

Теория, модели, инфраструктуры и языки спецификации командного поведения автономных агентов. Обзор (Часть 1)¹

Аннотация. Исследования в области автономных агентов и многоагентных систем в настоящее время во многом стимулируются потребностями современной робототехники, в которой одной из центральных задач является координация коллективного поведения в команде автономных роботов выполняющих некоторую миссию, которая должна выполняться без внешнего вмешательства. Именно этот аспект автономного поведения агентов является основным источником трудных проблем. Данная работа посвящена краткому обзору и анализу состояния исследований в рассматриваемой области. В частности, в работе описываются и анализируются известные результаты в области теории командной работы, в области моделей и предметно-независимых программных инфраструктур, поддерживающих процесс создания приложений, а также приводится сравнительный анализ существующих языков программирования, предназначенных для спецификации командных аспектов поведения агентов.

Ключевые слова: автономный агент, командная работа, теория командной работы, модель командной работы, инструментальное средство, язык спецификации.

Введение

Проблемы командного поведения автономных агентов и, в частности, командного поведения автономных роботов находятся в центре внимания исследователей и разработчиков в области искусственного интеллекта и робототехники, начиная с конца 1980-х годов, когда была опубликована фундаментальная работа [5]. Эта работа, в которой введены базовые понятия командной работы агентов, сформулированы и строго доказаны основные положения, о характере взаимодействия агентов команды и информационном обмене между ними, и по настоящее время во многом определяет направление исследований по этой проблеме.

Объектом исследований в ней является скоординированное командное поведение автономных агентов при совместном выполнении ими некоторой миссии. Поясним смысл терминов, использованных в предыдущем предложе-

нии. В зависимости от типа приложения содержание понятия миссии автономного агента, а также миссии команды агентов, в общем случае, может варьироваться. Однако наиболее часто под *миссией* автономного агента понимается решение им задачи, направленной на достижение одной или нескольких целей, при этом достижение цели или целей обычно отождествляется с достижением одного или нескольких целевых состояний агента.

Аналогично понимается миссия команды автономных агентов с тем отличием, что в этом случае она подразумевает достижение одной общей цели или нескольких таких целей (целевых состояний) совместными скоординированными усилиями членов команды. Именно такое понимание миссии агента используется в данной работе.

В основе понятия *автономности* агента лежит его способность самостоятельно анализировать ситуацию и на основании этого анализа

¹ Данная работа является инициативной и выполнена при частичной поддержке гранта № 10-07-00026 Российского фонда фундаментальных исследований.

выбирать и реализовывать стратегию поведения, которая приводит к наилучшему выполнению миссии, возложенной на него. Подчеркнем, что свойство автономности агента означает, что извне ему может только ставиться задача. Что касается выбора стратегии ее решения и выбора поведения, которое ведет к достижению цели, то это уже относится к области компетенции самого автономного агента.

В общем случае существуют различные ситуации, когда агенты должны координировать свое поведение. В некоторых случаях каждый агент имеет свою миссию и свои цели, однако он по каким-либо причинам не в состоянии решить задачу самостоятельно, а потому вынужден прибегать к помощи других агентов. Например, он может не обладать необходимой информацией и запрашивать ее у других агентов. Он может не обладать нужной функциональностью или нуждаться в ресурсах, которых у него нет. Этот случай взаимодействия агентов характерен тем, что у агентов нет никаких обязательств по отношению друг к другу, а потому любому агенту такого множества может быть отказано в помощи. Этот случай не относится к коллективному поведению агентов.

Обычно коллективное поведение агентов определяется дополнительными соглашениями, которые принято называть *взаимными обязательствами* (*commitments*). В зависимости от типа взаимных обязательств объединения агентов в группы принято разделять на три типа:

- *Альянс*—это объединение агентов в группу, в котором взаимные обязательства агентов группы относительно слабы. В этом случае агенты обязуются "помогать" друг другу, но не в ущерб собственным интересам.

- *Коалиция*—это объединение агентов, которые объединяются в группы с достаточно сильными взаимными обязательствами, например, чтобы помочь друг другу выстоять в конкурентной борьбе с другими агентами и/или коалициями. Агенты коалиции обычно имеют *свои цели и свои локальные показатели качества* их достижения, но они имеют также четко оговоренные взаимные обязательства, определяющие те ситуации или условия, в которых они обязаны помогать друг другу, а также договоренности о том, каким образом они должны это делать. Условия, при которых агенты коалиции прекращают помогать друг другу, также обычно четко оговариваются. Их приня-

то называть *соглашениями* (*conventions*). Примером коалиции является многоагентная система, управляющая деятельностью транспортных компаний, которые объединяются для получения выгодных контрактов в условиях конкурентной борьбы [18].

- *Команда*—это объединения агентов для совместного решения общей задачи (миссии), когда ни у одного агента команды не существует собственных *глобальных* целей, отличных от целей команды, а динамически формируемые локальные цели индивидуальных агентов направлены на достижение общих целей команды. Команда агентов действует как один агент. Примером такого объединения агентов являются команды роботов, играющих в футбол [23], группа военных вертолетов, выполняющих некоторую боевую задачу [26], команда агентов (посредников хакеров), которая выполняет распределенную скоординированную атаку на компьютерную сеть [8], и др.

Данная работа посвящена анализу состояния исследований в области командной работы автономных агентов. Напомним, что основная особенность командной работы агентов при выполнении некоторой миссии состоит в том, что эта миссия выполняется без внешнего вмешательства. Команда агентов должна самостоятельно решать такие задачи, как построение плана выполнения миссии путем декомпозиции глобальной цели на частные подцели и распределение частных задач по их достижению между агентами команды с учетом их возможностей и способностей по решению тех или иных задач. Команда должна выполнять мониторинг процесса работы и отслеживание зависимостей между задачами, решаемыми различными членами команды и обеспечивать их синхронизацию, своевременно предсказывать и разрешать конфликты, нарушения ограничений и обрабатывать нештатные ситуации. Все эти задачи должны решаться в реальном времени и в распределенном варианте, а информация, которой обмениваются агенты команды, должна постоянно поддерживать *самоорганизующийся* стиль исполнения командной работы по достижению общей цели. Примером такой самоорганизации команды агентов является поведение агентов при выполнении общего плана, предусматривающего, например, динамическое распределение задач, решаемых агентами в процессе выполнения миссии, когда всегда находится такой

агент команды, который определит необходимость выполнения очередного действия и выполнит его [25].

Можно привести много практически важных задач, которые в настоящее время могут быть эффективно решены только с использованием модели командной работы автономных физических и программных агентов. Примерами являются командная работа автономных роботов, охраняющих подводное пространство порта, роботов военного назначения, решающих совместно различные задачи на театре военных действий, командная работа беспилотных летательных аппаратов распределенного наблюдения за земной поверхностью, различные задачи противоборства в киберпространстве и другие. Заметим также, что модель командной работы автономных агентов рассматривается как перспективная модель различных автономных операций типа гуманитарных, военных, спасательных и т.п.

Сложности практической реализации идей командной работы автономных агентов, например, в приложениях упомянутого выше типа обусловлены многими факторами, которые являются общими практически для любого приложения в области командной работы автономных агентов. Перечислим их.

1. Непредсказуемая динамика внешней среды, сознательное противодействие соперничающей стороны (если она имеется).

2. Ограниченность знаний каждого агента о состоянии внешней среды, о команде противника и о других агентах команды, что влечет необходимость распределенной оценки текущей ситуации.

3. Многообразие вариантов сценария достижения одной и той же цели и вариантов распределения ролей между автономными агентами, зависимость множества возможных сценариев от текущего состояния и предыстории командной работы, от наличия ресурсов и от специализации агентов. Это обуславливает необходимость динамической оценки ситуации, динамического выбора продолжения сценария и распределения ролей в реальном времени.

4. Динамика сетевой организации команды агентов, когда коммуникационные средства, которыми располагают члены команды для поддержания взаимодействия, имеют ограниченную дистанцию достижимости, что обуславливает необходимость использования рас-

пределенной инфраструктуры для поддержки парных взаимодействия агентов и соответствующих коммуникационных протоколов.

Центральными задачами в области командной работы были и остаются задачи динамического планирования и перепланирования многошагового процесса выполнения миссии команды, динамического распределения ролей и подзадач между ее членами в зависимости от текущей ситуации, а также задачи координации поведения агентов команды в процессе исполнения миссии. Эти задачи чрезвычайно сложны, и, несмотря на большие усилия исследователей, результаты, достигнутые здесь, они нуждаются в дальнейшем развитии. Критический обзор таких результатов составляет основное содержание данной работы.

Далее раздел 2 представляет краткий обзор основных результатов в области теории командной работы автономных агентов. В разделе 3 (часть 2) рассматриваются существующие модели командной работы и программные инфраструктуры, которые благодаря наличию в них компонент повторного использования помогают существенно облегчить и ускорить разработку приложений в области командной работы автономных агентов. В этих разделах рассматриваются только предметно независимые аспекты командной работы. В отличие от этого, раздел 4 (часть 2) посвящен анализу состояния исследований в области языков спецификации приложений, в частности, языков спецификации роботов. В этом случае роботы являются исполнительными устройствами, которые исполняют поведение, задаваемое автономными агентами команды. В заключении по работе резюмируются ее цели.

1. Теории командной работы

Про агентов команды, которые прилагают совместные усилия (сотрудничают) для достижения общей *конечной* (*persistent*) цели, функционируют в динамической внешней среде в присутствии помех и противодействия со стороны соперника (или команды соперников) принято говорить, что они образуют *команду агентов*. Основная проблема, решаемая при организации командной работы автономных агентов, состоит в том, чтобы обеспечить работу агентов как единой команды в ситуации, когда каждый агент имеет только ограниченную ин-

формацию о собственной команде, о сопернике и о внешней среде и реализует *собственные намерения* с помощью *индивидуальных действий*. Эти действия могут выполняться параллельно или последовательно с действиями других агентов.

К настоящему времени существует несколько теорий командной работы автономных агентов. Одна из них, известная под названием "Теория общих намерений" предложена в фундаментальной работе [5]. Она формулирует, главным образом, базовые понятия и общие рамки, определяющие командное поведение, характер взаимодействия и информационного обмена членов команды. Вторая теория предложена в работах [9], [10]. Она известна под названием "Теория общих планов". Некоторые базовые положения, связанные с понятиями общих намерений, общих обязательств и соглашений были сформулированы в работах [12] и [13]. Рассмотрим кратко основные идеи обеих теорий.

1.1. Теория общих намерений

Теория общих намерений оперирует двумя типами понятий. К первому типу относятся так называемые *ментальные* понятия, в число которых входят событие, убеждение, цели и взаимные убеждения. *События* обычно связаны с конкретными агентами или объектами внешнего мира и отражают то, что с ними происходит в реальном времени. Предполагается, что возможные миры, в которых существуют и действуют агенты, состоят из примитивных событий, имеющих в качестве атрибутов тип события и имя агента, генерирующего событие или агента, с которым они ассоциированы. *Убеждения*—это знания агента о внешнем мире. Они представляют собой утверждения, в истинности которых во всех достижимых мирах агент убежден (верит) в текущий момент времени. Убеждения агента могут изменяться во времени и могут быть неверными. *Цели*—это некоторые состояния агента, которые он стремится достичь. Они формально представляются в виде утверждений, которые истинны в целевых состояниях [5]. Обычно понятия желаемого состояния и цели можно рассматривать как эквивалентные. *Взаимные убеждения* складываются из убеждений группы агентов и формализуются как конъюнкции утверждений, отвечающих убеждениям отдельных агентов группы. По-

скольку различные агенты могут иметь различные убеждения, т.к. они используют различные источники информации, то взаимные убеждения — это та часть индивидуальных убеждений, в которых сходятся все агенты команды.

Понятия второго типа используются в теории общих намерений для описания *поведения* команды агентов. К ним относятся индивидуальные обязательства и соглашения агента, его индивидуальные намерения, а также общие (коллективные) обязательства и намерения команды агентов. Поясним их содержание.

Определение 1. (Индивидуальные обязательства) Индивидуальные обязательства агента называют также его индивидуальной *конечной (persistent) целью*. Говорят [5] что агент имеет индивидуальную конечную цель достичь p при условии q (p и q пропозициональные утверждения) тогда и только тогда, когда

1. агент убежден (на самом деле это может быть не так), что в текущем состоянии p ложно (цель не достигнута);
2. агент стремится к тому, чтобы цель p в конечном счете стала истинной;
3. агент должен придерживаться этой цели в соответствии с соглашениями.

Соглашениями (conventions) называются условия, при истинности которых агент может отказаться от своих индивидуальных обязательств, выраженных в терминах индивидуальной конечной цели. Агент обязан придерживаться этой цели (не может отказаться от нее) до тех пор, пока не возникнут условия, предусматривающие отказ от нее в соответствии с соглашениями. В частности, агент придерживается этой цели (своих индивидуальных обязательств) до тех пор, пока он не придет к убеждению, что или *цель p истинна* (достигнута), или *она никогда не может быть достигнута*, или цель нерелевантна, когда *условие q перестало быть истинным* (стало ложным). Пример недостижимости и нерелевантности цели демонстрируют следующие ситуации. Пусть команда подводных роботов решает задачу обеспечения безопасности морского порта путем патрулирования подводного пространства. Индивидуальная цель p подводного робота, обнаружившего неизвестный объект, состоит в том, чтобы сначала идентифицировать его, а потом действовать, так или иначе, в зависимости от результата. Для идентификации объекта подводный робот должен приблизиться к нему

на некоторое расстояние. Если оказывается, что неизвестный объект является подвижным, и его скорость превышает максимальную скорость, которую может развить обнаруживший его подводный робот, то индивидуальная цель p для него недостижима. Примером нерелевантности цели является ситуация, когда подводный робот преследует неизвестный подвижный подводный объект в собственной зоне патрулирования, но этот объект в какой-то момент времени вышел в нейтральные воды. Тогда цель, состоящая в том, чтобы идентифицировать этот объект, становится нерелевантной.

Отметим, что при этом другие цели агента, а также его обязательства по отношению к другим агентам должны быть непротиворечивыми с его индивидуальной конечной целью.

Определение 2 (Индивидуальные намерения агентов). Говорят, что агент *намерен* выполнить в определенных условиях некоторое действие, если он в этих условиях в соответствии с конечной целью должен выполнить это действие, и при этом, выполняя его, он должен быть абсолютно уверенным, что выполняет именно принятое намерение.

Индивидуальные намерения также сохраняют все свойства индивидуальных обязательств в части непротиворечивости с убеждениями агента и его другими целями. Например, его намерения не должны быть такими, которые влекут действия, которые могут стать причиной возникновения нерелевантности или невозможности достижения целей другими агентами.

Понятие *общих обязательств* (и намерений) должно определяться так, чтобы оно описывало действия *группы агентов, как одного агента*. Очевидно, что для понятий, связанных с командной работой, простое объединение обязательств и намерений отдельных агентов не отражает существа и особенностей командной работы агентов, так как различные агенты на основе своей локальной информации могут расходиться в своих убеждениях относительно истинности p и q . И это естественно, поскольку команда выполняет распределенный алгоритм, а потому этот алгоритм должен еще включать в себя еще и *протокол взаимодействия* членов команды агентов. Поэтому при определении понятий, связанных с коллективной работой команды агентов, необходимо прямо или косвенно вводить компоненты, отражающие специфику такого протокола взаимодействия.

Подчеркнем, что протокол взаимодействия агентов команды играет исключительной важную роль в общем алгоритме взаимодействия. Понятие *общих обязательств* определяется через понятие "слабой конечной цели" агента.

Определение 3. Говорят, что агент имеет *слабую конечную цель* p при условиях q , если выполняется любое из нижеследующих условий:

1. Агент сохраняет стандартную конечную цель p , которой он стремится достичь, т.е. у агента нет убеждения, что цель достигнута, и значит, он полагает, что цель p продолжает оставаться конечной целью команды;

2. Агент убежден, что цель или достигнута (утверждение p истинно), или никогда не будет достигнута, или она нерелевантная (утверждение q ложно), и при этом он имеет в качестве текущей подцели сделать эту информацию о статусе цели *общим убеждением* членов команды.

Отметим, что отличие *слабой конечной цели* агента от его *конечной* цели состоит в том, что агент ее сохраняет даже тогда, когда она уже или достигнута, или она нерелевантная или недостижима, т.е. тогда, когда стремиться к ее выполнению ему уже не нужно. Но он ее сохраняет до тех пор, пока он не доведет текущую информацию о статусе цели до всех членов команды, и потому он стремится сделать эту информацию доступной для всех агентов команды. Конечно, на практике возможен также случай, когда он не в состоянии довести до членов команды информацию о статусе цели, например, ввиду отсутствия с ними связи, однако этот случай в теории общих намерений не рассматривается.

С учетом сказанного определение *общей (объединенной) конечной цели* (общих обязательств) формулируется нижеследующим образом [5].

Определение 4 (Общие обязательства). Общие обязательства команды агентов отождествляются с ее общей конечной целью. *Команда агентов* имеет *общую конечную цель* достигнуть p при условии q в том случае, если выполнены все нижеследующие условия:

1. *Все члены команды* убеждены, что выражение p в текущем состоянии ложно (т.е. конечная цель не достигнута).

2. Все члены команды знают, что *все* они хотят, в конце концов, сделать выражение p истинным (т.е. достигнуть цели).

3. Справедливо, и все члены команды это знают, что до тех пор, пока *все* они не придут к

общему убеждению, что или выражение p истинно, или что выражение p никогда не будет истинным, или что q ложно (цель нерелевантная), они будут убеждены, что p является слабой конечной целью в условиях q для всех членов команды.

Поясним это определение. Все члены команды с самого начала имеют общее обязательство достичь в конце концов цели p . Однако со временем члены команды (ввиду возможного отсутствия информации) не могут быть уверены, что все они по-прежнему имеют ту же цель. Они *допускают*, что кто-то из членов команды уже располагает информацией о том, что цель либо достигнута, либо не может быть достигнута, либо она нерелевантная, и соответствующий агент может сообщить об этом другим членам команды. Поэтому если некоторый агент приходит к убеждению, что цель p уже перестала быть целью, и что другие члены команды еще этого не знают, то он должен отказаться от этой цели сам и сделать это известным для всех членов команды. Последнее очень важно, поскольку оно определяет необходимость и важность соответствующего *протокола взаимодействия* членов команды.

Чтобы не возникло путаницы в понимании индивидуальной конечной цели агента и общей конечной цели команды агентов, подчеркнем, что любая индивидуальная цель агента является компонентой дерева целей команды, а потому индивидуальная цель агента является частью общей цели команды. Поскольку нерелевантность или недостижимость индивидуальной цели агента делает недостижимым или нерелевантным достижение общей цели в соответствии с построенным планом, призванным реализовать дерево целей, то эта информация может интерпретироваться и как недостижимость общей конечной цели в рамках принятого дерева целей, и как непригодность текущего плана для достижения цели. И потому информация о недостижимости или нерелевантности индивидуальной конечной цели интерпретируется и как недостижимость общей конечной цели, а потому должна доводиться до всех членов команды агентов. Эта взаимосвязь индивидуальной конечной цели агента и общей конечной целью команды формулируется в работе [5] следующим образом.

Свойство 1. Если команда агентов имеет общую конечную цель p , то каждый агент

имеет p в качестве своей индивидуальной конечной цели.

Общее намерение команды определяется аналогично ее общей конечной цели.

Определение 5. (Общее намерение). Говорят, что команда агентов имеет *общее намерение выполнить некоторое действие* при принятых соглашениях о том, когда это намерение перестает быть таковым, если и только если все члены команды имеют общую конечную цель выполнить это действие при этих соглашениях. Более того, выполнив действие, они все должны быть совершенно уверены, что они делали именно это действие.

Рассмотрим некоторые важные свойства понятия *общего намерения*². Одно из этих свойств устанавливает взаимосвязь между индивидуальным намерением агента и общим намерением команды.

Свойство 2. Если команда намерена выполнить совместно некоторое действие, и при этом единственный агент имеет убеждение, что это действие должен выполнить именно он, то он имеет индивидуальное намерение выполнить это действие.

Аналогичное свойство имеет место также и для действий, которые в соответствии с общим намерением несколько агентов должны выполнить параллельно. Более точно,

Свойство 3. Если команда агентов намерена выполнить совместно сложное действие, состоящее из параллельно выполняемых действий отдельных агентов команды, то каждый агент (до тех пор, пока общее намерение имеет место) индивидуально имеет намерение выполнить свою часть общего действия, предусмотренного общим намерением.

При этом агенты должны быть совершенно уверены, что они выполняют это действие совместно с остальными агентами, вовлеченными в действия по реализации общего намерения.

Еще один важный аспект выполнения общего намерения касается случая, когда агенты должны совместно выполнять действия, упорядоченные во времени. Рассмотрим этот аспект сначала для случая намерений и действий отдельного агента. В программировании обычно

² В работе [5] все рассматриваемые свойства формулируются в виде теорем, которые доказываются.

последовательность действий определяется по выполнению соответствующих условий, контролируемых, например, оператором "if-then-else". Агент может выполнять последовательность действий в существенно более гибкой манере. На каждом шаге исполнения последовательности агенту достаточно знать два факта, а именно, какой шаг только что был выполнен и что в текущий момент он исполняет оставшуюся часть последовательности. Такая стратегия выполнения последовательности называется "пошаговым исполнением" [5]. В нем каждый шаг определяется контекстом, в котором агент имеет определенную уверенность (убеждение). Это свойство пошагового исполнения формулируется в работе [5] в виде следующего утверждения:

Свойство 4. Если агент намерен выполнить последовательность действий в манере пошагового исполнения, то он намерен также выполнить каждый из шагов в контексте более общего намерения.

Случай совместного исполнения последовательности действий в пошаговой манере более сложен, однако он существенно базируется на идее пошагового контекстуального исполнения индивидуального действия каждым агентом. По сути, при совместном исполнении последовательности действий каждый из агентов нуждается только в информации о том, когда наступает очередь его действия. При этом не требуется, чтобы все агенты команды приходили к совместному убеждению, что "такой-то из агентов должен исполнять в данных условиях такое-то действие". Отсюда вытекает следующее свойство общего намерения по совместному исполнению некоторой последовательности действий [5]:

Свойство 5. Если команда агентов намерена совместно выполнить некоторую последовательность действий, то агент, исполняющий некоторое действие общей последовательности, должен иметь индивидуальное намерение исполнить его как часть более общего намерения. При этом он всегда должен знать, когда заканчивается исполнение предшествующей части последовательности действий, когда он исполняет свое очередное действие и когда оно заканчивается.

Однако во многих случаях командной деятельности предполагается, что агенты выполняют

свои последовательности действий совместно в пошаговой манере, т.е. каждый агент выполняет последовательность действий, которые могут быть упорядочены с шагами действий других агентов. Примером является, например, игра роботов в футбол [23]. Этому случаю отвечает нижеследующее утверждение о свойствах совместного намерения группы агентов [5]:

Свойство 6. Если команда агентов намерена исполнить последовательность действий в совместной пошаговой манере, то агенты, исполняющие некоторый шаг, должны иметь совместное намерение выполнить его как часть более общего намерения.

Теория общих намерений построена как строгая теория командной работы, в основу которой положена логическая модель, основанная на идеях модальной логики. Она формулирует основные понятия и их свойства, которые в соответствующих логиках (логике намерений, и т.п.) могут быть формализованы. Именно поэтому в ней обращается такое внимание на используемые понятия, их взаимосвязи и свойства. Заметим, что одна из наиболее распространенных моделей (архитектур) агентов и многоагентных систем строится именно на основании подобной модели, а именно на основании модели убеждения–желания–намерения (*belief-desire-intention*, BDI-модель). По-видимому, именно BDI-модель, главным образом, повлияла на общую идею построения теории общих намерений.

Теория общих намерений была фактически первой моделью, которая сформулировала проблему командной работы агентов и очертила ее рамки. И хотя она была опубликована еще в 1989–91 годах, до сих пор она остается в центре внимания специалистов. В частности, во многих новых разработках моделей и инструментариев для поддержки процессов создания приложений в области командной работы агентов в той или иной степени используются ее идеи. Работа [5] остается по-прежнему одной из наиболее цитируемых работ. Она была первой работой, которая обратила внимание специалистов на то, что в организации командной работы важную роль играет корректно построенный протокол общения агентов.

Однако как каждая классическая работа, она имеет и ряд недостатков. Любая разумная постановка задачи командной работы должна

принимать во внимание динамику плана: планы всегда развиваются во времени и в пространстве. Агенты обычно начинают с частичного плана и расширяют его вплоть до получения полного плана по мере его исполнения в динамически изменяющемся контексте. Поскольку убеждения агента могут быть ложными, а мир может изменяться как в процессе планирования, так и в процессе исполнения действий в соответствии с частичным планом, то может потребоваться пересмотр плана. По этим же причинам все компоненты модели, указанные выше (*индивидуальные обязательства агентов, индивидуальные намерения агентов, общие обязательства и общее намерение*), могут быть неполными: агент может иметь частичный план действий, может быть не назначен агент для некоторого действия, а потому может отсутствовать индивидуальный план для него или групповой план; агент принял некоторое новое обязательство, но еще не определил, какое намерение должно быть в связи с этим принято. Неполной может быть также информация, полученная в результате коммуникаций с внешней средой и с другими агентами, что не дает возможности полностью определиться с планом. Принятый план может оказаться неработоспособным и должен быть пересмотрен. Однако эти аспекты, связанные с динамикой планирования командной работы, в теории общих намерений не отражены. Они более глубоко представлены в описываемой в следующем подразделе *Теории общих планов*.

1.2. Теория общих планов

Теория общих планов ("*Shared Plans Theory*", [10]) строится несколько иначе, чем рассмотренная ранее теория общих намерений. Она не использует общие для команды агентов ментальные понятия "общее намерение", "общая долгосрочная цель". Ее базовыми понятиями являются *групповой план* и *индивидуальные ментальные понятия агентов*. Поясним эти понятия.

Под групповым планом понимается план совместного выполнения некоторого множества действий группой агентов. Частными случаями группы является команда агентов в целом, а также отдельный агент. Групповой план, как он понимается в работе [10], не является простой суммой индивидуальных планов агентов груп-

пы. Помимо множества действий отдельных участников группы агентов, согласованных с множеством условий (время, место, ресурсы и т.д.), он содержит в себе еще и другие компоненты, которые, по сути, превращают множество взаимодействующих агентов в команду. Прежде чем пояснить суть группового плана и индивидуальных намерений агентов, поясним терминологию, которая используется авторами работы [10].

Под *действием* α авторы понимают абстрактную, достаточно сложную сущность, которая описывается различными свойствами, например, *тип действия, время действия, агент или группа агентов, исполняющая действие, другие объекты*, вовлеченные в выполнение этого действия и т.д.

Одним из центральных понятий рассматриваемой теории является понятие функции "*предписание*"³. Функция *предписание* ассоциируется с каждым действием и задает множество инструкций, определяющих, каким именно образом действие должно выполняться. В общем случае предписание представляется в виде иерархической структуры, узлами которой являются предписания более низкого уровня. Предписание, соответствующее каждому узлу, задается описанием группы действий β_i ($1 \leq i \leq n$) и ограничениями ρ_j ($1 \leq j \leq m$), наложенными на эти действия. Совместно они определяют действие, поставленное в соответствие узлу.

Предполагается, что агент имеет библиотеку предписаний для действий различных типов, и эта библиотека со временем обновляется. Различные агенты могут иметь различные библиотеки. Обозначим предписание, соответствующее действию α , символом R_α , причем $R_\alpha = \{\beta_i, \rho_j\}$. Заметим, что предписание R_α может иметь неозначенные переменные, в частности, время выполнения действия. Листьями структуры, описывающей предписание, являются конкретные неделимые действия. Они называются действиями базового уровня. Другие действия называются сложными действиями. Предполагается, что агент, исполняющий некоторое действие, имеет убеждение в том, отно-

³ Перевод с английского термина "recipe", что может иначе переведено как "рецепт", "инструкция" и т.п.

сится ли действие к базовому уровню, или оно является сложным. Полагается также, что эти убеждения соответствуют действительности.

Основные особенности группового (командного) плана действий и его отличия от множества планов индивидуальных действий агентов, таковы [10]:

1. Групповой план действий требует, чтобы группа (команда) агентов пришла к согласию выполнять предписания (множества инструкций), которому они будут следовать в групповых действиях ("*mutual belief of a (partial) recipe*"). Это означает, что должен существовать механизм выбора этого предписания, а также механизм для комбинирования (объединения) индивидуальных знаний агентов об этом предписании (путем коммуникаций, путем проверки условий применимости тех или иных индивидуальных действий и т.д.).

2. Агенты должны принять на себя обязательства по отношению не только к своим индивидуальным действиям, но также и по отношению к действиям группы в целом (*individual intentions that the action be done*).

3. Аналогично, агент должен принять на себя обязательства по отношению к действиям других агентов (индивидуальное намерение выполнить—*individual "intending that"*).

4. План групповой деятельности может иметь в качестве компонент как планы индивидуальных агентов для назначенных действий, так и планы подгрупп (*individual or collaborative plans for the sub-actions*).

В результате группа должна иметь механизм для принятия решений о том, кто именно будет выполнять те или иные отдельные действия.

Данная теория далее для краткости называется далее *SP*. Она использует индивидуальное ментальное понятие агента *намерение выполнять* (*intending that*), которое очень похоже на традиционное понятие *намерение*. Отличие состоит в том, что это намерение относится к его действиям, связанным либо с сотрудничеством с членами команды, либо с выполнением некоторого совместного действия команды. Оно определяется аксиоматически и имеет целью предписать отдельному агенту некоторые действия (включая коммуникации с другими агентами), которые стимулируют членов команды, группы агентов из команды или всю команду выполнять поставленные задачи.

В этой теории различают два вида планов. Один из них—это так называемый *Полный общий план* (далее для краткости *FSP—Full Shared Plan*), а другой—это *Частичный общий план* (*PSP—Partial Shared Plan*). Полный общий план *FSP* выполнить действия α относится к ситуации, в которой детально описаны все аспекты совместных действий α команды агентов. При этом описание ситуации должно включать совместные убеждения⁴ агентов команды и общее согласие команды выполнять α согласно некоторому исчерпывающе описанному предписанию R_α . Это предписание содержит в себе описание всего множества действий β_i и множества условий, когда они должны выполняться в составе действий α .

Формально $FSP = FSP(P, GR, \alpha, T_p, T_\alpha, R_\alpha)$ определяется следующими атрибутами [10]: P —план, GR —группа агентов, выполняющая план, α —командная цель, T_p —время, к которому относится план, T_α —время выполнения действия по достижению командной цели α согласно предписанию R_α . План $FSP(P, GR, \alpha, T_p, T_\alpha, R_\alpha)$ остается актуальным для исполнения (истинен) тогда и только тогда, когда выполнены следующие условия:

1. Все члены группы GR совместно убеждены, что каждый из них полагает⁵, что пропозициональное выражение $D_0(GR, \alpha, T_\alpha)$ истинно, т.е. что группа GR выполняет α во время T_α .

2. Все члены группы GR совместно убеждены, что R_α есть предписание для выполнения α .

3. Для каждого шага β_i в R_α

- Некоторая подгруппа агентов GR_k ($GR_k \subseteq GR$) имеет полный общий план *FSP* для действия β_i , представленного предписанием R_{β_i} . В частном случае подгруппа может состоять из одного агента, который имеет полный индивидуальный план, анало-

⁴ Напомним, что *совместные (взаимные) убеждения*: складываются из убеждений группы агентов и формализуются как конъюнкция утверждений, отвечающих убеждениям отдельных агентов группы.

⁵ Синонимы: верит, убежден.

гичный полному плану FSP для подгруппы агентов.

- Другие члены группы GR убеждены, что существует предписание, такое, что группа GR_k может выполнять предписание β_i и имеет FSP для β_i , но при этом другие члены команды могут не знать предписание R_{β_i} .

Таким образом, теория общих планов пытается описать все множество взаимосвязанных намерений и убеждений агентов, вовлеченных в командную работу. Однако на практике команда не имеет полного общего плана, а располагает только частичным общим планом PSP , который представляет собой только некоторый "срез" ментального состояния команды в частной ситуации, возникающей в процессе командной работы. Коммуникации агентов имеют целью как раз восполнить условия реализации полного общего плана. В некоторых случаях, например, в динамической предметной области, полный общий план невозможно построить вообще.

Существенны следующие три частных случая, когда командная работа выполняется в условиях частичного общего плана. Первый из них отвечает случаю, когда имеется соперничающая команда, действия которой невозможно предусмотреть. Тогда предписание R_α и не может быть специфицировано полностью. Оно эволюционирует в результате реакции команды на действия соперника, когда члены команды приходят к новому взаимному убеждению на очередном шаге β_i , что далее ведет к формированию некоторого нового частичного общего плана. Это характерно для приложений, в которых моделируются боевые действия, игра в футбол и т.п.

Другой случай возникает в ситуации, когда целесообразно пересмотреть подгруппу, назначенную в плане для выполнения некоторого действия. Например, если некоторый сценарий задан в терминах ролей, а не в терминах имен конкретных агентов, группа агентов-исполнителей сценария может назначаться в соответствии с динамическим распределением ролей. Например, агент с текущей ролью нападающего при игре в футбол может оказаться слишком далеко от той зоны, в которой он в выбранном сценарии атаки должен находиться, но есть другой агент, возможно, с ролью защитника, который больше подходит для вы-

полнения выбранного сценария. Тогда его целесообразно включить в группу исполнителей выбранного сценария. Технически переназначение может производиться лидером команды или по инициативе самого потенциально более подходящего исполнителя.

В третьем случае агент или группа агентов к моменту, когда им следовало бы начать исполнения предписания R_{β_i} , еще не пришла к общему убеждению, что нужно исполнять шаг β_i . Поэтому члены группы еще не обменялись сообщениями, необходимыми для начала совместных действий в соответствии с предписанием R_{β_i} . Это может также произойти вследствие конфликтов индивидуальных и командных намерений членов команды.

Теория общих планов имеет строгое математическое обоснование (к сожалению, она представлена в слишком формализованном виде, так что местами она достаточно тяжела для понимания и, в особенности, для содержательной интерпретации). Хотя сама по себе эта теория уступает по популярности теории общих намерений, тем не менее, некоторые идеи этой теории оказались весьма продуктивными и используются при разработке моделей командной работы в комбинации с теорией общих намерений.

Теория общих планов, однако, тоже обладает рядом недостатков, которые приводят к трудностям ее практического использования. В ней не конкретизируется способ построения конкретного общего плана, полагая, что эта проблема является внешней по отношению к построенной теории. С одной стороны, это задает определенную свободу выбора алгоритма построения такого плана. Но, с другой стороны, большинство известных методов не рассчитано на некоторые специфические особенности командной работы агентов, особенно, в условиях противодействия, а эти особенности оказываются зачастую ключевыми. Например, может оказаться, что некоторый агент по той или иной причине оказывается не в состоянии выполнить некоторое назначенное ему действие. Например, у него может быть исчерпан некоторый необходимый ресурс. Он может быть выведен из строя. Состояние внешней среды может измениться так, что действие в новых условиях не может быть выполнено. Возможно, что это действие может быть выполнено другим агентом,

но эта возможность в теории общих планов не предусмотрена. В другой ситуации требуемое действие может оказаться невыполнимым для всех агентов. Тогда необходимо вообще выбрать новый план достижения цели, который будет включать в себя иные действия агентов команды, и эти действия должны распределяться между агентами в режиме реального времени. То же самое касается условий согласования действий. Но возможность изменения общей схемы сценария достижения цели вообще не предусматривается рассматриваемой теорией, хотя это может потребоваться во многих случаях. Заметим, что в теории общих намерений может иметь место аналогичная ситуация. Там заранее определяется условие, при котором цель либо нерелевантна, либо недостижима, и тогда агенты отказываются от ее достижения. При этом возможность изменения самой цели и/или сценария ее достижения (это может привести к изменению вида утверждений p и/или q) тоже не рассматривается. Следует также заметить, что коллективная обработка непредвиденных ситуаций не рассматривается в обеих теориях.

Недостатки теорий командной работы на практике обычно компенсируются некоторыми дополнительными средствами. Часто возникающие проблемы могут быть успешно разрешены благодаря использованию коммуникаций. Другой потенциально эффективный путь разрешения потенциальных проблем выполнения командной миссии—это обогащение (экспертных) знаний, которые используются агентами для построения плана и специальных структур их представления. Некоторые варианты подобных решений рассматриваются далее (вне теории командной работы) при обзоре и анализе моделей и инфраструктур, поддерживающих быструю разработку приложений в области командной работы. Следует отметить, что теории командной работы, отличные от работ [5, 10], в доступной литературе не описаны.

Заключение

Агентские технологии получают все большее признание и распространение среди интеллектуальных информационных технологий. Одной из областей активного использования этих технологий является современная робототехника. Однако взгляды на важность агентских технологий для различных задач робо-

техники за рубежом и в России долгое время, к сожалению, были различными.

Первые применения агентской архитектуры в робототехнике за рубежом относятся еще к началу 1990-х годов. В России такой интерес только что наметился, и пока реально разработанные отечественные робототехнические системы, использующие многоагентную архитектуру, отсутствуют⁶.

Однако современные мировые тенденции ушли гораздо дальше, и в настоящее время речь идет о создании команд автономных виртуальных и физических роботов для выполнения достаточно сложных миссий. Предполагается, что команде автономных роботов только ставится цель миссии, а построение плана, его выполнение, а также динамическое перепланирование командной работы в зависимости от воздействия внешней среды, состояния членов команды и противодействия оппонирующей команды — это уже должно решаться командой автономно. И этот взгляд на возможности агентских технологий для обеспечения описанного стиля командного поведения роботов не является фантазией или делом далекого будущего. Именно такое поведение уже демонстрируют команды роботов, играющих в футбол, на чемпионате мира среди роботов, проводимом ежегодно уже в течение 14 лет.

Именно этим, т.е. реально большим различием во взглядах зарубежной и российской науки на роль агентских технологий в командной работе роботов, мотивирована настоящая работа. В ее первой части дано введение в современную теорию командной работы автономных агентов. Во второй части будут рассмотрены модели, инструментальные средства и языки спецификации командного поведения агентов, управляющих поведением роботов при совместном выполнении ими автономной миссии.

Литература

1. Aalst, W., Hofstede, A.: YAWL: Yet Another Workflow Language. In: Information Systems, vol. 30(4), pp. 245–275 (2005).
2. Aalst, W.: The Application of Petri Nets to Workflow Management. In: Journal of Circuits, System and Computers, vol. 8(1), pp. 21–66 (1998).

⁶ С "высоких трибун" часто звучат значительно более оптимистические утверждения, но пусть они останутся на совести авторов.

3. Agent Communication Language Specification. <http://www.fipa.org/repository/aclspeecs.html>
4. Barbuceanu, M., Fox, M.: The Architecture for an Agent Building Shell. In: Wooldridge, M., Muller, J., Tambe M. (eds.) *Intelligent Agents*, vol. 2, Lecture Notes in Artificial Intelligence, 1037, Springer-Verlag, Heidelberg (1996).
5. P.Cohen and H.J.Levesque. Teamwork. *Nous*, 35 (1991).
6. Farinelli, A., Iocchi, L., Nardi, D., Ziparo, V.A.: Assignment of Dynamically Perceived Tasks by Token Passing in Multi-Robot Systems. In: *Proceedings of the IEEE, Special Issue on Multi-Robot Systems* (2006).
7. Farinelli, A., Iocchi, L., Nardi, D., Ziparo, V.A.: Task Assignment with dynamic perception and constrained tasks in a Multi-Robot System. In: *Proceedings of International Conference on Robotics and Automation (ICRA'05)* (2005).
8. Gorodetsky V., and Kotenko, I. Scenarios Knowledge Base: A Framework for Proactive Coordination of Coalition Operations. (Eds. M.Pechoucek and A.Tate) *Proceedings of the Third International Conference on Knowledge Systems for Coalition Operation (KSCO-04)*, Florida, USA. October, 2004, pp. 83-88.
9. Grosz, B.: Collaborating Systems. *AI Magazine*, 17(2), 1996.
10. B.Grosz and S.Kraus. Collaborative Plans for Complex Group Actions. *Artificial Intelligence*, 86, 1996, 269-358 (1996).
11. Java Agent Development Framework. <http://jade.tilab.com/>.
12. Jennings, N.R., Varga, L.Z., Aarnts, R.P., Fuchs, J., and Skarek, P.: Transforming standalone expert systems into a community of cooperating agents. *Int. Journal of Engineering Applications of Artificial Intelligence*, 6(4), 1993, pp. 317-331 (1993).
13. Jennings, N.: Controlling Cooperative Problem Solving in Industrial Multi-Agent Systems Using Joint Intentions. In: *Artificial Intelligence*, vol. 75, pp. 195--240 (1995).
14. Jennings, N., Norman, T., Faratin, P.: ADEPT: An Agent-Based Approach to Business Process Management. In: *ACM SIGMOD Record*, vol. 27(4), pp. 32—39 (1998).
15. Konolige, K.: COLBERT: A Language for Reactive Control in Saphira. In: *KI '97 Proceedings of the 21st Annual German Conference on Artificial Intelligence: Advances in Artificial Intelligence*, Lecture Notes in Computer Science, vol. 1303, pp. 31—52. Springer-Verlag, London, (1997).
16. Löttsch, M.: XABSL - A Behavior Engineering System for Autonomous Agents. Diploma thesis. Humboldt-Universität zu Berlin (2004) Available online: <http://www.martinloetzsch.de/papers/diploma-thesis.pdf>.
17. Palamara, P.F., Ziparo, V.A., Iocchi, L., Nardi, D., Lima, P.: Teamwork Design Based on Petri Net Plans. *RoboCup International Symposium* (2008).
18. E.Oliveira, J.Fonseca, A.Steiger-Garcia. Resource selection and cost estimation on a MAS for civil construction companies. In *Proceedings of the International Workshop "Distributed Artificial Intelligence and Multi-agent-Systems"*, St. Petersburg, Russia, pp. 162-173 (1997).
19. Pynadath, D., Tambe, M., Chauvat, N., Cavedon L.: Toward Team-Oriented Programming. In: Jennings, N., Lesperance, Y. (eds.) *Proceeding ATAL '99 6th International Workshop on Intelligent Agents VI, Agent Theories, Architectures, and Languages*, pp. 233–247. Springer, London, (2000).
20. Pynadath, Tambe, M.: An automated teamwork infrastructure for heterogeneous software agents and humans. In: *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS): Special Issue on Infrastructure and Requirements for Building Research Grade Multi-Agent Systems* (2002).
21. Rich, C., Sidner, C.: COLLAGEN: When Agents Collaborate with People. In: *Proceedings of the International Conference on Autonomous Agents (Agents'97)* (1997).
22. Risler, M., von Stryk, O.: Formal Behavior Specification of Multi-Robot Systems Using Hierarchical State Machines in XABSL. *AAMAS08 Workshop on Formal Models and Methods for Multi-Robot Systems*. Estoril (2008), see also <http://www.xabsl.de>.
23. RoboCup initiative, <http://www.robocup.org/>.
24. Sycara, K., Paolucci, M., Giampapa, J., van Velsen, M.: The RETSINA Multiagent Infrastructure. *Autonomous Agents and Multi-agent Systems*, vol. 7(1) (2003).
25. Sycara, K., Sukthankar, G.: Literature Review of Teamwork Models. CMU-RI-TR-06-50, November 2006, Robotics Institute Carnegie Mellon University. (2006) http://www.ri.cmu.edu/pub_files/pub4/sycara_katia_2006_1/sycara_katia_2006_1.pdf.
26. M.Tambe. Towards Flexible Teamwork. *Journal of Artificial Intelligence Research*, 7, 1997, pp. 83-124 (1997).
27. Tambe, M., Shen, W.-M., Mataric, M., Pynadath, D., Goldberg, D., Modi, P.J., Qiu, Z., Salemi, B.: *Proceedings of AAAI Spring Symposium on Agents in Cyberspace*, pp 136- 141 (1999).
28. Teamcore research group. <http://teamcore.usc.edu>.
29. Weske, M.: *Business Process Management: Concepts, Languages, Architectures*. Springer (2007).
30. Ziparo, V.A., Iocchi, L.: Petri Net Plans. In: *Fourth International Workshop on Modeling of Objects, Components, and Agents*, pp. 267-289. Turku (2006).
31. Ziparo, V., Iocchi, L., Nardi, D., Palamara, P.F., Costelha, H.: Petri Net Plans: a Formal Model for Representation and Execution of Multi-Robot Plans. In: *AAMAS*, Estoril, (1), pp. 79–86. (2008).

Городецкий Владимир Иванович. Доктор технических наук, профессор, главный научный сотрудник лаборатории интеллектуальных систем Санкт-Петербургского института информатики и автоматизации РАН. Окончил Ленинградскую военно-воздушную инженерную академию им. А.Ф. в 1960 году и Ленинградский государственный университет в 1970 году. Автор более 350 публикаций. Область научных интересов: многоагентные системы, извлечение знаний из данных, самоорганизация и др. E-mail: gor@iiias.spb.su, <http://space.iiias.spb.su/ai/gorodetsky>