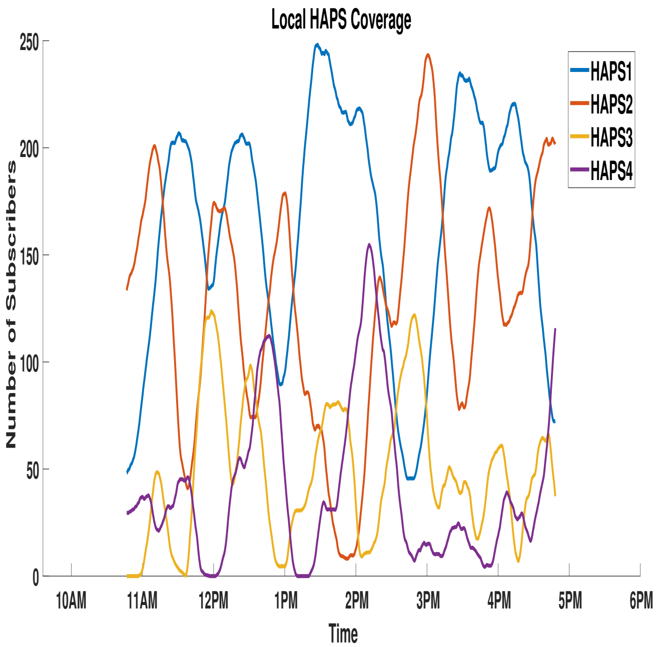

Fig. 6. Local HAPS Coverage with Reinforcement Learning Algorithm

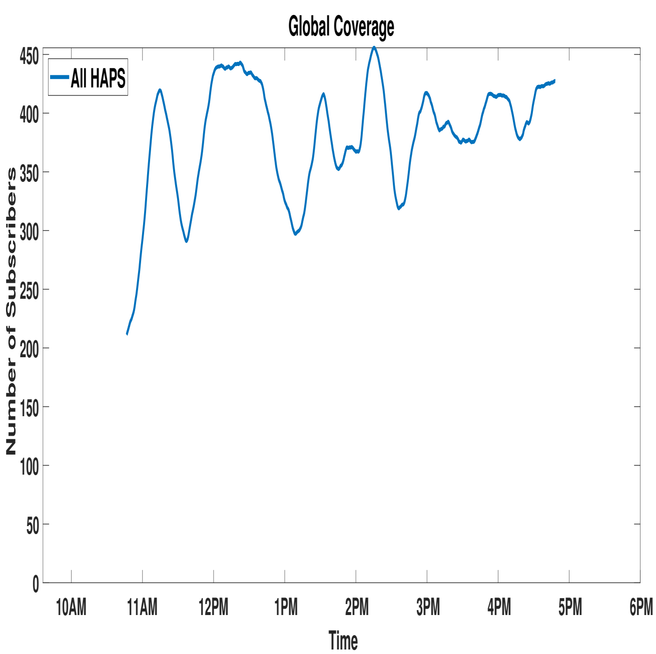

Fig. 7. Global HAPS Coverage with Reinforcement Learning Algorithm

C. *Applying decaying Learning Rate and Adaptive epsilon-greedy in the RL Algorithm*

The RL algorithm was deployed initially using static hyper-parameters, specifically learning rate (1) and epsilon (0.1). However, to further test the capability of the RL method a dynamic epsilon greedy value and decaying learning rate was applied. At the beginning high values were used and systematically reduced as the simulation progressed. Epsilon values range between 0 and 1; 1 indicates high exploration mode (akin to a random walk); while 0 means extreme exploitation. The figure below shows the result of running the experiment with decaying hyper-parameters. The local HAPS coverage (see figure 8) shows more aggressive exploration at the beginning as the epsilon value was set to 1 and decreased to 0.1 towards the end. The same method is used to adjust the learning rate, higher at the beginning and lower towards the end. Note that this run recorded more zero values for local coverage results (indicative of more exploration). However, the global result (see figure 9) showed better global performance reaching peak user coverage of above 450 users. The effectiveness of RL methods is highlighted in this experiment; between the two methods only RL came close enough to covering all users within the simulation period. The main objective in multiple HAPS coordination for communications area coverage is to achieve maximum user coverage.

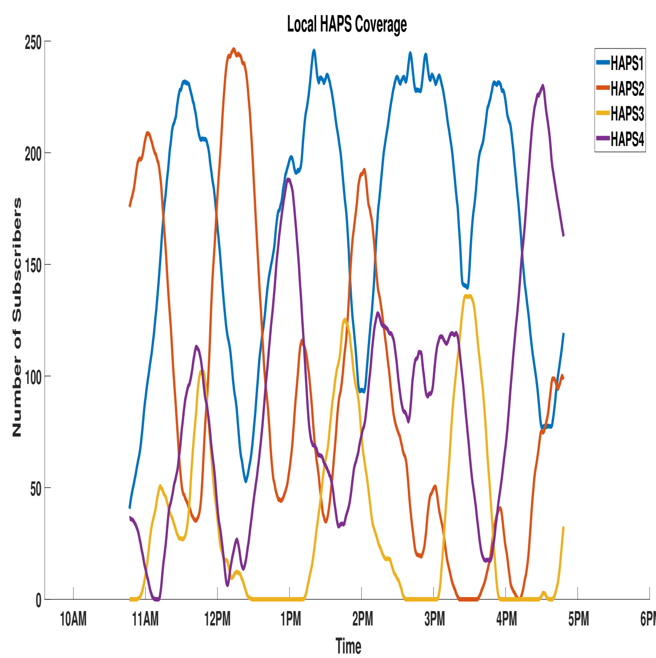

Fig. 8. Local HAPS Coverage with decaying RL Hyper-parameters

D. *Performance Analysis - Swarm Intelligence Algorithm*

In this experiment the swarm intelligence based algorithm is tested under the same set of conditions but with the HAPS allowed to navigate around as required. In figure 10, HAPS 1 and 4, can be seen to have started increasing

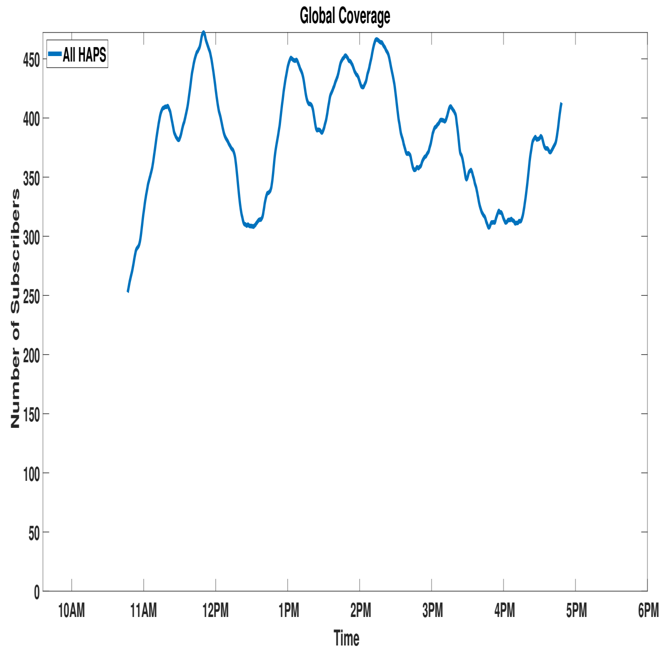

Fig. 9. Global HAPS Coverage with decaying RL Hyper-parameters

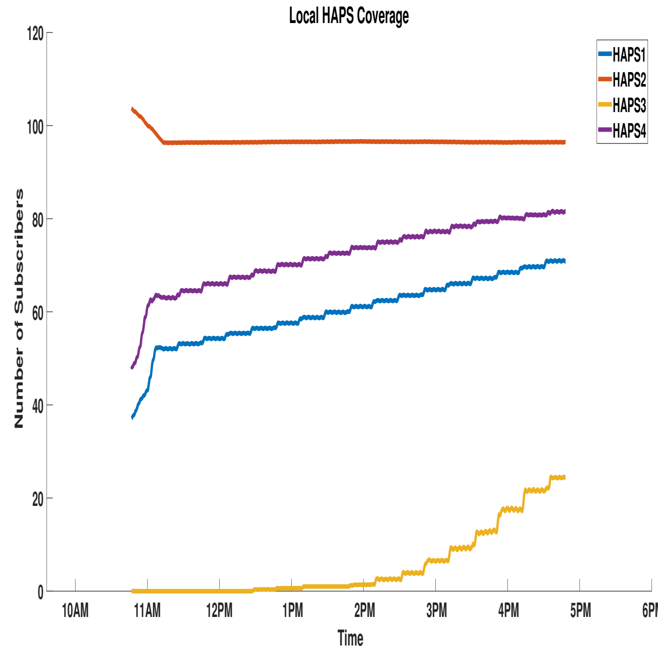

Fig. 10. Local HAPS Coverage with Swarm Intelligence based Algorithm

their coverage almost immediately, a positive development. More significantly HAPS 3, which could not cover any user in the previous 'do nothing' condition, increased its local user coverage from zero to almost 30 users. The cooperative nature of the SI algorithm is the reason HAPS 3, can discover how to navigate to 'rich forage' (high user density) based on the broadcast (feed-back loop) from the swarm. HAPS 2, due to its vantage location easily converges and maintains coverage performance of more than 90 users. The global coverage performance captured in figure 11, reached over 270 users, with a strong upward trend. This positive improvement trend is a sharp contrast to the 'flattening' performance trajectory of the 'do nothing' scenario. In comparison to the RL technique, the SI method did not show any unpredictable coverage dips but converged to a solution in less than 60 minutes of simulation time.

VI. Conclusions and Future Work

This paper compares the performance of RL and SI algorithms in the autonomous coordination of a swarm of HAPS for communications area coverage. It was observed that the SI algorithm showed faster convergence and more stable user coverage profile due to the simple rules-based logic. However, the RL algorithm (applying dynamic epsilon-greedy technique and decaying learning rate) achieved higher overall peak user coverage rates but with some coverage dips due to individual HAPS exploration strategy. RL based techniques demonstrate inherent coordination resilience due to independence from feedback loops and cross-agent communications. This work therefore, concludes that in designing coordination algorithms, swarm intelligence based approaches may be more efficient and reliable but with less optimal coverage results; while RL algorithms will achieve better coverage peaks but at the risk of occasional dips.

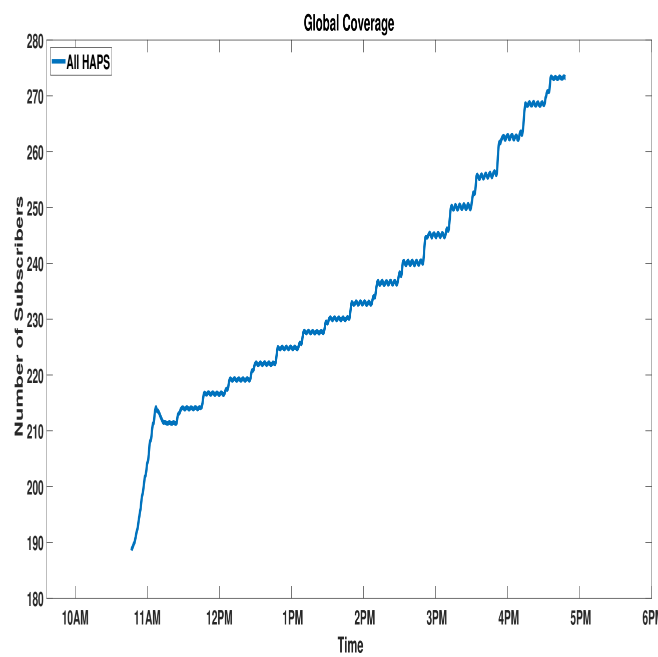

Fig. 11. Global HAPS Coverage with Swarm Intelligence based Algorithm

Future work will consider how to improve the performance of both RL and SI based algorithms by focusing on the observed weaknesses e.g. improve SI based algorithms to reach higher coverage rates, and resolving the unpredictable dips in RL solutions while still maintaining stable user coverage. Another interesting consideration is to develop a hybrid solution which may combine the strengths of both algorithms in ways relevant to this research work.

REFERENCES

[1] International Telecommunications Union (ITU). Terms and definitions. *Radio Regulations Articles*, 2016.

[2] Flavio Araripe d' Oliveira, Francisco CristovÃLourenÃ de Melo, and Tessaleno Campos Devezas. High-Altitude Platforms - Present Situation and Technology Trends. *Journal of Aerospace Technology and Management*, 8:249 – 262, 09 2016.

[3] David Grace and Mihael Mohorcic. *Broadband Communications via High Altitude Platforms*. Wiley, 2011.

[4] ITU. Identifying the Potential of New Communications Technologies for Sustainable Development. Technical report, Broadband Commission For Sustainable Development: Working Group on Technologies in Space and the Upper-Atmosphere, 2017.

[5] Ogbonnaya Anicho, Philip B Charlesworth, Gurvinder S Baicher, and Atulya Nagar. Integrating Routing Schemes and Platform Autonomy Algorithms for UAV Ad-hoc & Infrastructure Based Networks. In *28th International Telecommunication Networks and Applications Conference (ITNAC)*. 28th International Telecommunication Networks and Applications Conference (ITNAC), IEEE, November 2018.

[6] Yong Zeng, Rui Zhang, and Teng Joon Lim. Wireless Communications with Unmanned Aerial Vehicles: Opportunities and Challenges. *IEEE Communications Magazine*, 54(5):36–42, May 2016.

[7] V. Hehtke, J.J. Kiam, and A. Schulte. An Autonomous Mission Management System to Assist Decision Making for a HALE Operator. *Deutscher Luft-und RaumfahrtKongress*, 2017.

[8] Ting B. Chen. *Management of Multiple Heterogenous Unmanned Aerial Vehicles Through Capacity Transparency*. PhD thesis, Queensland University of Technology, 2016.

[9] Hai Chen, Xin min Wang, and Yan Li. A Survey of Autonomous Control for UAV. *IEEE Computer Society*, 2009.

[10] Philip B. Charlesworth. *A Game Theoretic Approach to Coordinating Unmanned Aerial Vehicles with Communications Payloads*. PhD thesis, Cardiff University, 2015.

[11] Zhongliang Zhao and Torsten Braun. Topology Control and Mobility Strategy for UAV Ad-Hoc Networks: A Survey. *Joint ERCIM eMobility and MobiSense Workshop*, 2012.

[12] Abhijit Gosavi. *A Tutorial for Reinforcement Learning*. Springer, 2017.

[13] Huy Xuan Pham, Hung La, David Feil-Seifer, and Luan Nguyen. Cooperative and Distributed Reinforcement Learning of Drones for Field Coverage. 03 2018.

[14] L Busoniu, B Schtter, and R Babuska. Multiagent Reinforcement Learning with Adaptive State Focus. In K Verbeeck, K Tuyls, A Nowe, B Manderick, and B Kuijpers, editors, *BNAIC 2005*, pages 35–42. Proceedings of the 17th Belgium-Netherlands Conference on Artificial Intelligence, October 2005.

[15] Adekunle Adepegba, Suruz Miah, and Davide Spinello. Multi-Agent Area Coverage Control using Reinforcement Learning. In *Proceedings of the Twenty-Ninth International Florida Artificial Intelligence Society Conference*, 2016.

[16] Shao-Ming Hung and Sidney Givigi. A Q-Learning Approach to Flocking with UAVs in a Stochastic Environment. *IEEE Transactions on Cybernetics*, 47(1):186–197, January 2017.

[17] Hung Nguyen, Lam Bui, Matthew Garratt, and Hussein Abbass. Apprenticeship Bootstrapping: Inverse Reinforcement Learning in a Multi-Skill UAV-UGV Coordination Task. In M Dastani, G Sukthankar, E Andre, and S Koenig, editors, *Proceedings of the 17th International Conference on Autonomous and Multiagent Systems*, pages 2204–2206, July 2018.

[18] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2017.

[19] Amrita Chakraborty and Arpan Kumar Kar. *Swarm Intelligence: A Review of Algorithms*. Springer, 2017.

[20] Hazem Ahmed and Janice Glasgow. Swarm Intelligence: Concepts, Models and Applications. Technical report, School of Computing, Queen's University Kingston, Ontario, Canada, 2012.

[21] Diego Silva, Luiz Oliveira, Mariana Macedo, and Carmelo Filho. On the Analysis of a Swarm Intelligence Based Coordination Model for Multiple Unmanned Aerial Vehicles. 2012 Brazilian Robotics Symposium and Latin American Robotics Symposium, IEEE, 2012.

[22] Gervasio Varela, Pilar Caamano, Felix Orjales, Alvaro Deibe, Fernando Lopez-Pena, and Richard Duro. Swarm Intelligence Based Approach for Real Time UAV Team Coordination in Search Operations. In *Proceedings of the 2011 Third World Congress on Nature and Biologically Inspired Computing*, pages 372–377, 2011.

[23] Caio Monteiro, Diego Silva, and Carmelo Bastos-Filho. On the Analysis of a Swarm-Intelligence Coordination Model for Swarm Robots. The Second International Conference on Intelligent Systems and Applications, 2013.

[24] Yichen Hu. Swarm Intelligence. http://guava.physics.uiuc.edu/~nigel/courses/569/Essays_Fall2012/Files/Hu.pdf, 2018. Accessed: 2018-10-23.