УДК 517.977; 519.832.3

МОДЕЛИ СТРАТЕГИЧЕСКОГО ПОВЕДЕНИЯ В ЗАДАЧЕ О ДИФФУЗНОЙ БОМБЕ[1]

В.О. Корепанов, Д.А. Новиков

Для задачи группового планирования траекторий в конфликтной среде (на примере задачи о диффузной бомбе) рассматривается ее расширение до теоретико-игровой постановки с активным обороняющимся субъектом.

Ключевые слова: задача о диффузной бомбе, стратегическое поведение, рефлексия, игра рангов.

ВВЕДЕНИЕ

Задачи планирования траектории (ЗПТ, Path Planning) *в конфликтной среде* или в условиях риска заключаются в построении оптимальной траектории *подвижного объекта* (ПО) на плоскости или в пространстве в условиях наличия группы *сенсоров* и активных средств обороны. Уровень риска для ПО определяется его положением в пространстве, а также его скоростью и ориентацией. Точные аналитические методы на сегодняшний день получены для случаев, когда имеется один ПО и число сенсоров не больше двух (см. ссылки в статье [1]). Для более сложных случаев обычно применяются приближенные решения с дискретизацией времени и пространства и эвристические методы дискретной оптимизации [2].

С одной стороны, нападающий имеет возможность выбора траекторий (или даже принципов поведения) ПО, с другой стороны, уровень риска зависит также от наличия у обороняющего средств обнаружения/уничтожения, их состава и расположения; т. е. противодействие нападения и обороны может описываться как антагонистическая игра (игра двух лиц с противоположными интересами).

Уместно вспомнить, что теоретико-игровой подход давно (с 1940-х гг.) и успешно применяется в моделировании военных действий (см. например, обзор в статье [3]), при этом отдельный класс задач, связанных с ракетным нападением и противоракетной обороной, появляется почти в то же время — см. обзор в работе [4]. Основной математический аппарат их исследования — конечные антагонистические игры и теория вероятностей. При этом обычно подразумевается информированность обороняющегося в подконтрольной ему области о ПО (например, ракете), а неопределенность касается количества и последовательности появления ПО, наличия ложных целей, поражающей силе ПО (например, зависимости накопленного ущерба от числа ПО, достигших цели) и т. п.

Затем появляются так называемые *задачи патрулирования*, когда упор делается на то, что первая и главная задача обороняющегося — это обнаружение ПО. В моделях данного типа в общем случае может быть неизвестно направление, маршрут, точка входа и время проникновения ПО на защищаемую территорию; ищутся непрерывные траектории патрулирования, обеспечивающие максимально возможный уровень обнаружения ПО в заданных условиях. Основной аппарат анализа — дифференциальные игры, игры поиска — см. ссылки в статье [3].

В настоящей работе исследуется «групповая» постановка задачи — противодействие группы нападающих ПО и группы сенсоров обороняющейся стороны (см. систему классификаций, примеры и обзор в статье [1]). При вычислении мгновенной вероятности обнаружения ПО в заданной точке учитываются расстояния до сенсоров и моделируется *эффект сосредоточения* — ПО служат «сенсорами» друг для друга (таким образом моделируется эффект увеличения вероятности обнаружения нескольких ПО в случае их сближения друг с другом благодаря увеличению эффективной площади отражения и ограничениям на разрешающую способность сенсоров).

Основная цель настоящей работы состоит в разработке методики построения моделей *стратегического поведения* (принятия решений с учетом возможных действий и реакций партнеров/оппонентов) в задачах групповой обороны-нападения

[1] Работа выполнена при частичной финансовой поддержке РФФИ, грант № 15-07-09048 А.

и применении этой методики для построения и анализа теоретико-игровой модели противоборства обороняющегося и нападающего в задаче о диффузной бомбе. Мы разбили такую методику на шаги, соответствующие разделам работы. В § 1 формулируется задача планирования траектории, которая отражает модель противодействия обороняющегося и нападающего, будем называть ее *микромоделью*. Формулировка микромодели есть базовый шаг в построении моделей стратегического поведения. Следующему шагу — построению имитационной модели противоборства, позволяющей исследовать микромодель и получать данные для построения моделей стратегического поведения (если микромодель не позволяет это делать аналитически) — посвящен § 2. Далее, в § 3 исследуются свойства построенной имитационной модели, подтверждаются или опровергаются желаемые свойства, разумность выбора параметров. В § 4 мы строим критерии эффективности стратегического противодействия для каждой из сторон. И наконец, в § 5 выбираются исследуемые стратегии, по результатам имитационного моделирования строится конкретная модель стратегического противодействия и проводится ее анализ подходящим математическим аппаратом.

1. ЗАДАЧА ПЛАНИРОВАНИЯ ТРАЕКТОРИИ

Пусть заданы начальные положения K подвижных объектов на плоскости: $(x_i(0), y_i(0))$, $i = \overline{1, K}$. Их цель — оказаться в точке с координатами (x^*, y^*). Положение i-го ПО в момент времени $t \geq 0$ обозначим через $(x_i(t), y_i(t))$, модуль его скорости — через $v_i(t) = \sqrt{(\dot{x}_i(t))^2 + (\dot{y}_i(t))^2}$, а время первого попадания в точку (x^*, y^*) — через T_i.

Также имеется N неподвижных сенсоров с координатами (a_j, b_j), $j = \overline{1, N}$. Расстояние от i-го ПО до j-го сенсора в момент времени t обозначим через $\rho_{ij}(t) = \sqrt{(x_i(t) - a_j)^2 + (y_i(t) - b_j)^2}$, расстояние между i-м и k-м ПО — через $r_{ik}(t) = \sqrt{(x_i(t) - x_k(t))^2 + (y_i(t) - y_k(t))^2}$. Наблюдения сенсоров независимы, задача сенсоров — обнаружение ПО. *Риск обнаружения* i-го ПО к моменту времени T_i в линейном приближении описывается функционалом [5]:

$$R_i = \int_0^{T_i} \sum_{j=1}^{N} C_j \frac{(v_i(t))^m}{(\rho_{ij}(t))^s} dt, \qquad (1)$$

где C_j — настроечный коэффициент сенсора (чувствительность j-го сенсора).

Показатель степени s характеризует физическое поле, в котором осуществляется обнаружение [5], а показатель степени m — зависимость уровня интенсивности излучаемого сигнала от скорости движения объекта (например, сигнала первичного гидроакустического поля [1]).

В данной работе моделируется также и эффект сосредоточения ПО, так что риск обнаружения i-го ПО определим как

$$R_i^* = \int_0^{T_i} \left[\sum_{j=1}^{N} C_j \frac{(v_i(t))^m}{(\rho_{ij}(t))^s} + \sum_{k \neq i} c_k \frac{1}{(r_{ik}(t))^s} \right] dt, \qquad (2)$$

т. е. ПО друг для друга служат сенсорами (c_k — «взаимная чувствительность»); этим мы учитываем увеличение вероятности обнаружения при сосредоточении ПО, при этом $\forall\ k, j\ c_k \ll C_j$.

В ЗПТ ищется оптимальная траектория движения одного ПО в конфликтной среде. Если риск обнаружения описывается функционалом (1), то задача планирования траектории для группы ПО распадается на поиск оптимальных траекторий для каждого ПО в отдельности. В случае же, когда функционал риска описывается выражением (2), ЗПТ для группы ПО не декомпозируется на набор простых задач, решаемых для каждого ПО независимо. Будем называть такую задачу *задачей группового планирования траекторий* (ЗГПТ).

2. ИМИТАЦИОННАЯ МОДЕЛЬ

Аналитическое решение ЗПТ получено только для случаев, когда имеется один ПО и один или два сенсора (см. обзор в работе [1]) для риска, определяемого выражением (1). Для большего количества сенсоров при риске (1) применяются приближенные алгоритмы поиска траекторий [2, 5, 6]. Поэтому в случае нескольких сенсоров и риска (2) можно пытаться применить численные методы поиска оптимальных траекторий, но при нескольких взаимодействующих ПО (ЗГПТ) даже это затруднительно, поскольку для этого случая нет аналитических результатов по свойствам оптимальных стратегий и оценкам оптимальности траекторий. Поэтому, как и в работе [1], построим имитационную модель для исследования ЗГПТ и принципов поведения ПО.

Отметим, что риск обнаружения (2) можно использовать как критерий минимизации при поиске оптимальной траектории, но для имитационного моделирования он не подходит, так как не является в точности вероятностью обнаружения, а является ненормированной кумулятивной величиной — не позволяет считать мгновенную вероят-

ность обнаружения. Тем не менее, исходя из вида риска обнаружения (2), определим вероятность p_i обнаружения i-го ПО группой сенсоров в момент времени t при наличии других ПО:

$$p_i = \min\left\{ \sum_{j=1}^{N} C_j \frac{(v_i(t))^m}{(\rho_{ij}(t))^s} + \sum_{k \neq i} c_k \frac{1}{(r_{ik}(t))^s};\ 1 \right\}, \quad (3)$$

где коэффициенты C_j и c_j в выражении (2) подобраны так, что $p_i \approx 1$ только в некоторой малой окрестности сенсоров или других ПО, и имеют такие размерности, что первое выражение под минимумом безразмерное.

Отметим, что исследование прикладных моделей («микромоделей» динамики и обнаружения ПО) не является целью данной работы, так как «игровая» модель может быть «надстроена» над любой моделью динамики обнаруживаемых объектов. Поэтому выражение для вероятности (3) получено простейшим преобразованием (2), так чтобы результат мог играть роль вероятности в вычислениях. Хотя, например, каждое отдельное слагаемое первой суммы под минимумом имеет смысл когда [6]: «отношение сигнал/помеха на входе наблюдателя мало в течение всего времени движения объекта по маршруту» (см. в статье [6] более приближенные к реальности выражения для мгновенных вероятностей обнаружения ПО группой независимых сенсоров). Аналогично, хотя риск (1) и отражает качественно риск обнаружения ПО в реальных приложениях, модель влияния сосредоточения ПО на риск обнаружения (2) построена из соображений простоты (конечно, при решении реальных практических задач целесообразно в явном виде анализировать отражательную способность группы находящихся поблизости друг от друга объектов).

Моделирование противодействия начинается с момента времени $t_0 = 0$, в который все ПО находятся в своих начальных точках: $(x_i(0), y_i(0)) = (x_i^0, y_i^0)$. Далее ПО начинают движение в соответствии с заложенными в них принципами поведения с постоянной по модулю скоростью [1]. Скорость ПО задается как 100 модельных единиц расстояния в 1 модельную единицу времени. Под *принципом поведения* подразумевается закон изменения скорости ПО (по абсолютной величине и направлению) в зависимости от локальной информации, полученной ПО: расположение соседних ПО, информация о риске обнаружения в некоторой окрестности своего положения или данные по расположению сенсоров обороняющегося. В зависимости от постановки задачи, ПО могут знать или не знать расположение сенсоров, их характеристики, знать или измерять текущий риск, вероятность обнаружения в некоторой окрестности своего расположения и т. п. [1, 0].

Далее, в каждый «такт» времени: t_0, $t_0 + \Delta t$, $t_0 + 2\Delta t$ и т. д. (где Δt — интервал времени обнаружения, 1/4 модельной единицы времени) обороняющаяся сторона производит попытку обнаружить каждый ПО: с учетом его текущего положения и скорости реализуется случайная величина «обнаружение» X_i: $P(X_i = 1) = p_i$ (см. выражение (3)), $P(X_i = 0) = 1 - p_i$. Если $X_i = 1$, то i-й ПО обнаружен и с вероятностью уничтожения $p_d \in [0; 1]$ (в рассматриваемом далее примере $p_d = 0{,}5$) он уничтожается (далее не участвует в моделировании). Обнаруженный, но не уничтоженный, ПО продолжает движение, также как и необнаруженные ПО. Моделирование останавливается, когда каждый ПО либо уничтожен, либо достиг цели. Число K^g дошедших до цели ПО — критерий эффективности стратегии нападающего, которая состоит в выборе принципов поведения ПО.

Имитационная модель реализовывалась на базе модели, построенной для работы [1] с помощью системы AnyLogic 6.4; параметры микромодели[2] и стратегий нападающего такие же, как в упомянутой работе; показатель степени $s = 1$.

3. «ДЛИНА» ТРАЕКТОРИИ И ВЕРОЯТНОСТЬ УНИЧТОЖЕНИЯ

В работе [1] при формулировке принципов поведения ПО использовалось понятие *критического уровня* δ вероятности обнаружения за один такт времени, т. е. ПО пытались обойти по кратчайшему пути так называемую *опасную область* — линию уровня δ, если же это не получалось без удаления от цели, то переключались в режим прямого движения к цели. В данном разделе рассматривается вопрос об эффективности такого поведения: нужно ли ПО обходить опасную область, и нужно ли им стремиться двигаться максимально близко к ее границе. Мы нашли некоторое подтверждение этого принципа поведения с помощью имитационного исследования влияния длины траектории и ее удаленности от сенсоров на вероятность достижения цели — *вероятность необнаружения на траектории*. В общем же случае, как предложено в работе [8], следует рассматривать многокритериальную задачу оптимизации с учетом вероятности обнаружения, длины траектории и т. д. (см. далее).

Пусть единственный ПО выбирает среди множества простейших принципов поведения, в рамках которых траектории являются дугами различ-

[2] Файл модели AnyLogic доступен по адресу https://github.com/moskvo/DiffuseBomb.

Рисунки к статье В.О. Корепанова, Д.А. Новикова

«МОДЕЛИ СТРАТЕГИЧЕСКОГО ПОВЕДЕНИЯ В ЗАДАЧЕ О ДИФФУЗНОЙ БОМБЕ»

(см. с. 38 – 44)

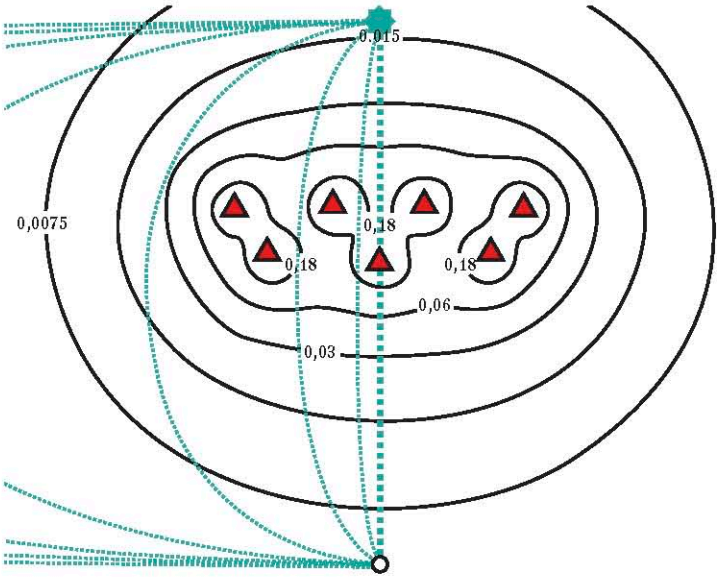

Рис. 1. Траектории движения ПО (пунктирные линии)

СТРАТЕГИИ ОБОРОНЫ

СТРАТЕГИИ НАПАДЕНИЯ

Прорыв	36,29	26,57	43,23	56,12	54,75	49,06	0,22
Обход	32,38	39,97	35,10	31,18	33,53	45,76	
Кооперация	41,47	45,22	39,76	37,13	48,06	57,81	
Рефлексия	38,85	43,16	38,34	35,26	46,82	54,03	
Адаптация	42,43	51,35	41,96	36,99	39,21	41,67	0,78
	0,76			0,24			

Рис. 4. Матричная игра обороны-нападения

РАНГ РЕФЛЕКСИИ ОБОРОНЯЮЩЕГОСЯ

РАНГ РЕФЛЕКСИИ НАПАДАЮЩЕГО

	0	1	2	3	4	
0	41,47	37,13	37,13	45,22	45,22	
1	42,43	36,99	36,99	51,35	51,35	0,78
2	36,29	56,12	56,12	26,57	26,57	0,22
3	36,29	56,12	56,12	26,57	26,57	
4	42,43	36,99	36,99	51,35	51,35	
	0,76	0,24				

Рис. 5. Игра рангов

ной длины от ее начальной точки до цели. Модуль скорости ПО постоянен. Расположение цели и сенсоров такие же, как в работе [1] (рис. 1, см. 3-ю страницу обложки). Пусть начальное расположение ПО есть (x^*, $y^* + \Delta$), где Δ — расстояние до цели, а сенсоры расположены симметрично относительно отрезка, соединяющего начальное положение ПО с целью, поэтому в силу симметрии можно рассматривать только траектории, огибающие опасную область с одной стороны (например, слева — см. рис. 1). Таким образом, ПО, фактически, выбирает только «длину» своей траектории, тем самым определяя и время движения по ней. Рассмотрим два различных типа полей, по сигналам в которых может быть обнаружен ПО, физически соответствующих гидроакустическому полю в мелком море ($s = 1$) и гидроакустическому полю в глубоком море в пассивном режиме ($s = 2$). На рис. 1 показаны: начальное положение ПО (внизу рисунка), положение цели (звездочка), расположение сенсоров (треугольники), возможные пути-дуги (зеленые пунктирные линии), линии уровня вероятности обнаружения (сплошные линии, соответствующие значениям δ: 0,18; 0,06; 0,03; 0,015 и 0,0075).

Результаты имитационного моделирования показаны на рис. 2. Для каждой траектории проводилось 50 тыс. запусков модели с целью получения усредненных данных. Длина траектории измеряется в модельных единицах.

Для исследуемых физических полей и траекторий зависимость вероятности необнаружения на траектории (вероятности достижения цели) от длины траектории имеет максимум, хотя для случая $s = 2$ он выражен не столь сильно. Точка максимума достигается при различных значениях длины пути ПО: для $s = 2$ «оптимальная» длина пути больше. Интересно, что «оптимальные» траектории обходили сенсоры и не удалялись слишком сильно от них, т. е. разумны стратегии ПО, заключающиеся в избегании попадания в опасную область.

На основании проведенного моделирования можно выдвинуть гипотезу, что, если путь оптимальной длины существует для случаев $s \geq 3$, то его длина будет тем больше, чем больше s; если же пути оптимальной конечной длины не существует, то выгодно как можно дальше обходить сенсоры.

Отметим, прежде всего, что в рассматриваемой модели мы анализировали вероятность обнаружения по каждому из полей отдельно; но ситуация может быть более сложной, когда обнаружение ведется одновременно в нескольких полях и необходимо соответственно пересчитывать вероятности необнаружения на траектории [8, 9]. Далее, отсутствовали ограничения на время достижения цели, длину траектории и т. п. В общем случае следует рассматривать несколько критериев — вероятность

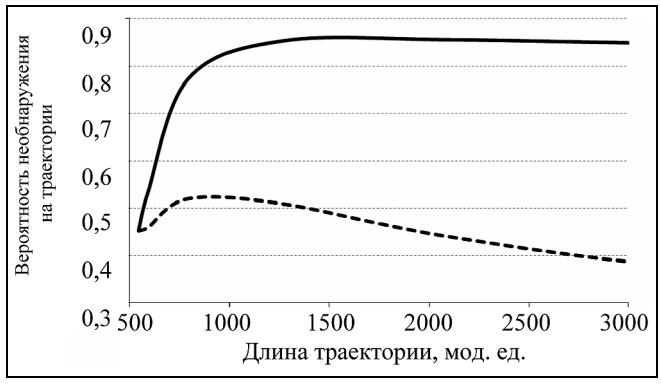

Рис. 2. Зависимость вероятности необнаружения на траектории от ее длины: ---- — $s = 1$; —— — $s = 2$

необнаружения на траектории, длина траектории, время движения и др., выбирая один из них в качестве экстремизируемого (на множестве Парето-эффективных комбинаций их значений — например, возрастающие участки кривых на рис. 2), а другие используя в качестве ограничений. И, наконец, перспективным направлением возможных дальнейших исследований можно считать построение и изучение моделей для исследования ЗГПТ (например, на основе дискретных марковских цепей), позволяющих получать аналитически характеристики различных траекторий.

4. КРИТЕРИИ ЭФФЕКТИВНОСТИ ДЛЯ ОБОРОНЯЮЩЕГОСЯ И НАПАДАЮЩЕГО В ЗГПТ

Будем рассматривать самый простой вариант критерия эффективности (K^g) — количество ПО, достигших цели. Обороняющаяся сторона старается это значение уменьшить, а нападающая — увеличить. В общем случае критерии эффективности для нападающего и обороняющегося могут различаться. Например, нападающему может быть важно, чтобы время достижения цели у всех неуничтоженных ПО было одинаково, или, наоборот, чтобы между моментами достижения цели был определенный промежуток; обороняющийся может стремиться к тому, чтобы каждое поражение цели произошло как можно позже, тем самым, например, увеличивая время ее функционирования и т. п.

5. ИГРА ОБОРОНЫ-НАПАДЕНИЯ

5.1. Стратегии игроков

В работе [1] исследованы пять различных принципов поведения ПО (в том числе с их стратегической рефлексией) в случае множества сенсоров, группы ПО и риске, описываемом выражением (2). В настоящей статье в рамках той же имитаци-

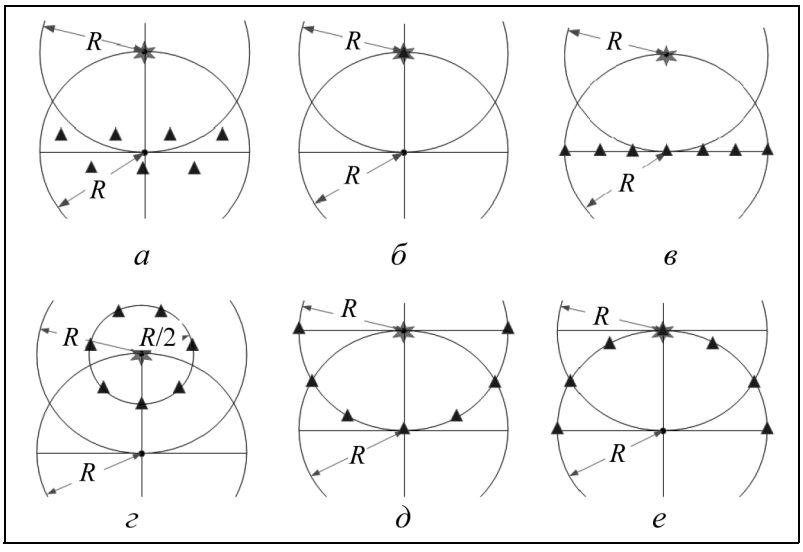

Рис. 3. Стратегии обороны: сенсоры расположены «в шахматном порядке» (*а*), сосредоточены в точке цели (*б*), расположены в линию (*в*), вокруг цели (*г*), полукругом (*д*) и «подковой» (*е*); звездочка — цель, треугольники — сенсоры, ПО движутся снизу

онной модели добавляется активность обороняющейся стороны.

Модель противодействия нападающего и обороняющегося сначала опишем как конечную антагонистическую (матричную) игру в нормальной форме. В игре участвуют две стороны (два игрока) — нападающий и обороняющийся. Начальное расположение группы ПО и положение цели заданы и являются общим знанием [10] для игроков (в рассматриваемом случае, так как в других вариантах может быть неопределенность и/или асимметричная информированность игроков относительно этих параметров).

Стратегия нападающего заключается в выборе принципов поведения ПО из конечного набора возможных их вариантов. Стратегия обороняющегося — в расстановке имеющихся сенсоров на плоскости (опять же, выбирается один из конечного числа вариантов). Доступные игрокам стратегии служат для них общим знанием. Свои стратегии игроки выбирают одновременно и независимо. Таким образом, обороняющийся изначально не знает принципов поведения группы ПО, а нападающий изначально не знает расположения сенсоров.

В работе [1] было предложено пять стратегий (принципов поведения) ПО: прорыв, обход, кооперация, рефлексивная и адаптивная стратегии. Стратегия же обороны была одна (рис. 3, *а*). Параметр R определяется по расположению сенсоров как расстояние от цели до прямой, разделяющей сенсоры на первую и вторую линии обороны. Рассмотрим шесть возможных стратегий обороны при неизменном числе сенсоров — 7 (рис. 3).

Конечность числа стратегий игроков постулируется априори. Понятно, что действий оппонентов на самом деле бесконечно много. Но, прежде всего, в реальности не все действия применимы. Далее, таким способом можно оценивать именно стратегии сторон как классы действий, качественно приводящие к однотипным ситуациям противоборства. Вероятно, отбор лучших стратегий можно проводить в духе метода ветвлений в дискретной оптимизации: сначала грубо разбить пространство действий на несколько классов, отобрать лучшие из них, затем их разбить на более мелкие, повторить отбор и т. д.

Итак, у каждого игрока имеется конечный набор стратегий (пять у нападающего, шесть — у обороняющегося), получается конечная антагонистическая игра в нормальной форме — матричная игра, в которой нападающий выбирает строку, а обороняющийся — столбец матрицы выигрышей. Для оценки выигрышей игроков (заполнения матрицы выигрышей) для каждой комбинации их стратегий нужно провести имитационное моделирование.

5.2. Матричная игра обороны-нападения

Для каждой из комбинаций шести стратегий обороны и пяти стратегий нападения осуществлялось 50 тыс. запусков имитационной модели с целью получения усредненных данных. В результате экспериментов получена 5×6 матрица выигрышей

$$A = \begin{pmatrix} 36{,}29 & 26{,}57 & 43{,}23 & 56{,}12 & 54{,}75 & 49{,}06 \\ 32{,}38 & 39{,}97 & 35{,}10 & 31{,}18 & 33{,}53 & 45{,}76 \\ 41{,}47 & 45{,}22 & 39{,}76 & 37{,}13 & 48{,}06 & 57{,}81 \\ 38{,}85 & 43{,}16 & 38{,}37 & 35{,}26 & 46{,}82 & 54{,}03 \\ 42{,}43 & 51{,}35 & 41{,}96 & 36{,}99 & 39{,}21 & 41{,}67 \end{pmatrix}.$$

Элементы этой матрицы — доли (в процентах) достигших цели ПО. Нападающий заинтересован в максимизации этого показателя, обороняющийся — в минимизации.

Для удобства восприятия и дальнейшего изложения представим матрицу A более наглядно (рис. 4, см. 3-ю страницу обложки). Числа справа и снизу матрицы соответствуют вероятностям стратегий в смешанном равновесии Нэша. Цветом выделены ситуации, возможные в равновесии Нэша, красным — наиболее вероятная ситуация, зеленым — наименее вероятная. Серые строки — до-

минируемые стратегии нападения. Обратим внимание, что у нападающего имеются две доминируемые чистые стратегии (соответствующие строки выделены серым цветом на рис. 4): «Обход» и «Рефлексия», которые доминируются стратегией «Кооперация». У обороняющегося доминируемых чистых стратегий нет.

В рассматриваемой игре нет равновесия Нэша в чистых стратегиях, равновесие Нэша в смешанных стратегиях единственно:

$$(x^N, y^N) = ((544/2527, 0, 0, 0, 1983/2527), (1913/2527, 0, 0, 614/2527, 0, 0)).$$

На рис. 4 (см. 3-ю страницу обложки) ситуации, входящие в носитель равновесия Нэша в смешанных стратегиях, выделены желтым, зеленым и красным цветом. Выигрыш нападающего в смешанном равновесии Нэша примерно равен 41,1. В равновесии Нэша нападающий должен применять стратегии «Прорыв» и «Адаптация» с вероятностями, примерно равными (1/5, 4/5), а обороняющийся — стратегии «Две линии сенсоров» (см. рис. 3, *а*) и «Сенсоры вокруг цели» (см. рис. 3, *г*) с вероятностями, примерно равными (4/5, 1/5).

Гарантирующие стратегии игроков — (3, 1) (стратегии — строки и столбцы - пронумерованы, начиная с 1). Выигрыш нападающего, если каждый из игроков выбрал свою гарантирующую стратегию, равен примерно 41,47.

Следует признать, что один из недостатков (типовой для конечных игр) данного подхода, в рамках которого анализируется конечное число априори зафиксированных стратегий нападающего и обороняющегося, состоит в сложности исследования «устойчивости» (чувствительности) получаемых результатов — например, матрица выигрышей и, следовательно, равновесные стратегии, могут в общем случае значительно изменяться при небольших вариациях множеств допустимых стратегий (например, при изменении координат расположения сенсоров, начальных координат и скоростей ПО и др.). Возможный выход заключается, прежде всего, в анализе диапазонов значений элементов матрицы выигрышей, при которых равновесие не изменяется, а также в попытке поиска соответствующих аналитических зависимостей.

5.3. Стратегическая рефлексия: игра рангов

Смешанное равновесие Нэша служит не очень «удобной» концепцией решения игры — если игровая ситуация разыгрывается только один раз, то действовать вероятностно представляется неразумным; может реализоваться «невыгодная» ситуация, например — (1, 1) и т. д. С другой стороны, выбор каждым из игроков своей гарантирующей стратегии не является равновесием Нэша. Поэтому исследуем полученную матричную игру (см. рис. 4) с точки зрения стратегической рефлексии, характерной для высших уровней иерархии принятия решений.

Стратегическая рефлексия — процесс и результат размышления игрока о принципах принятия решений оппонентами [10]. Рассмотрим «базовую» модель стратегической рефлексии (схема которой является общей [11]), опирающейся на понятие *ранг рефлексии*. Будем считать, что если игрок использует гарантирующую стратегию, то его ранг рефлексии равен 0. Если игрок считает, что его оппонент будет использовать свою гарантирующую стратегию, то его ранг рефлексии равен 1 и т. д. Игрок ранга k считает, что его оппонент имеет ранг рефлексии $(k - 1)$. Действие игрока ранга k: $x_i^k = BR_i(x_{-i}^{k-1}) = \arg\max_{k \in A_i} a_{kh}$, $i \in$ {«Н» — нападающий; «О» — обороняющийся}, где BR — наилучший ответ (Best Response), x_{-i}^{k-1} — действие оппонента ранга $(k - 1)$, $h = x_{-i}^{k-1}$, a_{kh} — элемент матрицы выигрышей A на пересечении k-й строки и h-го столбца, A_i — множество номеров стратегий игрока (в рассматриваемом примере задачи о диффузной бомбе $A_{\text{Н}} = \{1, ..., 5\}$, $A_{\text{О}} = \{1, ..., 6\}$).

Если игрок не знает точно ранг рефлексии оппонента, то ему можно выбрать, например, гарантирующий ранг рефлексии: $r_{\text{Н}} = \arg\max_{k = 0, 1, ...} \min_{h = 0, 1, ...} a_{kh}$; $r_{\text{О}} = \arg\min_{h = 0, 1, ...} \max_{k = 0, 1, ...} a_{kh}$.

В нашем случае гарантирующие ранги рефлексии соответствуют гарантирующим стратегиям игроков в исходной игре (в общем случае это не так), т. е.: $x_{\text{Н}}^{r_{\text{Н}}} = 3$; $y_{\text{О}}^{r_{\text{О}}} = 1$.

Если оба игрока осознают, что выбор ранга рефлексии является их стратегией, тогда возникает *игра рангов* [12]. В случае конечных игр, игра рангов, в которой игроки выбирают свои ранги рефлексии, также является конечной игрой; кроме того, в такой игре максимальный целесообразный ранг рефлексии ограничен [10]. Игра рангов, построенная для исследуемой матричной игры, представлена на рис. 5 (см. 3-ю страницу обложки). Числа справа и снизу матрицы соответствуют вероятностям стратегий в смешанном равновесии Нэша. Цветом выделены ситуации, возможные в равновесии Нэша, красным — наиболее вероятная ситуации, зеленым — наиболее вероятная. Серые строки и столбцы — повторяющиеся стратегии.

Серым цветом обозначены повторяющиеся строки и столбцы, которые можно исключить из рассмотрения. Равновесия Нэша в чистых стратегиях в рассматриваемой игре рангов нет (и быть не может для данного случая, см. работу [12]), равно-

весие Нэша в смешанных стратегиях в игре рангов единственно и, что интересно, совпадает с единственным равновесием Нэша в смешанных стратегиях для исходной матричной игры.

Таким образом, рассмотрение игры рангов позволило сократить размерность игры, отразить эффекты стратегического поведения, но не изменило равновесия. Тем не менее, тот интересный факт, что равновесие Нэша в смешанных стратегиях не изменилось при переходе от исходной матричной игры к соответствующей ей игре рангов, ставит вопрос о том, как вообще игра рангов может менять смешанные равновесия Нэша. Тем более, что для случая чистых стратегий ответ на этот вопрос уже получен [12].

ЗАКЛЮЧЕНИЕ

Можно сформулировать методику построения моделей стратегического поведения в задачах групповой обороны-нападения (приведем формулировку, находясь на позициях нападающего; с точки зрения обороняющегося это делается аналогично).

1. Перечислить возможные стратегии поведения нападающего (принципы поведения ПО могут включать в себя стратегии маскировки, использование ложных целей, отвлекающих маневров и т. п.) и обороняющегося.

2. Для каждой допустимой стратегии обороняющегося реализовать п. 3.

3. При фиксированной стратегии обороняющегося для каждой допустимой стратегии нападающего решить оптимизационную ЗПТ или в рамках имитационного моделирования решить ЗГПТ, проведя многокритериальный анализ и сравнение характеристик различных траекторий. Получить оценку выигрышей нападающего и обороняющегося.

4. Сформировать на основании результатов п. 2 и 3 и исследовать теоретико-игровую модель противодействия нападающего и обороняющегося.

5. На основании результатов п. 4 сформировать и исследовать модели, учитывающие стратегическую и/или информационную рефлексию нападающего и/или обороняющегося, включая задачи информационного противоборства [10].

Одна из возможных реализаций данной методики приведена выше для модели диффузной бомбы. В принципе же, число возможных вариантов невообразимо велико. С одной стороны, имея базовую теоретико-игровую модель (простейшую матричную игру), можно усложнять ее, применяя весь богатейший и разветвленный аппарат современной теории игр [13] (рассматривать игры в развернутой форме, повторяющиеся, иерархические и другие игры). С другой стороны, теоретико-игровые модели можно усложнять путем рассмотрения нетривиальной взаимной информированности игроков, наличия природной и/или целенаправленно создаваемой оппонентом неопределенности, стратегической рефлексии [10] и др.

Авторы признательны за ценные замечания и плодотворные обсуждения д-рам техн. наук А.А. Галяеву, В.Г. Лебедеву, Е.П. Маслову, И.М. Рудько и Е.Я. Рубиновичу.

ЛИТЕРАТУРА

1. *Корепанов В.О., Новиков Д.А.* Задача о диффузной бомбе // Проблемы управления. — 2011. — № 5. — С. 66—73.
2. *Zabarankin M., Uryasev S., Pardalos P.* Optimal Risk Path Algorithms / Cooperative Control and Optimization. — Dordrecht: Kluwer Acad., 2002. — P. 271—303.
3. *Новиков Д.А.* Иерархические модели военных действий // Управление большими системами: сборник трудов. — 2012. — № 37. — С. 25—62.
4. *Eckler A.R., Burr S.A.* Mathematical Models of Target Coverage and Missile Allocation. — Alexandria: Military Operations Research Society, 1972. — 254 p.
5. *Абрамянц Т.Г., Маслов Е.П., Яхно В.П.* Уклонение подвижного объекта от обнаружения группой наблюдателей // Проблемы управления. — 2010. — № 5. — С. 73—79.
6. *Абрамянц Т.Г., Маслов Е.П., Рудько И.М., Яхно В.П.* Уклонение подвижного объекта от обнаружения группой наблюдателей при малых отношениях сигнал/помеха // Информационно-управляющие системы. — 2011. — № 2. — С. 2—7.
7. *Абросимов В.К.* Групповое движение интеллектуальных летательных аппаратов в антагонистической среде: монография. — М.: Наука, 2013. — 168 с.
8. *Добровидов А.В., Кулида Е.Л., Рудько И.М.* Выбор траектории движения объекта в конфликтной среде // Проблемы управления. — 2011. — № 3. — С. 64—75.
9. *Добровидов А.В., Кулида Е.Л., Рудько И.М.* Оптимизация траектории движения объекта по вероятностному критерию в режиме пассивной гидролокации в анизотропной среде // Проблемы управления. — 2014. — № 4. — С. 31—37.
10. *Novikov D., Chkhartishvili A.* Reflexion and Control: Mathematical Models. — London: CRC Press, 2014. — 298 p.
11. *Новиков Д.А.* Модели стратегической рефлексии // Автоматика и телемеханика. — 2012. — № 1. — С. 3—23.
12. *Губанов Д.А., Чхартишвили А.Г.* О стратегической рефлексии в биматричных играх // Управление большими системами. — 2008. — № 21. — С. 49—57.
13. *Myerson R.* Game Theory: Analysis of Conflict. — Cambridge; Massachusetts; London: Harvard University Press, 2001. — 600 p.

Статья представлена к публикации членом редколлегии Е.Я. Рубиновичем.

Корепанов Всеволод Олегович — канд. техн. наук, ст. науч. сотрудник, ☎ (495) 334-90-51, ✉ vkorepanov@ipu.ru,

Новиков Дмитрий Александрович — чл.-корр. РАН, зам. директора, ☎ (495) 334-75-69, ✉ novikov@ipu.ru,

Институт проблем управления им. В.А. Трапезникова РАН, г. Москва.