A Dual-Network Health State Estimator and Decision Policy for Unmanned Combat Teams

N. Léchevin, C.A. Rabbath, and M. Lauzon

Abstract— We propose a one-step lookahead rollout policy in closed-loop with a health state estimator to ensure effective cooperation among unmanned combat teams despite intermittent wireless communications breakdowns. To ensure effective cooperation despite network faults, the proposed scheme relies on dual networks. On the one hand, a Sensory Information Management Network (SIM-Net) provides the most probable distribution on the location and classification of the adversarial ground units by fusing mobile sensor measurements obtained by a team of surveillance vehicles. On the other hand, a Routing and Munitions Management Network (RMM-Net) enables unmanned combat vehicle (UCV) communications, which are required for their effective path planning and for the distribution of the rollout decision policy over the formations. Simulation results demonstrate the effectiveness of the proposed health state estimator and decision policy.

I. Introduction

In [1], we proposed a decision policy for the routing and munitions management (RMM) of multiple UCVs evolving in an imperfectly known and adversarial environment. Intermittent wireless communications breakdowns were not addressed in [1], despite the facts such events are bound to occur in urban environments and have the potential to adversely affect mission success if not handled correctly. In this paper, we propose to remedy this situation. We study the following scenario. A blue team is composed of surveillance vehicles (SVs) and UCVs. The SVs are remote from the battlefield whereas the UCVs are engaged in the urban theatre. The blue team faces a red team, which consists of ground units distributed over the urban area. The blue team is deployed from a base as a set of UCV formations having to reach a tactical target within a specific time window. To achieve this mission objective, a base policy for the (blue) team is calculated prior to mission and embedded onboard the vehicles. Yet, during mission, the actual ground units location and classification (GULC) may change, and hence may differ from that assumed prior to mission. Therefore, an online policy improvement step is performed during the course of the mission by utilizing the most likely GULC, as obtained from a recursive Bayesian filter (RBF). The expected cost-to-go function of the policy improvement step is approximated, by means of Monte-Carlo simulations, and the decision-making is then updated [1]. To carry out the online computations required to solve the RMM problem, the health state must be exchanged among the formations through the wireless network connecting the UCVs, labeled as RMM-Net. In the RMM problem, the health state refers to the amount of munitions available onboard the UCVs. No matter how reliable is the network, it is subject to failures, delays and information loss due to urban obstacles. Failure to transmit the health state through RMM-Net is expected to significantly reduce the effectiveness of the decision-making.

In this paper, we assume that the sensors found onboard the SVs and the UCVs are part of a second network, denoted as SIM-Net. Furthermore, we assume that SIM-Net is available at times when the RMM-Net may be down, thereby providing a certain level of redundancy. We thus assume that the probability of information loss between the SVs and the UCVs is likely to be smaller than that characterizing information loss among the UCVs. We propose to leverage the availability of SIM-Net to calculate a probabilistic distribution on the possible number of healthy UCVs in each formation of the blue team through the design of a RBF-based health state estimator which feeds the RMM computing nodes, in each formation, in case of intermittent losses of RMM-Net communications. Numerical simulations show that the RBF-based health state estimator is capable of handling intermittent communication losses while resulting in performances that are close to those achieved by the policy in the idealized case when RMM-Net communications are exempt from failures.

II. Modeling, Definitions and Assumptions

A. Urban theater

The urban theater is composed of the red (R) and blue (B) teams. R comprises ground units and decoys located along urban routes with objective of protecting a strategic target location T. B is made up of UCVs and SVs. UCVs have for objective to reach T, simultaneously, from a common deployment base (B_a). Synchronous arrival at T is required to maximize weapons effects on T. For the same reason, maximizing the number of available munitions of B at the time the UCVs reach T is desired. A set of SVs (O) collects observation data from their onboard sensors. Sensing is also performed by the UCVs. The urban terrain is meshed by a $m_1 \times m_2$ grid (G_{UT}), whose nodes are numbered from 1 to $m_1 m_2$. A path is defined as a set of connected edges that link B_a to T. The time is discretized as $t_k = kT_s$, $k \in \mathbb{N}^+$,

The authors are with Defence Research and Development Canada - Valcartier, 2459 Pie-XI N., Québec, Qc, Canada G3J. {nicolas.lechevin.numerica, camille-alain.rabbath, marc.lauzon}@drdc-rddc.gc.ca.

with period $T_s = e/v_\nu$ between two successive discrete time instants, where e denotes the edge length of each mesh and v_ν is vehicle speed.

Definition 1 (Blue team) At the onset of mission, $B = \{1, ..., p\}$ is composed of $p \in \mathbb{N}$ homogeneous formations. Each formation $\nu \in B$ comprises n_p UCVs. A UCV carries $m_p \in \mathbb{N}$ munitions. Every vehicle of a formation ν is assumed to move with constant speed v_ν along paths of G_{UT}. B is able to divide into p smaller formations or into groups of formations. Conversely, grouping into larger formations, or into B itself, is allowed. Formation motion occurs along paths from B_a to T. Each formation can fire at most 1 munition, or salvo, per edge. If $p_e \leq p$ formations are engaged in the same edge, a maximum of p_e munitions, or salvo, can be shot. $O = \{1, ..., p'\}$ is composed of p' homogeneous SVs with onboard sensors. SV o is assigned to a subset $\mathcal{P}o$ of G_{UT}, defined such that Area($\mathcal{P}o_i$) = Area($\mathcal{P}o_j$) and $\mathcal{P}o_i \cap \mathcal{P}o_j = \emptyset$ for all $o_i, o_j \in O$ and $\cup_{i=1}^{p'} \mathcal{P}o_i = G_{UT}$.

Definition 2 (Red team) There is a maximum of one ground unit or decoy between any two adjacent nodes in the urban theater. Whenever p_e formations, $p_e \in [1, p]$, are engaged along an edge where there is a ground unit, the latter is allowed to shoot at most one munition. The firing of the ground unit can destroy $v \geq 0$ vehicles of the p_e formations.

Each $\nu \in B$ is characterized by state $\{N_k^\nu, X_k^\nu\} \in \{0, 1, ..., n_p\} \times \{1, ..., m_1 m_2\} = E^\nu$, which expresses, at t_k, the number of vehicles remaining and the assigned node. One vehicle of ν destroyed by a red unit that is located between two adjacent nodes over $[t_k, t_{k+1})$ corresponds to $N_{k+1}^\nu - N_k^\nu = -1$. The control signal of $\nu \in B$ is $U_k^\nu = (U_{1,k}^\nu, U_{2,k}^\nu) \in \{j \in \{1, ..., m_1 m_2\}\} \times \{0, 1\} = \mathcal{U}_k^\nu$, where $U_{1,k}^\nu = X_{k+1}^\nu$ and $U_{2,k}^\nu \in \{0, 1\}$ denote the location of ν at t_{k+1} and the choice of one vehicle of ν to attack by using one munition over $[t_k, t_{k+1})$, or not to attack ($U_{2,k}^\nu = 0$), respectively. Let N_k, U_k, $U_{1,k}$, $U_{2,k}$ and $\mathcal{U}_k$ denote the p-tuple $(N_k^1, ..., N_k^p)$, $(U_k^1, ..., U_k^p)$, $(U_{1,k}^1, ..., U_{1,k}^p)$, $(U_{2,k}^1, ..., U_{2,k}^p)$, and $(\mathcal{U}_k^1, ..., \mathcal{U}_k^p)$, respectively. A red unit located on (i, j), $i, j \in \{1, ..., m_1 m_2\}$, is characterized at t_k by $Y_{ij,k} \in \{0, 1\}$, where 0 stands for a destroyed status whereas 1 denotes an operational unit. The control signal for a red unit at (i, j) is $V_{ij,k} \in \{0, 1\}$, where 0 corresponds to the unit not attacking and 1 stands for an attack. See [1] for a detailed presentation of the MDP modeling of the problem.

B. Measurement Process to Estimate GULC

Assumption 1 (GULC) Although the actual urban theater is characterized by a fixed GULC, the knowledge of the adversarial team configuration by the command-and-control center (C3) is defined, at t_1, by a number of γ' configurations $s_\gamma, \gamma \in [1, \gamma']$. It is assumed that the C3 does not have sufficient information to derive a reliable distribution over the set of GULC $\{s_\gamma, \gamma \in [1, \gamma']\}$. Each ground unit or decoy is located along anyone of the σ edges of the $m_1 \times m_2$ grid G_{UT}, where $\sigma = 2m_1 m_2 + m_1 + m_2$. A configuration s_γ thus represents, at any t_k, a σ-tuple $s_\gamma = \{s_{\gamma,ij}; i \neq j, 1 \leq i \leq m_1, 1 \leq j \leq m_2, s_{\gamma,ij} = s_{\gamma,ji}\}$, where i and j denote all possible nodes of G_{UT}. $s_{\gamma,ij}$ characterizes the occupancy status of edge (i, j) by an element of R. More precisely, let $s_{\gamma,ij} = 1, 2$, or 3 when there is a ground unit, a decoy, or no unit, respectively, along (i, j). Note that card $s_\gamma = \sigma$.

Definition 3 (Observation, detection and classification) SVs have the capability to observe G_{UT}, while UCVs can observe only the local area; i.e., a UCV $\nu_l, l \in \{1, ..., p\}$, which is located along (i, j), can detect and classify, over $[t_k, t_{k+1})$, ground units and decoys located on edges that are ahead of ν_i and along the axis determined by (i, j). Denote $\mathcal{N}_{Veh}$ as the set of edges that can be sensed by $Veh \in \{B, O\}$; in particular, $\mathcal{N}_o = G_{UT}$. Let $\mathcal{N}_{\mathcal{P}o_i}$ be the sensory set of SV o_i restricted to $\mathcal{P}o_i$. Let $\delta_{ij,k}$ be the distance between $Veh \in \{B, O\}$ and the element of R, if any, located on (i, j) and characterized by $s_{\gamma,ij}$. The ability of Veh to achieve a correct detection and classification ($z_{ij,k}^{Veh} = s_{\gamma,ij}$) of an element of R, if any, is specified by probabilities $p_d^{Veh}(\delta_{ij,k})$ and $p_c^{Veh}(\delta_{ij,k})$, respectively, which are decreasing functions of $\delta_{ij,k}$.

The observation variable $Z_{ij,k}^{Veh}$ of $Veh \in \{B, O\}$ along (i, j) can be assigned, at t_k, one of four values $z_{ij,k}^{Veh} = 1, 2, 3$, and nd, where the first three values have the same meaning as that of $s_{\gamma,ij}$. However, nd indicates that the detection process has failed. The measurement model of $Veh \in \{B, O\}$ is defined by means of the detection probability $p_d^{Veh}(\delta_{ij,k})$, the classification probability $p_c^{Veh}(\delta_{ij,k})$, and the sensor likelihood function $L(z_{ij,k}^{Veh} \mid s_{\gamma,ij})$ and can be expressed, at t_k, as

$$\begin{aligned} & p(Z_{ij,k}^{Veh} = z_{ij}^{Veh} \mid S_{ij,k} = s_{\gamma,ij}) = p(z_{ij,k}^{Veh} \mid s_{\gamma,ij}) \\ & = \begin{cases} p_d^{Veh}(\delta_{ij,k}) L(z_{ij,k}^{Veh} \mid s_{\gamma,ij}) & \text{if} \quad z_{ij,k}^{Veh} \neq nd, \\ 1 - p_d^{Veh}(\delta_{ij,k}) & \text{if} \quad z_{ij,k}^{Veh} = nd, \end{cases} \\ & L(z_{ij,k}^{Veh} \mid s_{\gamma,ij}) = \begin{cases} p_c^{Veh}(\delta_{ij,k}) & \text{if} \quad z_{ij,k}^{Veh} = s_{\gamma,ij}, \\ \frac{1 - p_c^{Veh}(\delta_{ij,k})}{2} & \text{if} \quad z_{ij,k}^{Veh} \neq s_{\gamma,ij}. \end{cases} \end{aligned} \tag{1}$$

Let $Z_k^{Veh} = \{Z_{ij,k}^{Veh}; i \neq j, 1 \leq i \leq m_1, 1 \leq j \leq m_2, Z_{ij,k}^{Veh} = Z_{ji,k}^{Veh}\}$ and $Z_k = \{Z_k^{Veh}; Veh \in \{B, O\}$. z_k^{Veh} and z_k are defined similarly with $z_{ij,k}^{Veh}$ replacing $Z_{ij,k}^{Veh}$. Finally, let $\mathcal{Z}_k$ and z_k denote $\{Z_1, ..., Z_k\}$ and $\{z_1, ..., z_k\}$, respectively. In the sequel, S_k stands for the configuration state-space variable at t_k. Based on (1), the information state vector, subsequently used by the policy, is represented by the distribution $P(S_k = s_\gamma \mid \mathcal{Z}_k = \mathrm{z}_k)$ expressed over all possible configurations $s_\gamma, \gamma \in [1, \gamma']$. Unless specified otherwise, the short notation $P(S_{\gamma,k} \mid \mathrm{z}_k)$ is used in the sequel.

III. Impact of Information Loss on Policy

In [1], we presented one-step lookahead rollout policies $\pi_{1,P_i}^u = \{\mu_{1,P_i}, ..., \mu_{N-1,P_i}\}$ that optimally control the blue team formations despite the adversarial environment for the scenario at hand. The stochastic game was solved in such a way that policies for B and R

play the role of minimizer and maximizer, respectively, $(\pi^{u*}_{1,P_i}, \pi^{v*}_{1,P_i}) = \arg \min_{\pi^u_{1,P_i}} \max_{\pi^v_{1,P_i}} J(N_1, P_i, S_{\gamma^+,1})$, of

$$\begin{aligned} J(N_1, P_i, S_{\gamma^+,1}) = E\{&\textstyle\sum_{\nu \in B}(\sum_{k=1}^{k=N-1} U^\nu_{2,k} \\ &+\textstyle\sum_{k=2}^{k=N} m_p I(N^\nu_k | N^\nu_{k-1}) - m_p n_p)\}, \\ I(N^\nu_k | N^\nu_{k-1}) = &\left\{ \begin{array}{l} 1 \text{ if } N^\nu_k < N^\nu_{k-1}, \\ 0 \text{ if } N^\nu_k = N^\nu_{k-1}. \end{array} \right. \end{aligned} \tag{2}$$

$S_{\gamma^+,1}$, discussed in the sequel, is the GULC utilized at t_1 to derive and to implement the policy. J in (2) represents the negative of the expected number of remaining munitions at T. The policy derived in [1] can be symbolically expressed as

$$U_k \quad = \quad \mu_{k,P_i}(N_k, X_k, P(S_{\gamma,k} \mid \mathrm{z}_k)) \tag{3}$$

where μ_{k,P_i} is a function from $\{E^1 \times ... \times E^p, P(S_{\gamma,k} \mid \mathrm{z}_k)\}$, with $i \in [0, \kappa]$ and $\gamma \in [1, \gamma']$, to $\mathcal{U}_k$. The policy is obtained in three steps. First, the GULC estimator yields the GULC-information state $P(S_{\gamma,k} \mid \mathrm{z}_k)$, $k > 1$, which constitutes a distribution over all possible GULCs conditioned to the observations. Second, base policy $(\pi^{u+}_{1,P_i}, \pi^{v+}_{1,P_i})$ is computed offline by considering $S_{\gamma^+,1}$ that is the most harmful to B. This worst-case scenario is obtained when the uncertain state characterizing the occupancy of an edge is consistently assigned a ground unit. The base policy serves as a reference at subsequent t_k, $k \in [2, i-1]$. Finally, for all t_k, $k > 1$, a one-step lookahead policy improvement is carried out over $[t_k, t_{k+1})$ w.r.t. π^{u+}_{1,P_i} because the estimate of $S_{\gamma,k}$ may evolve from $S_{\gamma^+,1}$. The cost-to-go function is approximated by utilizing, over $[t_k, t_{k+1})$, the maximum a priori (MAP) estimate of $S_{\gamma^+,k}$,

$$\gamma^+ = \underset{\gamma \in [1,\gamma']}{\arg\max} P(S_{\gamma,k} \mid \mathrm{z}_k). \tag{4}$$

The implementation of the policy is described as follows. $P(S_{\gamma,k} \mid \mathrm{z}_k)$, is obtained from data communicated through SIM-Net. The computing node of SIM-Net collects measurement data from other SVs and from the UCVs, and manages the sensor allocation in order to compute the multisensor-multitarget joint Bayes filter. The implementation of the policy is computed in a decentralized fashion over B through RMM-Net. The formations are networked so that they may communicate to each other their state N^ν_k. Expand (3) in [1] as

$$\begin{aligned} &\mu_{k,P_i}(N_k, X_k) = \arg\min\nolimits_{U_k \in \mathcal{U}_k} \max_{V_k}(\textstyle\sum_{\nu \in B} U^\nu_{2,k} + \\ &\tfrac{1}{p}\textstyle\sum_{j=1}^{p} \underbrace{\frac{1}{\eta_l}\sum_{i=1}^{\eta_l} W'_{k,P_i,\gamma^+,\nu_j}(N_k, X_k, U_k, V_k)[i]}_{\widetilde{W'}_{k,P_i,\gamma^+,\nu_j}(N_k,X_k,U_k,V_k)}), \end{aligned} \tag{5}$$

where the double summation represents the Monte-Carlo-simulation-based approximation of the expectation of the cost-to-go function. $[i]$ denotes the realization of the cost-to-go function, $W'_{k,P_i,\gamma^+}(N_k, X_k, U_k, V_k)$, at the ith step, which is determined by means of the base policy $(\pi^{u+}_{k+1,P_i}, \pi^{v+}_{k+1,P_i})$ and of MDPs. Each formation $\nu_j \in [1, p]$ can thus compute $\widetilde{W'}_{k,P_i,\gamma^+,\nu_j}$ provided that ν_j can access (N_k, X_k). Once $\widetilde{W'}_{k,P_i,\gamma^+,\nu_j}, \nu_j \in [1, p]$, are obtained, a consensus computation is required (Fig. 1), so that each formation agrees on a single value, $\frac{1}{p}\sum_{j=1}^{p} \widetilde{W'}_{k,P_i,\gamma^+,\nu_j}(N_k, X_k, U_k, V_k)$, which represents the multiformation approximation of the expected cost-to-go function. Such consensus can be obtained by sharing $\widetilde{W'}_{k,P_i,\gamma,\nu_j}$ through the UCV formations communication links. When the number of formations is large, the average in (5) can be computed, e.g., with a consensus algorithm for multiagent networked systems [4]. Once the agreement is reached, each formation can compute from (5) the policy $\mu_{k,P_i} = [U^1_k, ..., U^p_k]$, which means that each formation v knows the policy, $U^{v'}_k$, applied to any other formation $v' \in B \backslash v$.

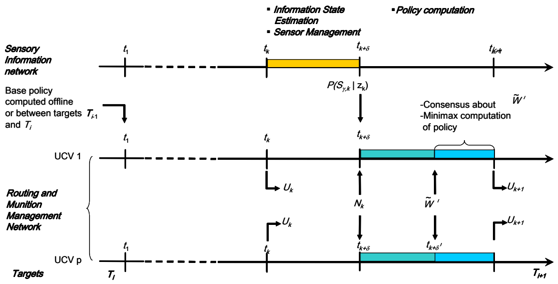
▪ Information State Estimation
▪ Sensor Management
▪ Policy computation
Sensory Information network
t_1
t_k
$t_{k+\delta}$
t_{k+1}
$P(S_{\gamma,k} \mid z_k)$
Base policy computed offline or between targets T_{i-1} and T_i
-Consensus about $\tilde{W}'$
-Minimax computation of policy
UCV 1
U_k
N_k
$\tilde{W}'$
U_{k+1}
Routing and Munition Management Network
$t_{k+\delta'}$
UCV p
Targets
T_i
T_{i+1}

Fig. 1. Computing task scheduling over the set of formations.

As suggested in Fig. 1, intermittent losses of data, due to RMM-Net communications breakdowns, will impede sharing $N^\nu_k, \nu \in B$, and $\widetilde{W'}_{k,P_i,\gamma^+,\nu_j}(N_k, X_k, U_k, V_k)$, among the formations, at $t_{k+\delta}$ and $t_{k+\delta'}$, respectively. Consequently, the rollout policy will be computed from an erroneous team health state, and performance will suffer. Note that $c_{c,k}$ in $\widetilde{W'}$, which represents the position of the formations at t_k, is known from each formation at t_k since $X_k = U_{1,k-1}$ is computed by each formation. The timely knowledge of N^ν_k is critical since destruction of UCVs can occur at anytime over $[t_k, t_{k+1})$, and has to be known by each formation at $t_{k+\delta}$. We propose in the sequel to exploit the ability of the SVs to detect and classify ground units to estimate N^ν_k, for all $\nu \in B$. This task is carried out by designing an additional RBF, denoted as the N_k-information state RBF. For this purpose, consider the following assumption.

Assumption 2 (Available information to SIM-Net) Consider any t_k. SVs are assumed to access N_{k-1} and V_{k-1} prior to an RMM-Net breakdown occurring over $[t_k, t_{k+1}]$. U_{k-1} is also available to SVs from the measurements; i.e., SVs can determine the number of healthy vehicles of the formations at t_k using a graph modeling the authorized discrete-time trajectories to be followed by the formations [5].

IV. SIM-Net-Based Estimation of N_k

The actual GULC is estimated by means of the GULC-information state RBF. A second RBF, labeled as N_k-information state RBF, yields the N_k-information that is sent to each formation connected through RMM-Net (Fig. 2). The N_k-information state RBF depends on the GULC-information state estimate at t_k, and on the last available "true" GULC, assumed to be N_{k-1} (Fig. 2). A preliminary step to obtain the GULC-information state estimate consists of solving a multisensor-multitarget allocation problem. The approach presented in the sequel aims at shaping the posterior distribution to minimize the estimate error.

A. GULC-Information State Estimate

The measurement model in (1) allows one to express the likelihood $P(z_k \mid S_{\gamma,k})$ as follows

$$P(z_k \mid S_{\gamma,k}) = \prod_{Veh \in \{B,O\}} \prod_{(i,j)\in\mathcal{N}_{Veh}} p(z^{Veh}_{ij,k} \mid s_{\gamma,ij}). \quad (6)$$

Assuming that sensor likelihoods are statistically independent,the posterior probability, at t_k, of a GULC, $S_{\gamma,k}$, given observation z_k is obtained from the RBF

$$P(S_{\gamma,k} \mid \mathrm{z}_k) = \frac{P(z_k \mid S_{\gamma,k})P(S_{\gamma,k} \mid \mathrm{z}_{k-1})}{\sum_{g=1}^{\gamma'} P(z_k \mid S_{g,k})P(S_{g,k} \mid \mathrm{z}_{k-1})}, \quad (7)$$

where the prior probability at t_1 is set to $P(S_{\gamma,1})$, as stated in Assumption 1. The conditioned distribution $P(S_{\gamma,k} \mid \mathrm{z}_k)$ obtained for all $\gamma \in [1, \gamma']$ constitutes the information state from which the MAP estimate in (4) is obtained. A possible drawback with the above formulation comes from the fact that the likelihood is a function of measurements obtained from sensors onboard every $Veh \in \{B, O\}$ about every possible target and decoy that lies within the sensing set $\mathcal{N}_{Veh}$. To solve this sensor allocation problem, define the sensory vector as follows. Denote ε_k, for all t_k, as the list $(\varepsilon_{B,k}, \varepsilon_{O,k})$, where $\varepsilon_{B,k} = (\varepsilon_{\nu_1,k}, ..., \varepsilon_{\nu_p,k})$, $\varepsilon_{\nu_i,k} = (\varepsilon_{\nu_i,1,k}, ..., \varepsilon_{\nu_i,\sigma_{\nu_i},k})$, $\varepsilon_{\nu_i,.,k} \in \mathcal{N}_{\nu_i,k}$, and $\sigma_{\nu_i,k} = \mathrm{card}(\mathcal{N}_{\nu_i,k})$ for all $i \in [1, p]$. $\varepsilon_{o,k}$, $\varepsilon_{o_i,k}$, and $\sigma_{o_i,k}$ are defined similarly. Associate to ε_k the sensory vector $\Sigma_k \in \{0, 1\}^{n_k}$, where $n_k = \sum_{i=1}^{p} \sigma_{\nu_i,k} + \sum_{i=1}^{p'} \sigma_{o_i,k}$. The ith entry of Σ_k is 1 whenever the ith entry of ε_k is retained in the computation of $P(z_k \mid S_{\gamma,k})$. The entry is 0 otherwise. The content of Σ_k indicates which sensor is used to compute

$$P(z_k \mid S_{\gamma,k}) = \prod_{u_k=1}^{n_k} p^{\Sigma_k(u_k)}(z_{u_k} \mid s_{\gamma,ij}), \quad (8)$$

where $\Sigma_k(u_k)$ stands for the u_kth entry of Σ_k. z_{u_k} corresponds to $z^{Veh}_{ij,k}$ where $(i, j) \in \mathcal{N}_{Veh,k}$ and Veh represents the vehicle, SV or UCV, that is associated to the u_kth entry of ε_k. $\Sigma_k(u_k)$ is determined by minimizing a measure of the degree of error of $\widehat{S}_{\gamma,k}$, which is accomplished by minimizing the Kullback-Leibler divergence [6]

$$\begin{aligned} I(z_k, \Sigma_k) &= \textstyle\sum_{g=1}^{\gamma'} P(S_{g,k}|z_k) \log \frac{P(S_{g,k}|z_k)}{P(S_{g,k}|z_{k-1})} \\ &= \textstyle\sum_{g=1}^{\gamma'} \frac{P(z_k|S_{g,k})P(S_{g,k}|z_{k-1}) \log \frac{P(z_k|S_{g,k})}{P(z_k|z_{k-1})}}{\underbrace{\displaystyle\sum_{g=1}^{\gamma'} P(z_k|S_{g,k})P(S_{g,k}|z_{k-1})}_{P(z_k|z_{k-1})}}, \end{aligned} \quad (9)$$

where $P(z_k|z_{k-1})$ depends on Σ_k through $P(z_k \mid S_{\gamma,k})$ given in (8). As z_k is not known ahead of time, the worst-case maximizer of (9) is selected as

$$\Sigma^*_k = \arg\max_{\Sigma_k} \min_{z_k} I(z_k, \Sigma_k). \quad (10)$$

B. N_k-Information State Estimate

Assume that RMM-Net is down over $[t_k, t_{k+\delta}]$, thus necessitating an estimate of N_k. This task is carried out by SIM-Net, which observes the formations, updates the RBF, and sends a distribution about N_k to the formations, by virtue of Assumptions 1 and 2. For the SVs, the measurement model used to estimate the number of healthy UCVs that are within a formation ν_i at $t_{k+\delta}$ is defined similarly to the process in (1). The only characterization of ν_i available at $t_{k+\delta}$ is $N^{\nu_i}_{k-1}$, which is assumed known. Here, one aims to estimate $N^{\nu_i}_k$. At $t_{k+\delta}$, $N^{\nu_i}_k \in \{0, 1, 2, ..., N^{\nu_i}_{k-1}\}$ and the measurement ζ^{sv}_k, by $sv \in O$ of $\nu_i \in B$ engaged along (i, j) belongs to $\{N^{\nu_i}_k, nd\}$. The measurement model of ν_i is thus

$$\begin{aligned} &p(\zeta^{\nu_i}_k \mid N^{\nu_i}_k) \\ &= \begin{cases} p^{sv}_d(\delta_{ij,k})L(\zeta^{\nu_i}_k \mid N^{\nu_i}_k) & \text{if} \quad \zeta^{\nu_i}_k \neq nd, \\ 1 - p^{sv}_d(\delta_{ij,k}) & \text{if} \quad \zeta^{\nu_i}_k = nd, \end{cases} \\ &L(\zeta^{\nu_i}_k \mid N^{\nu_i}_k) = \begin{cases} p^{sv}_c(\delta_{ij,k}) & \text{if} \quad \zeta^{\nu_i}_k = N^{\nu_i}_k, \\ \frac{1-p^{sv}_c(\delta_{ij,k})}{N^{\nu_i}_{k-1}} & \text{if} \quad \zeta^{\nu_i}_k \neq N^{\nu_i}_k. \end{cases} \end{aligned} \quad (11)$$

where p^{sv}_d and p^{sv}_c denote the probability of detection and classification of ν_i by sv. By classification, it is meant the process by which a value is assigned to the observation variable $\zeta^{\nu_i}_k$ given the actual state of ν_i being $N^{\nu_i}_k$. $\delta_{ij,k}$ is the distance that separates ν_i, engaged along (i, j), from sv. The following notation is adopted in the sequel: $\zeta_k = \{\zeta^1_k, ..., \zeta^{\nu_p}_k\}$, $\zeta_{1,k} = \{\zeta_1, ..., \zeta_k\}$. Following a Bayesian approach similar to that of the last subsection, the probability that the state $N_{j,k}$, at $t_j \in [t_k, t_{k+\delta})$, be equal to $\widetilde{N}_{j,k} \in \Pi^p_{i=1}\{0, 1, 2, ..., N^{\nu_i}_{k-1}\}$ given $\zeta_{1,j}$, which have been collected over $[t_k, t_j]$, is given by

$$\begin{aligned} &P(\widetilde{N}_{j,k}|\zeta_{1,j}) = \\ &\frac{P(\zeta_j|\widetilde{N}_{j,k})P(\widetilde{N}_{j,k}|\zeta_{1,j-1})}{\sum_{\overline{N}_{j,k}\in\Pi^p_{i=1}\{0,1,2,...,N^{\nu_i}_{k-1}\}} P(\zeta_j|)P(\overline{N}_{j,k}|\zeta_{1,j-1})}, \end{aligned} \quad (12)$$

where the joint likelihood is expressed as $P(\zeta_j|\widetilde{N}_{j,k}) = \Pi^p_{i=1} p(\zeta^{\nu_i}_j \mid N^{\nu_i}_{\widetilde{j,k}})$. The recursion in (12) is initialized by selecting $P(\widetilde{N}_{1,k}|\zeta_{1,1}) = P(\widetilde{N}_{1,k})$ from U_{k-1}, N_{k-1}, and the worst-case value of V_{k-1}, which are known by virtue of Assumption 2. From MDPs in [1] the probability that $d_i \geq 0$ vehicles of ν_i, which is part of l formations engaged along (i,j), be destroyed can be expressed as a probability function $P_{\nu_i}(d_i, l, \gamma, U_{k-1}, V_{k-1})$, with $\gamma \in [1, \gamma']$. Summing with respect to γ yields from (7), the marginal distribution

$$\begin{array}{l} P_{\nu_i}(d_i, l, U_{k-1}, V_{k-1} \mid \mathrm{z}_k) = \\ \sum_{\gamma \in [1,\gamma']} P(S_{\gamma,k} \mid \mathrm{z}_k) P_{\nu_i}(d_i, l, \gamma, U_{k-1}, V_{k-1}), \end{array} \tag{13}$$

which is selected to initialize (12); i.e.,

$$\begin{array}{l} P(\widetilde{N}_{1,k}) = [P(\widetilde{N}_{1,k}^{\nu_1}), ..., (\widetilde{N}_{1,k}^{\nu_p})], \\ P(\widetilde{N}_{1,k}^{\nu_i}) = P_{\nu_i}(d_i, l, U_{k-1}, V_{k-1} \mid \mathrm{z}_k), \end{array} \tag{14}$$

where $\widetilde{N}_{1,k}^{\nu_i} = N_{k-1}^{\nu_i} - d_i$ for $d_i \in \{0, 1, ..., N_{k-1}^{\nu_i}\}$. Once $P(\widetilde{N}_{j,k}|\zeta_{1,j})$ is obtained, the estimate $\widehat{N}_k$ utilized in (5) has to be selected. The MAP estimate, $\widehat{N}_k = \underset{\widetilde{N}_{j,k} \in \Pi_{i=1}^{p}\{0,1,2,...,N_{k-1}^{\nu_i}\}}{\arg\max} P(\widetilde{N}_{j,k}|\zeta_{1,j}),\ j > 1$, is one solution. A more advanced although still suboptimal approach consists in considering a risk-averse stochastic control perspective [2]. The MAP estimate is utilized to constrain the computational load. The closed-loop dynamics, composed of the MDPs [1], (4)-(5), and (6)-(10) is now augmented with the RBF (12)-(13) (Fig. 2). The N_k-information state RBF is utilized whenever RMM-Net breakdown prevents from communicating N_k^{ν}, $\nu \in B$. Although not directly applicable to our work, [7] provides an interesting analysis of the number of iterations needed so that RBF-based estimate achieves a given confidence level.

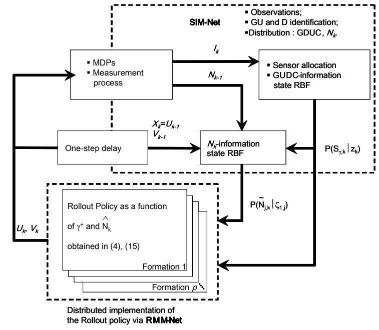
SIM-Net
▪ Observations;
▪ GU and D identification;
▪Distribution : GDUC, N_k
▪ MDPs
▪ Measurement process
I_k
N_{k-1}
▪ Sensor allocation
▪ GUDC-information state RBF
One-step delay
X_k=U_{k-1}
V_{k-1}
N_k-information state RBF
P(S$_{\gamma,k}$|z$_k$)
P($\tilde{N}_{j,k}$|$\zeta_{1,j}$)
Rollout Policy as a function of γ^+ and $\hat{N}_k$ obtained in (4), (15)
Formation 1
Formation p
U_k, V_k
Distributed implementation of the Rollout policy via RMM-Net

Fig. 2. Block diagram of proposed RBF-based rollout policy when N_k^{ν} cannot be communicated through RMM-Net

V. Distributed Computation of $\widetilde{W'}_{k,P_i,\gamma^+,\nu_j}$

Intermittent communication losses may also prevent from agreeing on a single value $\frac{1}{p}\sum_{j=1}^{p} \widetilde{W'}_{k,P_i,\gamma^+,\nu_j}(N_k, X_k, U_k, V_k)$ as mentioned in Section III. $\widetilde{W'}$ represents the cost-to-go function, which depends on the state of MDPs that model B and R [1]. $\widetilde{W'}$ is a linear combination of terms $m_p I(N_k^{\nu}|N_{k-1}^{\nu}) p_{i,j}^{\nu}$, where $p_{i,j}^{\nu}$ (Fig. 3) is the transition probability from N_k^{ν} to N_{k+1}^{ν} of the formation engaged along (i,j). $p_{i,j}^{\nu}$ depends on U_k^{ν} and $V_{ij,k}$.

First note that each $\nu \in B$ utilizes the same weighted graph such as that shown in Fig. 3. Then to avoid discrepancy about $\widetilde{W'}$, which is computed by the leader of each formation ν, an algorithm should be implemented in each ν such that the set of probabilities $\{p_{i,j}^{\nu},\ \nu \in B\}$, yields at t_k and at a given iteration of the Monte-Carlo simulation the same state transition, $N_k \rightarrow N_{k+1}$, regardless of the formation that computes this transition.

We propose that such an algorithm be composed of a uniform pseudorandom number generator (PRNG), such as the Mersenne twister [8], combined with the inverse transform sampling. The internal state s_o of PRNG must be the same in every $\nu \in B$ so that each formation provides by computation the same state transition, $N_k \rightarrow N_{k+1}$. Suppose, e.g., that, at t_k and at the ith iteration of the Monte-Carlo simulation, PRNG(s_o) provides in each formation the following pseudorandom number $s_k[i] \in [0,1]$. Then, $s_k[i]$ is compared to values that $\{p_{i,j}^{\nu},\ \nu \in B\}$ can adopt. Assume, e.g., that $P(N_{k+1}^{\nu} = N_k^{\nu}) = p_{0,i,j}^{\nu}$, $P(N_{k+1}^{\nu} = N_k^{\nu} - 1) = p_{1,i,j}^{\nu}$, ..., $P(N_{k+1}^{\nu} = N_k^{\nu} - d^{\nu}) = p_{d^{\nu},i,j}^{\nu}$. Then, every formation computes from $s_k[i]$ the state $N_{k+1} = \{N_{k+1}^{1}, ..., N_{k+1}^{\nu}, ..., N_{k+1}^{p}\}$ such that

$$N_{k+1}^{\nu} = \left\{ \begin{array}{lll} N_k^{\nu} & \text{if} & s_k[i] \in [0, p_{0,i,j}^{\nu}), \\ N_k^{\nu} - 1 & \text{if} & s_k[i] \in [p_{0,i,j}^{\nu}, p_{0,i,j}^{\nu} + p_{1,i,j}^{\nu}), \\ ... & & \\ N_k^{\nu} - d^{\nu} & \text{if} & s_k[i] \in [1 - p_{d^{\nu},i,j}^{\nu}, 1]. \end{array} \right. \tag{15}$$

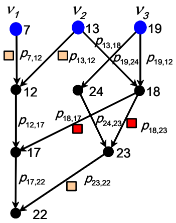

Fig. 3 Possible routes for formations ν_1,ν_2, and ν_3 from nodes (7,13,19) to 22. The weights found on the edges are conditional probabilities of MDPs.

VI. Numerical Examples

The information-state-estimator-based RMM strategy is applied to a group of three formations, which has to reach three targets sequentially in an adversarial environment. Each formation comprises three UCVs. The number of munitions per vehicle is 3 at onset of mission. It is assumed that up to three vehicles can be destroyed by a single ground unit, when in proximity. Numerical values for the MDPs are given in [5]. The theater (Fig. 4) is characterized by six uncertain ground unit classifications and location; i.e., there are three possibilities for each uncertainty: decoy, true ground unit or no ground unit. The closed-loop dynamics (Fig. 2) is therefore excited by an initial GULC estimation error since the computation of the base policy relies on a worst-case scenario, where a ground unit replaces each uncertainty in the imperfectly known GULC. The communication breakdowns occur within $[t_3, t_{3+\delta})$, when

B moves from the base to the first target.

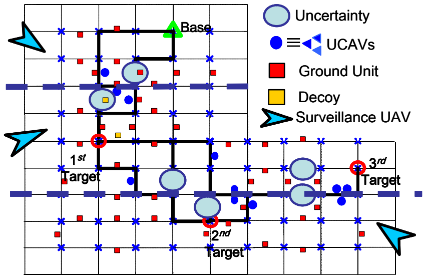

Fig. 4. Blue team evolution over the battlefield. Case of imperfect information (uncertainties).

Peformances are evaluated by carrying out 500 simulation runs. Five policies are tested. The lookup-table-based policy, denoted as PI [5], is applied to the ideal case of a perfectly known, time-invariant environment with RMM-Net exempt of communications breakdowns. The closed-loop rollout policy proposed in [1], denoted as PII, is implemented under the conditions of incomplete and partially known information on the GULC and breakdown-free RMM-Net communications. PIV corresponds to PII augmented with an estimator of N_k, which takes the last available data as the estimate; i.e., $\widehat{N}_k = N_{k-d}$ if communication breakdowns last d consecutive time intervals. The proposed dual-network health state estimator and decision policy (4)-(5), (6)-(10), and (12)-(13) is denoted as PIII and is simulated for the case of incomplete and imperfect information, and intermittent communication losses. The single-formation, single-route policy, denoted as PV, is simulated under imperfect information, although perfect communications. The results are given in Table 1. The first three columns indicate the average number ($\mathrm{a_{vr}}$) of munitions and the standard deviation ($\mathrm{s_{td}}$) on the number of munitions available at T_i. The last column gives the percent of total simulation runs for which no vehicle of B is capable of reaching T_3.

Table 1. Simulation results

	$\mathrm{T_1}$	$\mathrm{T_2}$	$\mathrm{T_3}$	$\mathrm{T_3}$ not
	$\mathrm{a_{vr}}$	$\mathrm{a_{vr}}$	$\mathrm{a_{vr}/s_{td}}$	reached
PI	17.8	12.8	9.7/4.4	1.8%
PII	17.7	11.9	8.5/4.3	4.6%
PIII	17.5	11.6	8.5/4.5	5.0%
PIV	17.5	11.4	7.5/3.9	8.5%
PV	13.6	7.7	3.8/3.9	28.4%

The number of remaining munitions at each target is greater when the multiformations of B are given the opportunity to divide and to aggregate (PI-PIV) as opposed to being constrained to follow a single route (PV). As shown in the rightmost column of Table 1, the third target is very likely to be reached by at least one vehicle when equipped with either PI, PII, PIII, or PIV. PI incurs the fewest number of losses. This makes sense as it is the optimal policy executed under ideal conditions of operation. The realistic scenario characterized by the combination of imperfect information about GULC and of intermittent communication losses, incurs a slight decrease of performance in PIII when compared to that obtained with PII. However, closing the loop with PIII provides superior performance than those obtained with PIV, which handles information loss in N_k by utilizing a d-step hold device, and with the single-formation, single-route policy PV. Finally, Fig. 5 shows the number of simulation runs, in percent of total number of runs, for which the number of munitions of B available at T_3 is greater than or equal to a prescribed threshold, denoted as t_N. For a given t_N, PIII allows reaching T_3 with a greater empirical frequency than that obtained with PIV and PV, and with an empirical frequency that is relatively close to that of PI and PII.

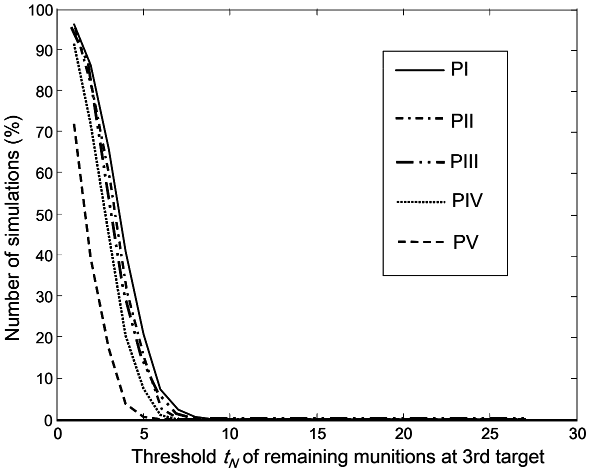

Fig. 5. Number of simulations yielding at T_3 a number of remaining munitions greater than or equal to t_N.

References

[1] N. Léchevin, C.A. Rabbath, and M. Lauzon, “A Networked Decision and Information System for Increased Agility in Teaming Unmanned Combat Vehicles,” in Proc. 46th IEEE Conf. on Decision and Control, New-Orleans, 2007.

[2] W.M. McEneaney, B.G Fitzpatrick, and I.G. Lauko, “Stochastic Game Approach to Air Operations,” IEEE Trans. Aerosp. Electron. Syst., Vol. 40, No. 4, pp. 1191-1216, 2004.

[3] C. Godsile and G. Royle, Algebraic Graph Theory. New York: Springer-Verlag, 2001.

[4] R. Olfati-Saber, J. Alex Fax, and R. Murray, “Consensus and Cooperation in Networked Multi-Agent Systems”, Proceedings of the IEEE, Vol. 95, No. 1, 2007.

[5] N. Léchevin, C.A. Rabbath, and M. Lauzon, “Cooperative and Deceptive Planning of Multiformations of Networked UCAVs in Adversarial Urban Environments,” AIAA Guidance, Navigation and Control Conf., Hilton Head, South Carolina, 2007.

[6] R. Mahler, “Objective functions for Bayesian ontrol-Theoretic Sensor Management, II: MHC-like Approximation,” in Recent Developments in Cooperative Control and Optimization, S. Butenko, R. Murphey, and P Pardalos (eds), Kluwer Academic, pp 273-316, 2004.

[7] L.F. Bertucelli and J.P. How, “Robust UAV Search for Environments with Imprecise Probability Maps,” in Proc. 44th IEEE Conf. on Decision and Control, Seville, Spain, 2005.

[8] M. Matsumoto and T. Nishimura, “Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator,” ACM Trans. on Modeling and Computer Simulation, Vol. 8, No. 1, pp.3-30, 1998.