Optimal index shooting policy for layered missile defense system

LI Longyue, FAN Chengli*, XING Qinghua, XU Hailong, and ZHAO Huizhen
Air and Missile Defense College, Air Force Engineering University, Xi’an 710051, China

Abstract: In order to cope with the increasing threat of the ballistic missile (BM) in a shorter reaction time, the shooting policy of the layered defense system needs to be optimized. The main decision-making problem of shooting optimization is how to choose the next BM which needs to be shot according to the previous engagements and results, thus maximizing the expected return of BMs killed or minimizing the cost of BMs penetration. Motivated by this, this study aims to determine an optimal shooting policy for a two-layer missile defense (TLMD) system. This paper considers a scenario in which the TLMD system wishes to shoot at a collection of BMs one at a time, and to maximize the return obtained from BMs killed before the system demise. To provide a policy analysis tool, this paper develops a general model for shooting decision-making, the shooting engagements can be described as a discounted reward Markov decision process. The index shooting policy is a strategy that can effectively balance the shooting returns and the risk that the defense mission fails, and the goal is to maximize the return obtained from BMs killed before the system demise. The numerical results show that the index policy is better than a range of competitors, especially the mean returns and the mean killing BM number.

Keywords: Gittins index, shooting policy, layered missile defense, multi-armed bandits problem, Markov decision process.

DOI: 10.21629/JSEE.2020.01.13

1. Introduction

These years, ballistic missile (BM) technology has spread to more and more countries. Nations all over the world are developing missiles capable of reaching enemies. One important mission of the strategic defense is to develop an integrated, layered ballistic missile defense (BMD) system to defend the homeland, deployed forces, allies, and friends from ballistic missiles attack [1]. The BMD system is based on a multilayer defense concept and therefore contains more than one defense weapon; it will include different types of defense weapons located on land or ships used to destroy BM [2]. A two-layer missile defense (TLMD) system includes exoatmospheric high-layer interceptors (EXOHI) and endoatmospheric low-layer interceptors (ENOLI), such as THAAD mid-range EXOHI and PAC-3 short-range ENOLI [3]. By now, the most difficult problem of terminal missile defense is that the time that can be used to intercept the target is very short. Short timelines not only stress the detection, tracking, and command and control of a defensive system, but also require a very efficient interceptor shooting policy [4]. In order to cope with the increasing threat of BMs in a shorter reaction time, we must optimize the shooting policy of the TLMD system, a well-designed shooting policy should be done. The aim of this study is to determine an optimal shooting policy for the TLMD system.

Research has been performed on system effectiveness, attack-defense game, framing analysis, decision analysis, reaction-diffusion differential system and interceptor’s allocation against incoming BMs, but very little research has been conducted on the interceptor’s shooting policy with complex multi-wave attack of BMs [5 – 16]. Gittins first proposed the Gittins index to calculate the optimal allocation over time of a single resource among a collection of bandits which are in competition for it. We call such problems multi-armed bandit (MAB) problems [17,18]. Gittins has proposed a series of calibration indices, which are now commonly referred to as Gittins indices; he showed that the index policy which always directs the key resource to the projects with the highest current index is optimal [19]. Agrawal considered MAB problems with switching cost, defined uniformly good allocation rules, and restricted attention to such rules [20]. The Gittins index require that passive projects should remain frozen inhibits applicability. Thus, Whittle introduced an important but intractable class of restless bandit problems which generalize the multi-armed bandit problems of Gittins by allowing state evolution for passive projects [21 – 23]. In recent
years, many researchers have made innovative research on
the theory of Gittins index and its application in the fields
of resource scheduling, task allocation and random de-
cision [24–28]. Glazebrook proposed a transformed index
which is used in procedures for policy evaluation [29].
Then, Glazebrook considered the optimal use of informa-
tion in shooting at a collection of targets, generally with
the object of maximizing the average number of targets killed
[30,31]. In [1,7], we presented a work for the modeling,
performance simulation, and channels optimal allocation
of the layered BMD $M/M/N$ queueing systems; and in
[2], a mathematical model and method for BMD intercep-
tor resource planning was presented.

All the literature mentioned above have provided some
inspiration for this thesis. In the process of missile defense
engagement, the probabilities of using EXOHI and ENOLI
are different, and the probabilities that EXOHI and ENOLI
kill a BM are different. Similarly, since the types of BMs
are different, their penetration probabilities and threat val-
ues are different. The TLMD system always attaches a
threat value to every BM it faces which could, for exam-
ple, reflect the damage which would be caused should
that BM penetrate its defenses [31–33]. The value of threat to
the defense area may vary according to the targets TLMD
chooses, that is, the threat value of BMs will be dynami-
cally changed after TLMD obtains more specific target in-
formation or knows that the target is intercepted. Thus, we
take TLMD’s goal to be maximizing the expected return of
BM s killed before the system demises and minimizing the
cost of BMs leakage, and then the improved index policy is
applied to the TLMD system. This paper is organized
as follows. Section 2 proposes a general model for TLMD
shooting decision processes. Section 3 illustrates the pre-
cise statements and proofs of the index shooting policy for
the TLMD system. Section 4 is the discussion and expan-
sion. Section 5 provides other shooting policies for com-
parison. Section 6 provides numerical examples.

2. General model for TLMD shooting
decision-making

Nash originally put forward a kind of Markov decision the-
ory in which at each decision epoch the decision-maker
must choose one of the $N$ arms [34]. Subsequently, this
theory was further studied and developed by Fay, Glaze-
brook, Crosbie, Dumitriu, Katta and others [35–39]. Now,
we consider the case of shoot-look-shoot, a missile defense
(MD) shooter has to plan a series of engagements with a
finite fixed collection of $N$ BMs. Define “one engagement”
at least including one shot of the MD system (during which
the MD system or defense area may be destroyed by BMs
penetration), and one look of targets intercepting results. It
is assumed that the number of interceptors is not limited;
then, the main decision-making problem of the shooting
optimization is how to choose the next BM target which
needs to be shot according to the previous shooting pro-
cess and results, thus maximizing the expected return of
BM s killed or minimizing the cost of BMs penetration. The
shooting engagements can be described as the following
discounted reward Markov decision process:

$$\{(\Omega_j, \omega_j, P_j, R_j, Q_j, \beta), \ 1 \leq j \leq N\}. $$

The state space of the BMs at time $t \in \mathbb{N}$ is
$X(t) = \{X_1(t), X_2(t), \ldots, X_N(t)\}$, where $X_j(t)$ denotes
the state of BM $j$ ($1 \leq j \leq N$). Let $\Omega_j$ be the cog-
nitive space (countable) of the TLMD system for all possible
states of BM $j$, then

$$X_j(t) \in \Omega_j \cup \{\omega_j\} \tag{1}$$

where $X_j(t) = \omega_j$ denotes that by time $t$, the TLMD
system or defense area is destroyed when shooting BM
$j$. $X_{N+1}(t) = 0$ indicates that TLMD chooses to with-
draw from engagement or be destroyed at time $t$; otherwise
$X_{N+1}(t) = 1$, that is,

$$X_{N+1}(t) = \begin{cases} 0, & \text{system disengaged or be destroyed at time } t, \\ 1, & \text{otherwise} \end{cases} \tag{2}$$

Obviously we have $X_{N+1}(0) = 1$. At each $t \in \mathbb{N}$, if
$X_j(t) \neq \omega_j$ and $X_{N+1}(t) = 1$ are satisfied, then TLMD
must choose to shoot, so that let $a_j$ represent the action of
shooting BM $j$ in the TLMD $(t+1)$th engagement, and
$a_{N+1}$ means that TLMD will no longer proceed with sub-
sequent shooting operations. The expected return achieved
by the TLMD system takes action $a_j$ at $t \in \mathbb{N}$ is

$$\beta^t R_j \{X_j(t)\} $$

where $R_j : \Omega_j \rightarrow \mathbb{R}^+$ (bounded and non-negative) may
reflect the damage caused by a still alive BM’s penetration
or its threat value to the defense area. The discount rate
$\beta \in (0, 1)$ is usually a constant, which is used to simplify
the loss of the expected return under certain subjective and
objective factors. Function $Q_j$ is represented as

$$Q_j(x) = \begin{cases} 1, & x \in \Omega_j, \\ 0, & x = \omega_j \end{cases} \tag{3}$$

$Q_j(x)$ indicates that if the TLMD system or its pro-
tected area is destroyed, the return of the shooting is 0; sim-
ilarly, the return of stop shooting when there is no shooting
opportunity for the BM is 0 [40,41]. That is, no returns are

gained beyond time

$$T = \inf\{t; X_j(t) = \omega_j, 1 \leq j \leq N; X_{N+1} = 0\}. \tag{4}$$
If the system executes action \( a_j \) at \( t \), the probability that BM changes from \( X_j(t) \) to state \( X_j(t+1) \) can be determined by Markov law \( P_j \):

\[
P_j(x, y) = P\{X_j(t+1) = y|X_j(t) = x\}, \quad x, y \in \Omega_j \cup \{\omega_j\}.
\]

Note that the state space \( \Omega_j \) should contain the state \( \omega_j \) (BM \( j \) is killed), so both \( \omega_j \) and \( \omega_j \) are the absorption states under \( P_j \), that is, to take any shooting action, all states are transferred to their own states with a probability of 1.

In order to describe the expected return of shooting actions, now we define bounded functions

\[
R_j, Q_j, \bar{R}_j \in \Omega_j \cup \{\omega_j\} \cap I,
\]

let \( \bar{R}_j = R_j Q_j \), the expected return of taking action \( a_j (1 \leq j \leq N) \) at time \( t \in \mathbb{N} \) is

\[
\beta^t R_j \{X_j(t)\} \left( \prod_{k \neq j} Q_k \{X_k(t)\} \right) X_{N+1}(t) = \\
\beta^t \bar{R}_j \{X_j(t)\} \left( \prod_{k \neq j} Q_k \{X_k(t)\} \right) X_{N+1}(t)
\]

where \( \bar{R}_j = R_j Q_j \). The product term \( Q_k \) in (6) can be understood that if the system or protected area is destroyed during the engagements, then the return is 0. By introducing the discount rate \( \beta \), (6) can simplify the loss of expected return under certain subjective and objective factors. The value of \( \beta \) is usually set between 0 and 1, which is determined by the decision-maker himself. The expected return of taking action \( a_{N+1} \) at time \( t \in \mathbb{N} \) is \( \beta^t R_d \), where \( R_d \) denotes the return of taking of disengagement action.

### 3. Index shooting policy for the TLMD system

A policy for the TLMD system is actually a rule based on historical shooting effects to decide the next shooting action. If the shooting policy is represented by \( v \), \( v(t) \) means that the action is taken at \( t \in \mathbb{N} \). Based on (6), the expected return of the total shooting process under the strategy \( v \) can be expressed as

\[
E_v = \sum_{t=0}^{\infty} \beta^t \bar{R}_{v(t)} \{X_{v(t)}(t)\} \cdot \left( \prod_{j \neq v(t)} Q_j \{X_j(t)\} \right) X_{N+1}(t).
\]

The purpose of this study is to find the optimal shooting policy \( v^* \), which can maximize the expected return. The model in Section 1 is a Markov decision model, which is called the generalized bandits decision process. The generalized bandits problem is a type of MAB problem in which the return is independent between different bandits. The multiplicatively separable form of the objective means that it can be used as the framework for the index shooting processes. From [41], we know that there exists an optimal policy for generalized bandits problems. Let \( X_j(t) \) be the state of BM \( j (1 \leq j \leq N) \) at \( t \), \( \tau \) is the stopping time of the shooting process, suppose that at \( t = 0 \) BM \( j \) is in state \( x \), \( \bar{R}_j(x, \tau) \) denotes expected returns during \([0, \tau]\), then

\[
\bar{R}_j(x, \tau) = E \left( \sum_{t=0}^{\tau-1} \beta^t \bar{R}_j \{X_j(t)\} | X_j(0) = x \right).
\]

When the TLMD system or its protected area is destroyed, its shooting returns immediately terminate, and the index \( \bar{G}_j(x, \tau) \) is defined as

\[
\bar{G}_j(x, \tau) = \frac{\bar{R}_j(x, \tau)}{1 - E(\beta^t Q_j \{X_j(t)\} | X_j(0) = x)}
\]

where \( 1 - E(\beta^t Q_j \{X_j(t)\}) X_j(0) = x \) is the probability that the TLMD system or its protected area will be destroyed by BMs during \([0, \tau]\), so (9) can be understood as

\[
\bar{G}_j(x, \tau) = \frac{\text{expected returns during } [0, \tau]}{\text{probability that the system be destroyed during } [0, \tau]}.
\]

the index \( G_j(x) \) takes the maximum value for \( \bar{G}_j(x, \tau) \), that is,

\[
G_j(x) = \sup_{\tau} \bar{G}_j(x, \tau), \quad x \in \Omega_j.
\]

From (11) we can see that the index shooting policy is a strategy that can effectively balance the shooting returns and the risk that defense mission fails, and the goal is to maximize the return obtained from BMs killed before the system demises.

Considering the limitations of previous studies, we now analyze the TLMD system shooting \( N \) BMs. Assuming that the BMs have \( B \) types, the “type” here may reflect any BM characteristic which is relevant to determining outcomes as the conflict proceeds, such as range, technical level, and so on. In actual operations, the TLMD system sometimes cannot completely determine the type of the incoming BM. This uncertainty can be expressed by \( N \) independent prior probabilities \( \prod_{1}^{N} \prod_{1}^{N} \prod_{1}^{B} \), and \( \prod_{1}^{B} \) denotes that the probability of the BM \( j \) belongs to the type \( b (1 \leq j < N, 1 \leq b < B) \). At every \( t \in \mathbb{N} \), the system selects one BM for shooting until the system is destroyed or all BMs are successfully intercepted. Assume that in one engagement, the type of BMs will not change,
all shooting outcomes are assumed independent, and the TLMD system has an accurate judgment on whether the BM is successfully intercepted. Let \( r_{Hb} \) denote the probability that the EXOHI system successfully intercepts a \( b \) type BM, \( \theta_{Hb} \) denotes the probability that the BM survives and the EXOHI system (or its protected area) is destroyed. Similarly, let \( r_{Lb} \) denote the probability that the ENOLI system successfully intercepts a \( b \) type BM, and \( \theta_{Lb} \) denote the probability that the BM survives and the ENOLI system (or its protected area) is destroyed. For all EXOHI and ENOLI systems, we let \( \phi_b \) denote the probability that the \( b \) type BM penetrates successfully, but due to its own guidance error, interference or other reasons fail to destroy the protection area.

The objective function of the TLMD system shooting is to maximize the return obtained from BMs killed before its own demise (when \( A = 0 \), the objective is to maximize the number of BMs killed), and the return obtained from successfully intercepting a \( b \) type BM is expressed as \( \beta L \). Based on Bayesian theory, if the system and BM \( j \) are alive after \( N \) round shots, then the posterior probability \( P_{n,N} \) of BM \( j \) belonging to \( b \) type can be expressed as

\[
\prod_{b}^{j,n} = \frac{\prod_{b}^{j}((1 - r_{Hb})^{x}(1 - \theta_{Hb})^{x})(1 - r_{Lb})^{n-x}(1 - \theta_{Lb})^{n-x}(1 - \phi_{b})^{n}}{\sum_{b=1}^{B}((1 - r_{Hb})^{x}(1 - \theta_{Hb})^{x})(1 - r_{Lb})^{n-x}(1 - \theta_{Lb})^{n-x}(1 - \phi_{b})^{n}} = \\
\prod_{b}^{j} ((1 - r_{Hb})(1 - \theta_{Hb}))^{x}((1 - r_{Lb})(1 - \theta_{Lb}))^{n-x}(1 - \phi_{b})^{n}
\]

(12)

where \( 1 \leq b \leq B, 0 \leq x \leq n \), \( x \) indicates the number of shots organized by the EXOHI system in all \( n \) shots. We already know that the probability that the BM changes from state \( X_{j}(t) \) to state \( X_{j}(t+1) \) can be determined by Markov law \( P_{j} \), since the type of BM is unknown, the calculation of (12) needs to be divided into four situations.

**Situation 1** Neither the TLMD system nor BM \( j \) is killed and the BM remains in the conflict, and the system will still need to continue to shoot, then

\[
P_{j}(n,n+1) = \frac{\sum_{b=1}^{B} \prod_{j}^{b}((1 - r_{Hb})(1 - \theta_{Hb}))^{x+1}((1 - r_{Lb})(1 - \theta_{Lb}))^{n+1-x}(1 - \phi_{b})^{n+1}}{\sum_{b=1}^{B} \prod_{j}^{b}((1 - r_{Hb})(1 - \theta_{Hb}))^{x}((1 - r_{Lb})(1 - \theta_{Lb}))^{n-x}(1 - \phi_{b})^{n}}.
\]

**Situation 2** BM \( j \) is killed. The system or its protected area is not destroyed. It is considered as follows:

(i) BM \( j \) is killed by the EXOHI system, then

\[
P_{j}(n,\overline{w}_{j})_{H} = \frac{\sum_{b=1}^{B} \prod_{j}^{b} r_{Hb}(1 - r_{Hb})^{y+1}(1 - r_{Lb})^{n+1-x}(1 - \theta_{Lb})^{n+1-x}(1 - \phi_{b})^{n}}{\sum_{b=1}^{B} \prod_{j}^{b} (1 - r_{Hb})(1 - \theta_{Hb}))^{x}((1 - r_{Lb})(1 - \theta_{Lb}))^{n-x}(1 - \phi_{b})^{n}}
\]

(ii) BM \( j \) is killed by the ENOLI system, then

\[
P_{j}(n,\overline{w}_{j})_{L} = \frac{\sum_{b=1}^{B} \prod_{j}^{b} r_{Lb}(1 - r_{Hb})^{y+1}(1 - r_{Lb})^{n+1-x}(1 - \theta_{Lb})^{n+1-x}(1 - \phi_{b})^{n}}{\sum_{b=1}^{B} \prod_{j}^{b} ((1 - r_{Hb})(1 - \theta_{Hb}))^{x}((1 - r_{Lb})(1 - \theta_{Lb}))^{n-x}(1 - \phi_{b})^{n}}
\]
**Situation 3** The system (or its protected area) is destroyed. It is considered as follows:

(i) The EXOHI system is destroyed, then

\[
P_j(n, \omega_j)_H = \frac{\sum_{b=1}^{B} \prod_{j} \theta_{Hb}(1-r_{Hb})(1-\theta_{Hb})^{x}((1-r_{Lb})(1-\theta_{Lb}))^{n-x}(1-\phi_b)^{n}}{\sum_{b=1}^{B} \prod_{j} ((1-r_{Hb})(1-\theta_{Hb}))^{x}((1-r_{Lb})(1-\theta_{Lb}))^{n-x}(1-\phi_b)^{n}}.
\]

(ii) The ENOLI system is destroyed, then

\[
P_j(n, \omega_j)_L = \frac{\sum_{b=1}^{B} \prod_{j} \theta_{Lb}(1-r_{Hb})(1-\theta_{Hb})^{x}((1-r_{Lb})(1-\theta_{Lb}))^{n-x}(1-\phi_b)^{n}}{\sum_{b=1}^{B} \prod_{j} ((1-r_{Hb})(1-\theta_{Hb}))^{x}((1-r_{Lb})(1-\theta_{Lb}))^{n-x}(1-\phi_b)^{n}}.
\]

**Situation 4** Neither the system nor BM \(j\) is killed (due to guidance error, interference or other reasons, BM \(j\) fails to destroy its target), then

\[
P_j(n, j) = \frac{\sum_{b=1}^{B} \prod_{j} \phi_b (1-\phi_b)^{n}((1-r_{Hb})(1-\theta_{Hb}))^{x+1}((1-r_{Lb})(1-\theta_{Lb}))^{n+1-x}}{\sum_{b=1}^{B} \prod_{j} ((1-r_{Hb})(1-\theta_{Hb}))^{x}((1-r_{Lb})(1-\theta_{Lb}))^{n-x}(1-\phi_b)^{n}}
\]

where \(x\) is the number of shots organized by the EXOHI system, and \(n-x\) is the number of shots organized by the ENOLI system.

The expected return gained from BM \(j\) killed in the engagement in four situations (without discount rate) above is given by

(i) BM \(j\) is successfully intercepted by the EXOHI system, then

\[
R_j(n)_H = \frac{\sum_{b=1}^{B} \prod_{j} R_b r_{Hb}(1-r_{Hb})^{x} (1-\theta_{Hb})^{x+1}(1-r_{Lb})(1-\theta_{Lb})^{n+1-x}(1-\phi_b)^{n}}{\sum_{b=1}^{B} \prod_{j} ((1-r_{Hb})(1-\theta_{Hb}))^{x}((1-r_{Lb})(1-\theta_{Lb}))^{n-x}(1-\phi_b)^{n}}.
\]

(ii) BM \(j\) is successfully intercepted by the ENOLI system, then

\[
R_j(n)_L = \frac{\sum_{b=1}^{B} \prod_{j} R_b r_{Lb}(1-r_{Lb})^{x+1}(1-\theta_{Hb})^{x}(1-r_{Lb})(1-\theta_{Lb})^{n-x}(1-\phi_b)^{n}}{\sum_{b=1}^{B} \prod_{j} ((1-r_{Hb})(1-\theta_{Hb}))^{x}((1-r_{Lb})(1-\theta_{Lb}))^{n-x}(1-\phi_b)^{n}}.
\]

The above analysis assumes that the EXOHI system has \(x\) shots, and the ENOLI system has \(n-x\) shots. If \(n\) is unknown, \(\xi_{Hb}\) can be used as the probability that the decisionmaker uses the EXOHI system to intercept the BM, then \(\xi_{Lb} = 1 - \xi_{Hb}\) is the probability of using the ENOLI system to intercept the BM, then we have

\[
R_j(n)_H = \frac{\sum_{b=1}^{B} \prod_{j} R_b (\xi_{Hb}r_{Hb}(1-r_{Hb})(1-\theta_{Hb}))^{n-\xi_{Hb}} + \xi_{Lb} r_{Lb}(1-r_{Lb})(1-\theta_{Lb})^{n-\xi_{Lb}}(1-\phi_b)^{n}}{\sum_{b=1}^{B} \prod_{j} ((1-r_{Hb})(1-\theta_{Hb}))^{x}((1-r_{Lb})(1-\theta_{Lb}))^{n-x}(1-\phi_b)^{n}}
\]  \[
R_j(n)_L = \frac{\sum_{b=1}^{B} \prod_{j} (1-r_{Hb})(1-\theta_{Hb})^{n-\xi_{Hb}}(1-r_{Lb})(1-\theta_{Lb})^{n-\xi_{Lb}}(1-\phi_b)^{n}}{\sum_{b=1}^{B} \prod_{j} ((1-r_{Hb})(1-\theta_{Hb}))^{x}((1-r_{Lb})(1-\theta_{Lb}))^{n-x}(1-\phi_b)^{n}}
\]
where \( n \in \mathbb{N} \). The probability \( \xi_{Hb} \) and \( \xi_{Lb} \) are introduced in (13), so \( x \) can be replaced by \( n \cdot \xi_{Hb} \). This replacement is not absolutely necessary. If we can accurately know the number of predetermined shots of EXOHI and ENOLI systems for the BM \( j \), it is not necessary to replace it; if we cannot know, we can set the probability of using different systems to intercept the target according to the conflict process. In order to enhance the generality of the model, we replace \( x \) with the probability \( \xi_{Hb} \).

It is known by (9) that the shooting process also requires a stopping time \( \tau_r \), in which \( r \) is a positive integer, and \( X_j(0) = n \). \( \tau_r \) indicates that after the system continues \( r \) shots for the BM \( j \), the system will stop shooting after one of the two is destroyed. Then \( \tilde{R}_j(n, \tau_r) \) and \( E(\beta_r Q_j(X_j(\tau_r)) | X_j(0) = n) \) are

\[
\tilde{R}_j(n, \tau_r) = \sum_{b=1}^{B} \left( \prod_{b=1}^{i} ((1 - r_{Hb})(1 - \theta_{Hb}))^{n \cdot \xi_{Hb}} ((1 - r_{Lb})(1 - \theta_{Lb}))^{n \cdot \xi_{Lb}} (1 - \phi_{b})^{n} \right) \cdot \left( \sum_{m=0}^{r-1} (\beta^m R_b(\xi_{Hb} r_{Hb} ((1 - r_{Hb})(1 - \theta_{Hb}))^{m \cdot \xi_{Hb}} + \xi_{Lb} r_{Lb} ((1 - r_{Lb})(1 - \theta_{Lb}))^{m \cdot \xi_{Lb}} (1 - \phi_{b})^{n}) \right) \cdot \left( \sum_{b=1}^{B} \prod_{b=1}^{j} ((1 - r_{Hb})(1 - \theta_{Hb}))^{n \cdot \xi_{Hb}} ((1 - r_{Lb})(1 - \theta_{Lb}))^{n \cdot \xi_{Lb}} (1 - \phi_{b})^{n} \right)^{-1} - 1 \right) \]

(14)

\[
E(\beta_r Q_j(X_j(\tau_r)) | X_j(0) = n) = \sum_{b=1}^{B} \left( \prod_{b=1}^{j} ((1 - r_{Hb})(1 - \theta_{Hb}))^{n \cdot \xi_{Hb}} ((1 - r_{Lb})(1 - \theta_{Lb}))^{n \cdot \xi_{Lb}} (1 - \phi_{b})^{n} \right) \cdot \left( \sum_{m=0}^{r-1} (\beta^{m+1} (1 - \phi_{b})^{m+1} (\xi_{Hb} r_{Hb} ((1 - r_{Hb})(1 - \theta_{Hb}))^{m \cdot \xi_{Hb}} (1 - \theta_{Hb})^{(m+1) \cdot \xi_{Hb}} + \xi_{Lb} r_{Lb} ((1 - r_{Lb})(1 - \theta_{Lb}))^{m \cdot \xi_{Lb}} (1 - \theta_{Lb})^{(m+1) \cdot \xi_{Lb}} + \beta^r (1 - \phi_{b})^{r} ((1 - r_{Hb})(1 - \theta_{Hb}))^{r \cdot \xi_{Hb}} ((1 - r_{Lb})(1 - \theta_{Lb}))^{r \cdot \xi_{Lb}} \right) \right) \cdot \left( \sum_{b=1}^{B} \prod_{b=1}^{j} ((1 - r_{Hb})(1 - \theta_{Hb}))^{n \cdot \xi_{Hb}} ((1 - r_{Lb})(1 - \theta_{Lb}))^{n \cdot \xi_{Hb}} (1 - \phi_{b})^{n} \right)^{-1} \right) \]

(15)

We express our conclusion in Theorem 1.

**Theorem 1** There exist functions \( G_j : \Omega_j \rightarrow \mathbb{R}^{+} \) such that, while the TLMD system or its protected area remains alive, that is, \( X_j^*(t) \neq \emptyset \), then the system optimally engages BM \( j^* \) if

\[
G_j^*(X_j^*(t)) = \max_j G_j(X_j(t))
\]

\[
\max_j G_j(X_j(t)) = G_j^*(n) = \max_n \left( \sum_{b=1}^{B} \left( \prod_{b=1}^{i} ((1 - r_{Hb})(1 - \theta_{Hb}))^{n \cdot \xi_{Hb}} ((1 - r_{Lb})(1 - \theta_{Lb}))^{n \cdot \xi_{Lb}} (1 - \phi_{b})^{n} \right) \cdot \left( \sum_{m=0}^{r-1} (\beta^m R_b(\xi_{Hb} r_{Hb} ((1 - r_{Hb})(1 - \theta_{Hb}))^{m \cdot \xi_{Hb}} + \xi_{Lb} r_{Lb} ((1 - r_{Lb})(1 - \theta_{Lb}))^{m \cdot \xi_{Lb}} (1 - \phi_{b})^{n}) \right) \cdot \left( \sum_{b=1}^{B} \prod_{b=1}^{j} ((1 - r_{Hb})(1 - \theta_{Hb}))^{n \cdot \xi_{Hb}} ((1 - r_{Lb})(1 - \theta_{Lb}))^{n \cdot \xi_{Lb}} (1 - \phi_{b})^{n} \right)^{-1} \right) \right) \right)
\]

(16)
where \( n \in \mathbb{N}^+, 1 \leq j \leq N, r \geq 1, 1 \leq b \leq B \).

4. Discussion and expansion

According to Theorem 1, the optimal shooting policy is

\[ H_j(n) = G_j(n)_{|r=1} = \]

\[
\sum_{b=1}^{B} \left( \prod_{b} ((1 - r_{Hb})(1 - \theta_{Hb}))^{n \xi_{Hb}} \cdot (1 - \theta_{Lb})^{n \xi_{Lb}} (1 - \phi_b)^n \right) \cdot (R_b(\xi_{Hb} r_{Hb} + \xi_{Lb} r_{Lb})) \\
\sum_{b=1}^{B} \left( \prod_{b} ((1 - r_{Hb})(1 - \theta_{Hb}))^{n \xi_{Hb}} \cdot (1 - \theta_{Lb})^{n \xi_{Lb}} (1 - \phi_b)^n \right) \cdot (1 - \beta(1 - \phi_b) + \beta(\xi_{Hb} \theta_{Hb} + \xi_{Lb} \theta_{Lb})) \]

(17)

\( H_j(n) = G_j(n)_{|r=1} \) is a “one-step index” defined by [41]. \( H_j(n) \) can be understood as a one-step index for shooting \( b \) type BMs, which is denoted by

\[
H_j(n) \propto \frac{R_b(\xi_{Hb} r_{Hb} + \xi_{Lb} r_{Lb})}{1 - \beta(1 - \phi_b) + \beta(\xi_{Hb} \theta_{Hb} + \xi_{Lb} \theta_{Lb})}. \quad (18)
\]

The BM type in (18) follows a posteriori probability distribution, and the value of \( H_j(n) \) is closely related to the values of \( R_b, \xi_{Hb}, \theta_{Hb}, r_{Hb}, r_{Lb}, \theta_{Lb}, \theta_{Hb}, \theta_{Lb} \) and \( \phi_b \). Table 1 lists several important correlation conclusions.

| Constant | Variable | Variables declaration | Variation tendency | \( H_j(n) \) |
|----------|----------|-----------------------|-------------------|-------------|
| \( \xi_{Hb}, r_{Hb}, r_{Lb}, \theta_{Hb}, \theta_{Lb}, \phi_b \) | \( R_b \) | BM threat values | / \| / |
| \( R_b, \xi_{Hb}, \phi_b \) | \( r_{Hb}, r_{Lb} \) | Kill probabilities | / \| / |
| \( R_b, \xi_{Hb}, \phi_b \) | \( \theta_{Hb}, \theta_{Lb} \) | Demise probabilities | / \| / |

Table 1 lists only a few correlation conclusions. There are seven variables in (18). The increase or decrease of a single variable does not necessarily lead to the increase or decrease of \( H_j(n) \), so we still need to calculate to determine the change of \( H_j(n) \). The probability \( \theta_{Hb}, \theta_{Lb} \) and \( r_{Hb}, r_{Lb} \) are not the opposite of each other in this paper, and there is no absolute correlation between them. For example, \( \theta_{Hb}, \theta_{Lb} \) may be reduced with measures such as electronic interference, camouflage protection and others. \( \phi_b \) denotes the probability that the \( b \) type BM penetrates successfully, but due to the fact that its own guidance error, interference or other reasons fail to destroy the protection area, in general, \( \phi_b \) can be a constant. Now we analyze the boundary conditions, if the function \( H_j(n) \) is monotonically decreasing, the maximum value of (17) is obtained for all \( n \) when \( r = 1 \), and at this time \( G_j(n) = H_j(n) \). This boundary condition indicates that the optimal shooting policy of the system is always selecting the BM with the highest index to shoot. If the function \( H_j(n) \) monotonically increases, then for all \( n \), when \( r \to \infty \) (17) gets the maximum value,

\[
G_j(n)_{|r=\infty} = \]

\[
\sum_{b=1}^{B} \left[ \left( \prod_{b} ((1 - r_{Hb})(1 - \theta_{Hb}))^{n \xi_{Hb}} \cdot (1 - \theta_{Lb})^{n \xi_{Lb}} (1 - \phi_b)^n \right) \cdot (R_b(\xi_{Hb} r_{Hb} + \xi_{Lb} r_{Lb})) \cdot (1 - \beta(1 - \phi_b) + \beta(\xi_{Hb} \theta_{Hb} + \xi_{Lb} \theta_{Lb})) \right]^{-1}. \quad (19)
\]

When \( r \to \infty \), this boundary condition indicates that the optimal shooting policy of the system is to shoot continuously at each BM until the BM is successfully intercepted or the system is destroyed. If there are only two types of targets, that is, \( B = 2 \), then \( H_j(n) \) must be a monotone function, so we have Lemma 1.

**Lemma 1** For all \( n \), the function \( H_j(n) \) must be monotonically increasing or decreasing when \( B = 2 \).

**Proof** If \( B = 2 \), then
since the value of $H_j(n)/H_j(n+1)$ does not depend on $j$ or $n$, then $H_j(n) \geq H_j(n+1)$ or $H_j(n) \leq H_j(n+1)$ for all $j$, that is, $H_j(n)$ is a monotonically increasing or decreasing function.

Lemma 1 shows that when there are two BM types, the optimal shooting policy is one of the following two policies. (i) The system always shifts from shooting type 1 BMs to type 2 BMs or from shooting type 2 BMs to type 1 BMs, that is the system shooting the BMs (which are still alive) in a numerical order. (ii) The system keeps shooting on each type of BMs until the BMs are successfully intercepted or the system is destroyed. The motivation for this behavior comes from an index increase or decrease, which considers the type 1 BMs or type 2 BMs which have a higher shooting index.

5. Other shooting policies

In order to compare with the TLMD index shooting policy

$$
\frac{H_j(n)}{H_j(n+1)} = \frac{(1 - r_{Hb})(1 - \theta_{Hb})}{(1 - r_{Lb})(1 - \theta_{Lb})} = \frac{\xi_{Hb}((1 - r_{Hb})(1 - \theta_{Hb}))}{\xi_{Lb}((1 - r_{Lb})(1 - \theta_{Lb}))} = \frac{\xi_{Hb}((1 - \phi_1)}{\xi_{Lb}((1 - \phi_2)}
$$

proposed in this paper, we give three other shooting policies for the system, which are the myopic policy, random policy and round-robin policy [41–46].

(i) Myopic policy

The myopic policy, which maximizes only immediate throughput, has been established as a simple and robust opportunistic spectrum access policy. If the principle of selecting shooting BMs by the index policy is to maximize the long-term return of shooting, then the principle of the myopic policy is to maximize the immediate return of shooting, that is, the myopic policy guides decisionmakers to shoot next at whichever BM is still alive and offers the decisionmaker the highest expected one-stage return. If the BM $j$ belongs to the $b$ type with a prior probability of $\frac{j}{B}$, and the system adopts the myopic shooting policy, the shooting returns are

$$(1 - r_{Hb})(1 - \theta_{Hb}))^{n - \xi_{Hb}} \cdot (1 - r_{Lb})(1 - \theta_{Lb}))^{n - \xi_{Lb}}(1 - \phi_1)^n \cdot (R_b(\xi_{Hb}r_{Hb} + \xi_{Lb}r_{Lb}))
$$

The myopic policy is not absolutely the optimal shooting policy. Suppose two interceptors intercept two BMs, the killing probabilities are 1, 0.9 and 0.9, 0 respectively and $R_1 = R_2 = 1$. If the myopic policy is adopted, the first interceptor should be used to intercept BM 1, and the second interceptor will not be used, with a total return of 1. If the index policy is used, the second interceptor should be used to intercept BM 1. If the interception fails, then the first interceptor is used to intercept BM 1 or BM 2, and the total return is $0.9 \times (1 + 0.9) + 0.1 \times (0 + 1) = 1.81$. Thus the myopic policy is not the optimal policy. The myopic policy is aimed at the maximum return of shooting at the present time, with relatively small calculation and high demand for real time BM data, but it does not consider the effect of subsequent shooting and the probability of the TLMD system being destroyed, so it is suitable for shooting the identical BMs. Compared with the index policy, the myopic policy can be called a sub-optimal policy.

It can be seen that for each type of BM, when the BM threat value is an invariant constant, the myopic policy always chooses the most easily killed BM ($r_{Hb}$ or $r_{Lb}$ is relatively large) to intercept, and obviously can maximize the number of BM killing.

(ii) Exhaustive policy

Here the TLMD system adopts the exhaustive shooting policy among those in which it shoots continuously at each BM until either party to the engagement is killed. This kind of shooting policy requires a simple sorting of the BM. For example, after the threat ranking of the BM, the BM can be shot in sequence from a large to a small threat value, that is, shoot at BMs in the order of descending values of the threats:

$$
\frac{\sum_{b=1}^{B} \left( \prod_{i=1}^{j} R_b(\xi_{Hb}r_{Hb} + \xi_{Lb}r_{Lb}) \cdot (1 - \beta)(1 - r_{Hb})^{\xi_{Hb}}(1 - r_{Lb})^{\xi_{Lb}}(1 - \theta_{Hb})^{\xi_{Hb}}(1 - \theta_{Lb})^{\xi_{Lb}}\right)^{-1}}{\sum_{b=1}^{B} \left( \prod_{i=1}^{j} (1 - \beta)(1 - \theta_{b})(1 - \theta_{Hb})^{\xi_{Hb}}(1 - \theta_{Lb})^{\xi_{Lb}}\right)^{-1}}
$$
(iii) Round-robin policy

In special situations, the system may make irrational shooting decisions, such as round-robin shooting, if there is a lack of external BM indication information or strong electronic interference measures. Round-robin is one of the policies employed by process and decisionmakers in computing. The name of the policy comes from the round-robin principle known from other fields, where each person takes an equal share of something in turn. The round-robin policy is simple, easy to implement, and starvation-free. The round-robin policy is to shoot the surviving targets in a certain order, such as the target number from small to large, the threat value from large to small, and so on. The first BM is chosen at random.

6. Numerical examples

In this section, we give examples to verify the optimality of the TLMD index shooting policy. Considering the interception of ten BMs \((N = 10)\), there are five BM types \((B = 5)\), and the discount rate \(\beta = 0.95\). The parameters of the two examples are shown in Table 2. As shown in Table 2, the higher the BM threat value is, the more difficult the BM is to be killed, the higher the probability of the TLMD system or its protected area will be destroyed.

| Example | \(b\) | \(\phi_b\) | \(R_b\) | \(r_{H_b}\) | \(r_{L_b}\) | \(\theta_{H_b}\) | \(\theta_{L_b}\) | \(\xi_{H_b}\) | \(\xi_{L_b}\) |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1       | 0   | 50  | 0.95| 0.80| 0.25| 0.10| 1/3 | 2/3 | 1/3 |
| 2       | 0   | 60  | 0.90| 0.60| 0.10| 0.10| 1/3 | 2/3 | 1/3 |
| 1       | 0   | 100 | 0.90| 0.90| 0.12| 0.06| 1/4 | 3/4 | 1/4 |
| 1       | 0   | 125 | 0.80| 0.80| 0.05| 0.05| 1/4 | 3/4 | 1/4 |
| 0       | 0   | 0   | 0.95| 0.75| 0.09| 0.13| 1/4 | 3/4 | 1/4 |
| 0       | 0   | 150 | 0.60| 0.20| 0.40| 0.40| 1/4 | 3/4 | 1/4 |
| 3       | 0   | 200 | 0.45| 0.20| 0.45| 0.25| 1/4 | 3/4 | 1/4 |
| 4       | 0   | 750 | 0.65| 0.45| 0.25| 0.05| 1/4 | 3/4 | 1/4 |
| 5       | 0   | 1000| 0.45| 0.25| 0.45| 0.25| 1/4 | 3/4 | 1/4 |

In contrast, we use four shooting policies to calculate two examples for \(10^4\) times. It is assumed that \(\prod_b^j\) obeys the \(U(0, 1)\) uniform distribution in each calculation, that is

\[
\sum_{b=1}^{5} \prod_{j}^j = 1, \quad 1 \leq j \leq 10.
\]

To simplify the calculation, we let \(R_d = 0\), indicating that the TLMD system does not stop shooting when the BM is alive and the system is not destroyed. \(\phi_b = 0\) \((1 \leq b \leq 5)\) indicates that if the interception of the \(b\) type BM fails, the penetration of the BM will destroy the system. Table 3 is the data summary of shooting returns, Table 4 is the data summary of the number of BMs killed, and Fig. 1 and Fig. 2 are comparisons of shooting returns and the mean number of BMs killed. In general, it is believed that a better shooting policy is based on the immediate states of the BM and defense system, and it should be the immediate optimal policy (myopic policy); the exhaustive and round-robin policies are poor, because these two policies do not take into account the return too much. In fact, the two examples show that the index policy is better than the other three shooting policies, especially the mean returns and the mean killing BM number, which is in agreement with the conclusions of Theorem 1.

| Example 1 | Policy | System | Minimum | Mean | Maximum |
|-----------|--------|--------|---------|------|---------|
| Index     | EXOHI  | 0.00   | 199.25  | 2 077.31 |
| ENOLI     | 0.00   | 99.62  | 1 038.65 |
| Myopic    | EXOHI  | 0.00   | 133.00  | 2 196.55 |
| ENOLI     | 0.00   | 66.50  | 1 098.27 |
| Exhaustive| EXOHI  | 0.00   | 164.84  | 2 202.44 |
| ENOLI     | 0.00   | 82.42  | 1 101.22 |
| Round-robin| EXOHI  | 0.00   | 155.27  | 2 238.81 |
| ENOLI     | 0.00   | 77.63  | 1 119.40 |

| Example 2 | Policy | System | Minimum | Mean | Maximum |
|-----------|--------|--------|---------|------|---------|
| Index     | EXOHI  | 0.00   | 844.88  | 9 678.44 |
| ENOLI     | 0.00   | 281.63 | 3 226.15 |
| Myopic    | EXOHI  | 0.00   | 710.64  | 11 132.63 |
| ENOLI     | 0.00   | 236.88 | 3 710.88 |
| Exhaustive| EXOHI  | 0.00   | 734.69  | 9 625.36 |
| ENOLI     | 0.00   | 244.90 | 3 208.45 |
| Round-robin| EXOHI  | 0.00   | 734.12  | 9 863.72 |
| ENOLI     | 0.00   | 244.71 | 3 287.91 |

| Example 1 | Policy | System | Minimum | Mean | Maximum |
|-----------|--------|--------|---------|------|---------|
| Index     | EXOHI  | 0.00   | 199.25  | 2 077.31 |
| ENOLI     | 0.00   | 99.62  | 1 038.65 |
| Myopic    | EXOHI  | 0.00   | 133.00  | 2 196.55 |
| ENOLI     | 0.00   | 66.50  | 1 098.27 |
| Exhaustive| EXOHI  | 0.00   | 164.84  | 2 202.44 |
| ENOLI     | 0.00   | 82.42  | 1 101.22 |
| Round-robin| EXOHI  | 0.00   | 155.27  | 2 238.81 |
| ENOLI     | 0.00   | 77.63  | 1 119.40 |

| Example 2 | Policy | System | Minimum | Mean | Maximum |
|-----------|--------|--------|---------|------|---------|
| Index     | EXOHI  | 0.00   | 3.74    | 7.50 |
| ENOLI     | 0.00   | 1.25   | 2.50    |
| Myopic    | EXOHI  | 0.00   | 1.69    | 7.50 |
| ENOLI     | 0.00   | 1.06   | 2.50    |
| Exhaustive| EXOHI  | 0.00   | 3.02    | 7.50 |
| ENOLI     | 0.00   | 1.01   | 2.50    |
| Round-robin| EXOHI  | 0.00   | 2.63    | 7.50 |
| ENOLI     | 0.00   | 1.08   | 2.50    |
Problems involving allocating limited resources in a limited time-frame under uncertainty may closely resemble the proposed shooting process. These problems could be non-defense related.

7. Conclusions

In order to cope with the increasing threat of BM in a shorter reaction time, a well-designed shooting policy should be done. A class of decision processes called MAB has been previously deployed to develop optimal policies for decisionmakers who attach a Gittins index to each target and optimally shoot next at the target with the largest index value. Unfortunately, for missile defense problems, the index policies are in general not proposed. The goal of this paper is to find a policy for shooting decision-making in TLMD interceptors planning. Our numerical results show that the index policy is better than a range of competitors, especially the mean returns and the mean killing BM number. Specifically, problems involving allocating limited resources in a limited time-frame under uncertainty may closely resemble the proposed index shooting process. We believe that the methodology can be applied to other non-defense related problems.

References

[1] LI L Y, LIU F X, LONG G Z, et al. Performance analysis and optimal allocation of layered defense M/M/N queueing systems. Mathematical Problems in Engineering, 2016, 5: 1 – 21.
[2] LI L Y, LIU F X, LONG G Z, et al. Modified particle swarm optimization for BMDS interceptor resource planning. Applied Intelligence, 2016, 44(3): 471 – 488.
[3] BRACKEN J. Layered defense of deceptively based ICBMs. Naval Research Logistics Quarterly, 1984, 31(4): 653 – 670.
[4] DAVIS M T, ROBBINS M J, LUNDAY B J. Approximate dynamic programming for missile defense interceptor fire control. European Journal of Operational Research, 2017, 259(3): 873 – 886.
[5] BLAHA G A, PENDERGRAFT T C, RILEY F A. Exploring architectural options for a space based missile defense layer. Proc. of the AIAA SPACE 2007 Conference and Exposition, 2007: 1 – 10.
[6] HAN C Y, LUNDAY B J, ROBBINS M J. A game theoretic model for the optimal location of integrated air defense system missile batteries. INFORMS Journal on Computing, 2017, 28(3): 405 – 416.
[7] LI L Y, LIU F X, LONG G Z, et al. Intercepts allocation for layered defense. Journal of Systems Engineering and Electronics, 2016, 27(3): 602 – 611.
[8] BROWN G, CARLYLE M, DIEHL D, et al. A two-sided optimization for theater ballistic missile defense. Operations Research, 2005, 53(5): 745 – 763.
[9] PARK S. Systematic analysis of framing bias in missile defense: implications toward visualization design. European Journal of Operational Research, 2007, 182(3): 1383 – 1398.
[10] ENDER T, LEURCK R F, WEAVER B, et al. Systems-of-systems analysis of ballistic missile defense architecture ef-
fectiveness through surrogate modeling and simulation. IEEE Systems Journal, 2010, 4(2): 156 – 166.

[11] LI LI Y, CAO J Z, MEI Y Y. Reduction and normal forms for a delayed reaction-diffusion system with B-T singularity. Advances in Difference Equations, 2019, 204: 1 – 15.

[12] KARASAKAL O. Air defense missile-target allocation models for a naval task group. Computers and Operations Research, 2008, 35(6): 1759 – 1770.

[13] PARNELL G S, METZGER R E, MERRICK J, et al. Multi-objective decision execution of theater missile defense architectures. Systems Engineering, 2015, 4(1): 24 – 34.

[14] BOARDMAN N T, LUNDAY B J, ROBBINS M J. Heterogeneous surface-to-air missile defense battery location: a game theoretic approach. Journal of Heuristics, 2017, 23(6): 1 – 31.

[15] GORFINKEL M. A decision-theory approach to missile defense. Operations Research, 1963, 11(2): 199 – 209.

[16] OGBU F A. Simulated annealing (SA) and optimization: modern algorithms with VLSI optimal design, and missile defense applications. Journal of the Operational Research Society, 1991, 42(1): 97 – 98.

[17] GITTINS J C. Bandit processes and dynamic allocation indices. Journal of the Royal Statistical Society, 1979, 41(2): 148 – 177.

[18] GITTINS J C. Multi-armed bandit allocation indices. New Jersey: Wiley, 1979.

[19] GLAZEBROOK K D, RUIZ-HERNANDEZ D, KIRKBRIDE C. Some indexable families of restless bandit problems. Advances in Applied Probability, 2006, 38(3): 643 – 672.

[20] AGRAWAL R, HEGDE M V, TENEKETZIS D. Asymptotically efficient adaptive allocation rules for the multiarmed bandit problem with switching cost. IEEE Trans. on Automatic Control, 1988, 33(10): 899 – 906.

[21] WHITTLE P. Multi-armed bandits and the Gittins index. Journal of the Royal Statistical Society, 1980, 42(2): 143 – 149.

[22] WHITTLE P. Restless bandits: activity allocation in a changing world. Journal of Applied Probability, 1988, 25(A): 287 – 298.

[23] WHITTLE P. Optimal control: basics and beyond. Academic Pediatrics, 1996, 10(5): 289 – 292.

[24] ANDERSON C M. Ambiguity aversion in multi-armed bandit problems. Theory and Decision, 2012, 72(1): 15 – 33.

[25] GU M Z, LU X W. The expected asymtptotic ratio for preemptive stochastic online problem. Theoretical Computer Science, 2013, 495: 96 – 112.

[26] SONIN I M. A generalized Gittins index for a Markov chain and its recursive calculation. Statistics & Probability Letters, 2008, 78(12): 1526 – 1533.

[27] KUMAR U D. Optimal selection of obsolescence mitigation strategies using a restless bandit model. European Journal of Operational Research, 2010, 200(1): 170 – 180.

[28] SI P, JI H, YU F R. Optimal network selection in heterogeneous wireless multimedia networks. Wireless Networks, 2010, 16(5): 1277 – 1288.

[29] GLAZEBROOK K D, GREATRIX S. On transforming an index for generalized bandit problems. Journal of Applied Probability, 1995, 32(1): 168 – 182.

[30] GLAZEBROOK K D, WASHBURN A. Shoot-look-shoot: a review and extension. Operations Research, 2004, 52(3): 454 – 463.

[31] GLAZEBROOK K D, KIRKBRIDE C, MITCHELL H M, et al. Index policies for shooting problems. Operations Research, 2007, 55(4): 769 – 781.

[32] LI LI Y, LIU F X, ZHAO L F, et al. Optimality analysis of index policy for offense-defense shooting process. Acta Armamentarii, 2015, 36(5): 953 – 960.

[33] GAYER D P, JACOBS P A. Suppression of enemy air defense (SEAD) as an information duel. Naval Research Logistics, 1998, 49(8): 723 – 742.

[34] NASH P. A generalized bandit problem. Journal of the Royal Statistical Society, 1980, 42(2): 165 – 169.

[35] FAY N A, WALRAND J C. On approximately optimal index strategies for generalized arm problems. Journal of Applied Probability, 1991, 28(3): 602 – 612.

[36] GLAZEBROOK K D, GREATRIX S. On transforming an index for generalized bandit problems. Journal of Applied Probability, 1995, 32(1): 168 – 182.

[37] CROSBIE J H, GLAZEBROOK K D. Evaluating policies for generalized bandits via a notion of duality. Journal of Applied Probability, 2000, 37(2): 540 – 546.

[38] CROSBIE J H, GLAZEBROOK K D. Index policies and a novel performance space structure for a class of generalized branching bandit problems. Mathematics of Operations Research, 2000, 25(2): 281 – 297.

[39] DUMITRIU I, TETALI P, WINKLER P. On playing golf with two balls. SIAM Journal on Discrete Mathematics, 2003, 16(4): 604 – 615.

[40] PUTERMAN M L. Markov decision processes: discrete stochastic dynamic programming. New Jersey: Wiley, 1994.

[41] GLAZEBROOK K D, KIRKBRIDE C, MITCHELL H M, et al. Index policies for shooting problems. Operations Research, 2007, 55(4): 760 – 781.

[42] WANG K H, CHEN L, LIU Q, et al. On optimality of myopic sensing policy with imperfect sensing in multi-channel opportunistic access. IEEE Trans. on Communications, 2013, 61(9): 3854 – 3862.

[43] CETINKAYA S, PARLARB M. Optimal myopic policy for a stochastic inventory problem with fixed and proportional backorder costs. European Journal of Operational Research, 1998, 110(1): 20 – 41.

[44] LEE Y. Modified myopic policy with collision avoidance for opportunistic spectrum access. Electronics Letters, 2010, 46(12): 871 – 872.

[45] LIU Z, TOWSLEY D. Optimality of the round-robin routing policy. Journal of Applied Probability, 1994, 31(2): 466 – 475.

[46] PANDAY D, VANDANA. Improved round robin policy mathematical approach. International Journal on Computer Science & Engineering, 2010, 2(4): 948 – 954.

Biographies

LI Longyue was born in 1988. He received his master and doctor degrees from Air Force Engineering University (AFEU) in 2012 and 2016, respectively. Now he is a lecturer in Air and Missile Defense College of AFEU. His research interests are modeling and simulation of missile defense operations, intelligent optimization algorithm, and operational research theory. He has published more than 40 papers, presided over a number of projects of the National Natural Science Foundation of China, Shaanxi Natural Science Foundation, China Post-doctoral Fund, etc.

E-mail: lilong_yue@126.com
FAN Chengli was born in 1988. She received her master and doctor degrees from Air Force Engineering University (AFEU) in 2011 and 2015, respectively. Now she is a lecturer in Air and Missile Defense College of AFEU. Her research interests are intelligent decision-making, intelligent evolutionary algorithms, and deep machine learning. E-mail: ffq516@163.com

XING Qinghua was born in 1964. She received her master and doctor degrees from Air Force Engineering University (AFEU) in 1992 and 2003, respectively. Now she is a professor in Air and Missile Defense College of AFEU. Her research interests are intelligent decision-making, intelligent evolutionary algorithms, and decision optimization theory and method. E-mail: qh_xing@126.com

XU Hailong was born in 1981. He received his doctor’s degree from Air Force Engineering University (AFEU) in 2010. Now he is a lecturer in Air and Missile Defense College of AFEU. His research interests include machine learning, information retrieval, datamining and image processing. E-mail: xhl_81329@163.com

ZHAO Huizhen was born in 1990. She received her doctor’s degree from Air Force Engineering University (AFEU) in 2018. Now he is a lecturer in Air and Missile Defense College of AFEU. Her research interests include deep machine learning, decision optimization theory and method. E-mail: margeryzhao@outlook.com