The Adaptive Agent Architecture: Achieving Fault-Tolerance Using Persistent Broker Teams[*]

Sanjeev Kumar
Dept. of Computer Science
Oregon Graduate Institute
20000 NW Walker Road,
Beaverton, OR 97006, USA
+1-503-748-7803
skumar@cse.ogi.edu

Philip R. Cohen
Dept. of Computer Science
Oregon Graduate Institute
20000 NW Walker Road,
Beaverton, OR 97006, USA
+1-503-748-1326
pcohen@cse.ogi.edu

Hector J. Levesque
Dept. of Computer Science
University of Toronto
Toronto, Ontario,
M5S 3H5, Canada
+1-416-978-3618
hector@cs.toronto.edu

ABSTRACT

Brokers are used in many multi-agent systems for locating agents, for routing and sharing information, for managing the system, and for legal purposes, as independent third parties. However, these multi-agent systems can be incapacitated and rendered non-functional when the brokers become inaccessible due to failures such as machine crashes, network breakdowns, and process failures that can occur in any distributed software system.

We propose that the theory of teamwork can be used to create robust brokered architectures that can recover from broker failures, and we present the Adaptive Agent Architecture (*AAA*) to show the feasibility of this approach. The AAA brokers form a team with a joint commitment to serve any agent that registers with the broker team as long as the agent remains registered with the team. This commitment enables the brokers to substitute for each other when needed. A multi-agent system based on the AAA can continue to work despite broker failures as long as there is at least one functional broker in the system. Another team commitment enables the brokers to start new brokers and recruit them to the broker team. As a result, an AAA-based multi-agent system can maintain a specified number of functional brokers in the system despite broker failures, thus effectively becoming a self-healing system.

Teamwork is explained in terms of the theory of joint intentions. The previous theory assumes that team members remain in a team as long as the team exists. We extend the theory of joint intentions to allow dynamic but persistent broker teams whose members can change with time. In particular, we introduce the notion of joint commitment to a team wherein the individuals are committed to the team as an entity rather than to the members that constitute the team. We present a logical analysis of the AAA and show that the AAA brokers have the required individual commitments that result in robust behavior when the brokers act rationally.

Keywords

Teamwork and Cooperation, Fault-Tolerance, Foundations, Multi-Agent System Architectures, Middle Agents

---

[*] This research was supported by the DARPA CoABS program contract F30602-98-2-0098, A0 G352.

1 INTRODUCTION

Multi-agent systems are prone to failures that can occur in any distributed software system. Bugs and improperly handled exceptions in the agent program or in the supporting environment, machine crashes, network partitioning, and numerous other hardware and software faults can make agents unavailable suddenly for unforeseen periods. The traditional distributed systems literature provides various fault-tolerance techniques to recover from these failures. We have shown elsewhere [16] that most of these techniques are meant for specific failure situations and they require special infrastructural support. For example, the techniques of hot backups [1], object group replication [2], virtual synchrony [2], and N-version voting [3] need specific mechanisms for communication and synchronization among the replicas. It may not be possible to use these techniques in multi-agent systems without extensive modifications to the underlying agent infrastructure. On the other hand, a technique based on a multi-agent system concept may be amenable to implementation by adding a plan to the plan library of generic agents. We note that earlier work on teamwork [17] has shown agent teams to be more robust than a collection of agents in the face of adversity and unforeseen situations [13, 20]. The reason behind this robust behavior exhibited by teams is that the members of a team are committed not only to the success of their portion of the joint action but also to the success of the team as a whole. A team will try to recover from problems and will abandon the joint goal only when it is mutually believed by the team members that the goal is no longer possible. This discussion motivates us to investigate exploiting teamwork to achieve fault-tolerance.

Multi-agent systems often require brokers[1] or middle agents for accepting requests, locating capable agents, routing requests and responses, sharing of information, managing the system, registering agent capabilities, and for various other facilitation tasks [19]. Furthermore, with the advent of electronic commerce, middle agents are likely to become increasingly important for legal purposes as independent third parties [12]. However, brokered systems are brittle because the facilitator is a single point of failure. Our experience with Quickset [9], a multi-agent system based on the facilitated Open Agent Architecture [5, 18], reinforces the need for an agent architecture that can recover quickly from broker failures.

We hypothesize that the theory of teamwork can be used for recovering brokered multi-agent systems from failures due to sudden broker unavailability. We present the Adaptive Agent Architecture (*AAA*), a robust brokered architecture, to show the feasibility of our approach. The AAA uses teamwork (1) to recover a multi-agent system broker failures, and (2) to maintain a certain minimum number of functional brokers in the system even when some of the brokers become inaccessible. The performance characteristics of using teamwork for fault-tolerance have been presented in another paper [16]. In this paper, we will be mainly concerned with the logical characterization of the recovery techniques and relating the teamwork theory to its application in the AAA.

One departure from our previous theory [8, 17] is that a team can exist independently of the identity of its members. For example, the New York Yankees are a team even if all the players are traded or

---

[1] In this paper, we use the terms broker, middle-agent, and facilitator interchangeably as the work described in this paper may be extended to most middle agents.

it is sold to new owners. Our earlier work characterized a team in terms of joint commitment between the individuals in a team. In this paper, we introduce the notion of team commitment wherein the agents involved are committed to the team as an entity rather than to the individuals that constitute the team. We then use this concept of team commitment to establish the various commitments that lead to robust behavior in AAA-based multi-agent systems.

The next section discusses the theory of teamwork, the relation between teamwork and fault-tolerance, and different ways of implementing teams of software agents. The adaptive agent architecture is introduced and its logical characterization is discussed in section 3. We also walk through a scenario in which an AAA-based multi-agent system recovers from a broker failure, and explain the commitments at each step that lead to the recovery. In section 4, we present the definitions of team commitment and the related concepts and show how to logically establish the different commitments that follow from a team commitment. We discuss the related work and future work in section 5 and finally, we conclude in section 6 with a brief summary of the results.

2 TEAMWORK

A team is more than just a collection of individuals working together in close cooperation to achieve a common goal. The agents in a team must have a shared goal as well as a shared mental state [7, 8,17]. Absence of either of these characteristics implies lack of a team irrespective of the agents having a common goal, having coordinated actions, or even their being mutually helpful. When operating normally, teams of agents may be indistinguishable from a collection of agents coordinated by external forces. However, it is when things go astray that the benefits of teamwork are most apparent – team members work together to overcome adverse conditions without abandoning the team. Team activity is explained in terms of the theory of joint intentions [6, 7, 8, 17]. This theory characterizes an agent's behavior in a team in terms of its internal state described in modal logic, linear time temporal logic, and dynamic logic of action.

2.1 Theory of Teamwork

A joint persistent goal (*JPG*) formalizes the notion of joint commitment. The existence of a JPG between a group of agents is a sufficient condition for the formation of a team with respect to that JPG. The JPG prevents a team from falling apart in the wake of adversities and from changes in private beliefs of team members by requiring that agents cannot abandon a team unless there is mutual belief about the completion, feasibility or irrelevance of the jointly committed goal. Two agents have a joint intention (*JI*) to do an action *a* if they have a JPG to do *a* while being in a particular mental state. A joint intention requires the starting mutual belief that the team members are going to do the jointly intended action next. Joint intention brings about one of the following mental states. (1) The agents mutually believe that *a* has been done; (2) They mutually believe that *a* is impossible; (3) They mutually believe that *a* is irrelevant. In section 4, we will introduce persistent teams using the concepts of team persistent goal (TPG) and team intention (TI). These are similar to JPG and JI respectively, except that the commitments involve the team as an entity rather than the individuals that constitute the team.

2.2 Teamwork and Fault-Tolerance

We intuitively expect teams to be more robust than a simple collection of individuals. Earlier work in multi-agent systems on teamwork has shown agent teams to be more robust than a collection of agents in the face of adversity and unforeseen situations [13, 20]. The reason behind this robust behavior exhibited by teams is that the members of a team are committed not only to the success of their portion of the joint action but also to the success of the joint action as a whole. Therefore, team members do whatever it takes to achieve the jointly committed goal. When a team member gets into trouble, the others will come to its aid, and when a team member becomes unavailable, the others will try to achieve the jointly committed goal if it is still possible to do so. A team will try to recover from problems and will abandon the joint goal only when it is mutually believed by the team members that the goal is no longer possible. Hence, we may conclude that teams are inherently fault-tolerant.

A team of agents can successfully recover from failures such as sudden unavailability of a teammate if (1) there is at least one available sub-team capable of doing the task that the unavailable agent was supposed to do, or (2) there is at least one alternative method that can be jointly executed by the available team members to achieve the joint goal. The fault-tolerance in AAA-based multi-agent systems comes from the observation that in a large multi-agent system there will typically be more than one middle agent and that these middle agents may be able to substitute for each other when needed. Therefore, if these middle agents form a team with appropriate joint commitment, they *will* substitute for any middle agent that becomes unavailable. Hence, the multi-agent system can continue to work as long as there is at least one middle agent remaining in the broker team.

The definitions of joint commitment and team commitment capture the essence of team behavior and imply the required individual commitments. It is these individual commitments that impart the expected fault-tolerant behavior to a team, as it is the individual team members that ultimately act on behalf of the team. However, it must be realized that simply having the right commitments is not sufficient to achieve fault-tolerance: the team members must have the means to honor these commitments. The teamwork process can also be looked upon as a distributed coordination protocol that prescribes the desired end result at each step of a distributed interaction. It is up to an agent (or the agent designer) to decide the most appropriate rational action to taken in order to achieve those ends. The ultimate goal of the entire interaction is achieved when the end result for each intermediate step in any complete path has been achieved. The interaction protocol includes steps for situations when things go wrong and this in turn results in the fault-tolerant behavior.

2.3 Implementing a Team in Software

A group of software agents can form a normal team with respect to a particular goal by establishing a joint commitment (JPG) with each other with respect to that goal. The agents can also form a persistent team by establishing a team commitment (TPG) with respect to a goal[2]. In either case,

---

[2] Note that this team will be persistent with respect to change in team membership for all goals. However, in order for the team to persist beyond completion of one-time goals, the main team goal should be a maintenance goal.

implementing a team in software requires (1) establishing the initial mutual beliefs that lead to the joint commitments, (2) ensuring that the required individual commitments follow from the joint commitment, and (3) providing support for appropriate actions to enable the individual commitments to be honored. A generic way to meet these requirements is by using a planning engine that can reason over modal logic, temporal logic, and dynamic action logic and take appropriate actions. Rule-based approaches can also be used to implement teams as in [13]. Agent programs can also use an explicit model of teamwork, with operators representing the various constructs of the joint intention theory, to form teams. This approach is used in [20].

An agent programmer can use any other method, implicit or explicit, to incorporate the beliefs and the behavior required for JPG (or TPG) formation into the agent program. However, the semantics of the agent programs must guarantee that the interactions among these agents lead to the formation and the proper execution of the required joint commitments. A variation of the rule-based approach is to establish logically the various individual commitments that follow from the joint commitment under the different possible conditions, decide on the rational actions that the agent can take under those conditions, and then use it as software design specification for the agent program. This method is currently used to implement broker teams in the AAA.

3 THE ADAPTIVE AGENT ARCHITECTURE (*AAA*)

The AAA is a facilitated multi-agent system architecture under development and forms the basis of our research in fault-tolerance and agent communication languages. It interoperates with the Open Agent Architecture [5] and it is currently used in multi-agent systems such as Quickset [9] that place heavy demand on inter-agent communication and yet provide acceptable real time response.

3.1 Overview

The AAA is an agent infrastructure that can be used to build fault-tolerant brokered multi-agent systems. The AAA agent library has been developed in Java and it provides an agent shell for developing AAA agents. The library also provides a facilitator agent that serves as a broker, a matchmaker and a recruiter. Henceforth, in this paper, we will refer to the AAA facilitator as the AAA broker. The AAA brokers can be interconnected to form arbitrary networks and the agent library supports both facilitated and direct inter-agent communication. The AAA agents advertise their capabilities as well as an address for connection requests, with a broker during registration. TCP/IP is used for network transport and the TCP mechanisms and timeouts are used for detection of connection failures. The brokers as well as other agents can dynamically enter and leave AAA-based multi-agent systems. The AAA brokers form a team for the purpose of fault-tolerance and they share knowledge about who is connected to whom with the team members.

3.2 Logical Characterization

The AAA brokers form a persistent broker team, that we call T, when they register with each other. This broker team establishes a team intention (*TI*) about performing certain concurrent actions. The AAA has been designed to implement the specification of teamwork that follows from the team intentions for these actions and we argue that the observed fault-tolerance behavior of AAA-based multi-agent systems is a consequence this design.

AAA Mission Statement: *Whenever an agent registers with the broker team, the brokers have a team intention of connecting with that agent, if it ever disconnects, as long as it remains registered with the broker team.* More formally,

$$\models \forall y\ [(\text{agent } y) \wedge (\textbf{DONE}\ (\text{registered } y\ T)?)\ \supset (\textbf{TI}\ T\ a(y)\ \ (\text{registered } y\ T))]$$

where, $a(y) = (\ \textbf{WHILE}\ (\text{registered } y\ T)\ \ \textbf{DO}\ [\text{if } \neg(\text{connected } y\ T)\ \textbf{THEN}\ (\text{connect } y\ T)]\ )$

The AAA broker team also has a maintenance goal of having at least N brokers in the team at all times where N is specified during the team formation. This maintenance goal results in an achievement goal, expressed in terms of TI, to have N brokers in the broker team whenever the broker team mutually believes that the number of brokers in the team is less than the specified number. The AAA-brokers act rationally in response to this team intention by recruiting new brokers into the broker team. The design of the AAA brokers satisfies the resulting specification that we state in the following proposition.

Proposition 3.1: $\models\ (\textbf{GMB}\ T\ \neg(\text{numberOfBrokers } T\ N)) \supset (\textbf{TI}\ T\ (\text{numberOfBrokers } T\ N)?\ q)$
where, the escape condition q results from the maintenance goal.

Using the above propositions, along with other logical properties of the AAA, we can establish the commitments of the brokers in the team. These commitments result in fault tolerant behavior when the brokers act rationally and take appropriate actions to honor them. The commitments derived from the mission statement require the brokers in the broker team to locate and connect with any stranded agents that were connected to a broker teammate that is no longer accessible. As a result, the AAA-based agents that cannot function without a broker have access to brokers at all times, even when the broker serving them becomes suddenly unavailable. Proposition 3.1 requires the broker team to always have a certain number of brokers in the team despite broker failures. The broker team achieves this by getting new brokers started, on possibly different machines, and recruiting them into the broker team. The AAA agents are capable of spawning additional brokers but they can do so only when requested by the broker that is serving them. The end result is that new brokers get started on different machines on the network and the multi-agent system continues to function smoothly.

We state below the theorems that follow from the above propositions about commitments of an AAA broker team. We show how to prove these theorems in section 4.2.

Theorem 3.1: Whenever an agent registers with a broker, the broker has a commitment to make this fact mutually believed by the broker team.

Theorem 3.2: When an agent unregisters with a broker, the broker has a commitment to make this fact mutually believed by the broker team.

Theorem 3.3: When a broker discovers that an agent that is registered with the team is not connected, it has a commitment to make this fact mutually believed.

Theorem 3.4: When an agent that is registered with the broker team gets disconnected, the brokers have a team commitment to connect to that agent.

Theorem 3.5: When an agent that is registered with the broker team gets disconnected, all the brokers in the broker team have an individual commitment to connect to that agent.

Theorem 3.6: When a broker successfully connects to an agent that is registered with the broker team but got disconnected, it has a commitment to bring about mutual belief about this fact.

Theorem 3.7: When a broker that was committed to having a disconnected agent being reconnected to the team, learns that the agent has been reconnected to the broker team, it gives up its commitment to connect to that agent.

Theorem 3.8: When a broker believes that the broker team does not have a specified number of brokers, it has an individual commitment to have the required number of brokers in the broker team.

Theorem 3.9: When a broker successfully recruits a new broker into the broker team, it has an individual commitment to make this fact mutually believed by the broker team.

Next, we use these theorems to explain a recovery process when one of the brokers in an AAA-based multi-agent system becomes inaccessible.

3.3 A Recovery Scenario

Figure 1 illustrates a multi-agent system with three brokers and two other agents. The client agent periodically sends requests for which the distance agent is the only capable agent. The three brokers form a robust team characterized by the mission statement and proposition 3.1. This system can function only if both the client and the distance agents are registered with a broker. It has been specified that the multi-agent system should have at least three brokers at all times.

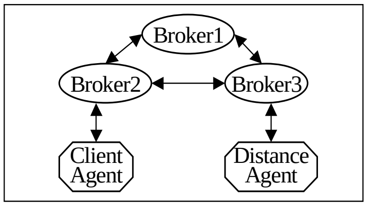

Figure 1

From theorem 3.1, the brokers have an individual commitment to bring about mutual belief when an agent registers with a broker. Therefore, when the client agent registers with Broker2, Broker2 informs this fact along with the name and address of the distance agent to Broker1 and Broker3. Similarly, when the distance agent registers with the Broker3, Broker3 informs this fact to Broker1 and Broker2.

After some time, we kill Broker3. When a broker teammate is no longer accessible, the other brokers believe that all the agents registered with that broker are disconnected. When Broker3 is killed, at least one of the remaining brokers, Broker1 or Broker2, discovers that Broker3 is no longer accessible to it and believes that the distance agent is not connected to the broker team. Therefore, from theorem 3.3, this broker has an individual commitment to bring about a mutual belief about its discovery. As a result, communication is predicted to take place among the brokers in the team.

From theorem 3.4, the broker team has a joint commitment to connect to the agent that it mutually believes is disconnected from the team. Moreover, from theorem 3.5, each of the remaining brokers has an individual commitment to contact the disconnected agent. The two brokers act rationally by attempting to contact the distance agent at the address given to them earlier by Broker3. If the distance agent accepts registration request from one of the brokers, it refuses subsequent registration requests from other brokers. Figure 2 illustrates the situation when broker1 has successfully contacted the distance agent.

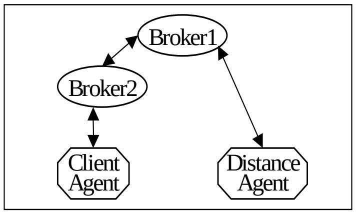

Figure 2

From theorem 3.6, Broker1 now has an individual commitment to inform the successful connection of the distance agent to its teammates. As a result, Broker1 will act rationally by communicating this information to Broker2, and from theorem 3.7, Broker2 will give up its attempt if it was still trying to contact the distance agent as the mutual goal has already been achieved. Moreover, from theorem 3.1, Broker1 needs to inform the registration and address information of the distance agent to Broker2. In the current AAA implementation, these two communication attempts are combined into one and just one message is sent from Broker1 to Broker2.

Broker1 requests the distance agent for agent-specific information such as its capabilities after successfully contacting it. Any pending request from the client agent that could not be completed due to the failure of Broker3 may be sent again to the distance agent and the system continues to work.

Recall that the broker team had a maintenance goal of having three brokers in the system. Therefore, when a surviving broker discovers that Broker3 is no longer accessible, it will have a commitment to bring about mutual belief about this fact and hence, communication will take place between the surviving brokers. From theorem 3.8, both the surviving brokers have an individual commitment to recruit a new broker to the team. Broker2 knows of an agent (the client agent) that can spawn a new broker and requests it to do so. After the new broker comes up, it contacts the broker on whose request it was started. Theorem 3.9 predicts that Broker2 will now have an individual commitment to bring about a mutual belief that the broker team has the required number of brokers and therefore, communication will take place among the brokers. We will see in the next section that communication is also required by the commitment about team membership in the definition of weak team goal. At the end of this process, a configuration similar to that in figure 1 will have been restored and hence the multi-agent system has recovered from broker failure.

4 PERSISTENT TEAMS

Teams in the real world may be one-time teams that are disbanded after the team goal has been achieved, or they may be persistent teams that continue to exist even when the team members change. Persistent teams are especially desirable from a fault-tolerance perspective as agents that fail will generally be replaced by other agents during the recovery process. We have seen in earlier sections that a team is formally defined in terms of a joint persistent goal (JPG) between the team members [17].

$$(\mathbf{JPG}\ x_1..x_n\ p\ q) \equiv (\mathbf{MB}\ x_1..x_n\ \neg p) \wedge (\mathbf{MG}\ x_1..x_n\ p) \wedge$$

$$\{\mathbf{UNTIL}\ \ [(\mathbf{MB}\ x_1..x_n\ p) \vee (\mathbf{MB}\ x_1..x_n\ \Box\neg p) \vee (\mathbf{MB}\ x_1..x_n\ \neg q)]\ \ (\mathbf{WMG}\text{[3]}\ x_1..x_n\ p)\}$$

It is apparent from this definition that the JPG becomes invalid when any of the team members $x_1..x_n$ leaves the team. Even the individual commitments implied by this JPG is allowed to be dropped because it is now impossible to establish any of the mutual beliefs in the *until* clause.

We introduce below the notion of joint commitment to a team rather than to the individuals that constitute the team. This commitment allows the members of a team to change but the same team continues to exist as long as the current members share the team goal and the required mental state.

4.1 Team Commitment

We use a special notation to represent propositions that refer to a group whose members satisfy certain properties. Let α_x be a formula with free variable x. We call $\alpha_x(\pi)$ a quasi-formula where π is a one-place predicate. If α is a formula or quasi-formula with respect to π, we use the notation $\langle\alpha\rangle$ to denote a formula defined by the following rule:

(1) If α is a formula, then $\langle\alpha\rangle = \alpha$

(2) If α is a formula with respect to π, and z does not appear in α, and $\alpha_z(\pi)$ is a formula, then $\langle\alpha\rangle = \forall z.\ \pi(z) \supset \alpha_z(\pi)$

For example,

$$\langle \text{BEL x } p\rangle = (\text{BEL x } p)$$

$$\langle \text{BEL } \Pi\ p\rangle = \forall z\ \ (\Pi(z) \supset (\text{BEL z } p))$$

$$\langle \text{BEL } \lambda z.(\text{member z T})\ \ p\rangle = \forall z\ \ ((\text{member z T}) \supset (\text{BEL z } \ p))$$

Using this notation, we redefine the alternating mutual belief concept [7] to allow groups as well as individuals.

Definition 4.1: $(\mathbf{ABEL}\ \text{n}\ \tau_1\ \tau_2\ p) \equiv \langle\mathbf{BEL}\ \tau_1\ \langle\mathbf{BEL}\ \tau_2 \ldots \langle\mathbf{BEL}\ \tau_1\, p\rangle \ldots\rangle$

A group τ has a group alternating belief about *p* up to n levels when everybody in the group believes that everybody in the group believes that everybody in the group believes *p* and so on up to depth n.

Definition 4.2: $(\mathbf{GABEL}\ \text{n}\ \tau\ \ p) \equiv (\mathbf{ABEL}\ \text{n}\,\lambda z.(\text{member z } \tau)\ \ \lambda y.(\text{member y } \tau)\ \ p)$

For example, $(\mathbf{GABEL}\ 3\ \text{T}\ \ p)$

$$= \langle\mathbf{BEL}\ \lambda z.(\text{member z T})\ \langle\mathbf{BEL}\ \lambda y.(\text{member y T})\ \langle\mathbf{BEL}\ \lambda z.(\text{member z T})\ p\rangle\rangle\rangle$$

$$= \forall z.(\text{member z T}) \supset (\mathbf{BEL}\ \text{z}\ \forall y.(\text{member y T}) \supset (\mathbf{BEL}\ \text{y}\ \forall w.(\text{member w T}) \supset (\mathbf{BEL}\ \text{w}\ p)))$$

We used variable w instead of z in the innermost formula to clarify that there is no binding between the free variables used in the quantifiers.

A group member has unilateral group mutual belief about *p* when it believes that the group has a mutual belief about *p*. It believes *p* and believes that everybody believes that everybody believes *p* and so on for all levels.

---

[3] It denotes weak mutual goal.

Definition 4.3: $(\mathbf{BGMB}\ z\ \tau\ p) \equiv (\mathbf{BEL}\ z\ [p \wedge \forall n\ (\mathbf{GABEL}\ n\ \tau\ p)])$

A group τ mutually believes *p* when all the members of the group have unilateral group mutual belief about *p*.

Definition 4.4: $(\mathbf{GMB}\ \tau\ p) \equiv \langle \mathbf{BGMB}\ \lambda z.(\text{member}\ z\ \tau)\ \ \tau\ \ p\rangle$

A group τ has *p* as a group mutual goal if the group mutually believes that everybody in the group has *p* as an individual goal.

Definition 4.5: $(\mathbf{GMG}\ \tau\ p) \equiv (\mathbf{GMB}\ \tau\ \langle \mathbf{GOAL}\ \lambda z.(\text{member}\ z\ \tau)\ \ p\rangle)$

Teams are specialized groups with certain constraints. A member z of a team τ is said to have a weak team goal with respect to the team when the following conditions hold:

(1) If the team member believes that the goal *p* is not yet achieved, it has an individual goal to eventually bring about *p*.

(2) If it believes that the goal has been achieved, is impossible to achieve, or is irrelevant then it has an individual goal to bring about the corresponding group mutual belief.

(3) If it believes that the group has a new member, or an existing member is no longer in the group then it has an individual goal to bring about the corresponding group mutual belief.

Definition 4.6:

$$\begin{aligned}(\mathbf{WTG}\ z\ \tau\ p\ q) \equiv\ & \{\, [\neg(\mathbf{BEL}\ z\ p) \wedge (\mathbf{GOAL}\ z\ \Diamond p)] \vee \\ & \ [(\mathbf{BEL}\ z\ p) \wedge (\mathbf{GOAL}\ z\ \Diamond(\mathbf{GMB}\ \tau\ p))] \vee \\ & \ [(\mathbf{BEL}\ z\ \Box\neg p) \wedge (\mathbf{GOAL}\ z\ \Diamond(\mathbf{GMB}\ \tau\ \Box\neg p))] \vee \\ & \ [(\mathbf{BEL}\ z\ \neg q) \wedge (\mathbf{GOAL}\ z\ \Diamond(\mathbf{GMB}\ \tau\ \neg q))]\, \} \wedge \\ & \{\forall y\ \mathbf{Q}_y \supset (\mathbf{GOAL}\ z\ \Diamond(\mathbf{GMB}\ \tau\ \neg(\text{member}\ y\ \tau))\} \wedge \\ & \{\forall y\ \mathbf{P}_y \supset (\mathbf{GOAL}\ z\ \Diamond(\mathbf{GMB}\ \tau\ (\text{member}\ y\ \tau))\}\end{aligned}$$

where, $\mathbf{Q}_y = (\mathbf{BEL}\ z\ \exists e\{\mathbf{DONE}\ (\text{member}\ y\ \tau)?;e;\neg(\text{member}\ y\ \tau)?\})$

$\mathbf{P}_y = (\mathbf{BEL}\ z\ \exists e\{\mathbf{DONE}\ \neg(\text{member}\ y\ \tau)?;e;(\text{member}\ y\ \tau)?\})$

and *e* is a singleton event.

A team τ has *p* as a weak team mutual goal if the team mutually believes that everybody in the team has *p* as a weak team goal.

Definition 4.7: $(\mathbf{WTMG}\ \tau\ p\ q) \equiv (\mathbf{GMB}\ \tau\ \langle \mathbf{WTG}\ \lambda z.(\text{member}\ z\ \tau)\ \tau\ \ p\ q\rangle)$

A team τ has a team persistent goal *p* when the following conditions hold:

(1) The team mutually believes that the goal *p* has not been achieved.

(2) The team has a mutual goal to achieve the goal *p*.

(3) The team has *p* as a weak team goal until there is team mutual belief about the completion, feasibility or irrelevance of *p*.

Definition 4.8:

$$\begin{aligned}(\mathbf{TPG}\ \tau\ p\ q) \equiv\ & (\mathbf{GMB}\ \tau\ \neg p) \wedge (\mathbf{GMG}\ \tau\ p) \wedge \\ & \{\mathbf{UNTIL}\ [(\mathbf{GMB}\ \tau\ p) \vee (\mathbf{GMB}\ \tau\ \Box\neg p) \vee (\mathbf{GMB}\ \tau\ \neg q)]\ \ (\mathbf{WTMG}\ \tau\ p\ q)\}\end{aligned}$$

A team jointly intends an action *a* if there is a team commitment to do the action while believing that the team is going to do the action next[4].

Definition 4.9: (**TI** τ *a* q) $\equiv$ [**TPG** τ (**DONE** τ (**GMB** τ (**HAPPENS** τ *a*))?;a) *q*]

The joint commitment and the joint intention defined above are with respect to the team as an entity rather than the individuals that constitute the team. This allows a team having these team commitments to continue as long as there is no mutual belief about the completion, impossibility or irrelevance of the team goal even if some of the team members leave and new members join the team.

It is obvious from the definitions of TPG and TI that teamwork requires establishment of mutual beliefs between a group of agents at various stages. It has been shown that the sending of an inform, followed by an acknowledgement that the prior inform was believed brings about mutual belief between two agents by default [4][5]. Similarly, group communication mechanisms such as broadcast and multicast, followed by appropriate acknowledgements can be used to establish mutual belief in a group by default.

We use the above formalism to establish the theorems stated section 3.2.

4.2 Commitments of the AAA Brokers

We first state the properties of the implementation that are needed to establish some of the theorems.

Proposition 4.1: An agent is connected with the broker team if it is connected with any member of the broker team.

$\forall$y (agent y) $\supset$ [(connected y T) $\equiv$ $\exists$z (member z T) $\wedge$ (connected y z)]

Proposition 4.2: An agent registers with the broker team by registering with any member of the broker team.

$\forall$y (agent y) $\supset$ [(**DONE** y (register y T)) $\equiv$ $\exists$z (member z T) $\wedge$ (**DONE** y (register y z))]

Proposition 4.3: An agent unregisters with the broker team by unregistering with any member of the broker team.

$\forall$y (agent y) $\supset$ [(**DONE** y (unregister y T)) $\equiv$ $\exists$z (member z T) $\wedge$ (**DONE** y (unregister y z))]

Proposition 4.4: When a broker teammate is no longer accessible to a broker, it believes that all the agents registered with that broker are disconnected from the team.

$\forall$x, y [(member x T) $\wedge$ (member y T) $\wedge$ $\neg$(accessible x y)
$\supset$ (**BEL** x $\forall$z.(connected z y) $\supset$ $\neg$(connected z T))]

---

[4] More accurately, TI is the commitment to do an action while believing throughout that the team is doing the action.

[5] Note that a circular data structure enables the compact representation of mutual belief [4].

The theorems in section 3.2 follow from either the AAA mission statement or proposition 3.1. We give detailed proof for one of the main theorems, and the other theorems can be proved using a similar approach or they directly follow from the basic theorems on commitments.

4.2.1 Team Commitment to Connect to a Disconnected Agent Follows From the AAA Mission Statement

Theorem 3.4 states that *when an agent registered with the broker team gets disconnected, the brokers in the team have a team commitment to connect to that agent.* More formally,

$$\forall y\ [(\text{agent } y) \wedge (\text{registered } y\ T) \wedge \neg(\text{connected } y\ T) \supset (\textbf{TPG}\ T\ (\textbf{DONE}\ (\text{connect } y\ T))\ q]$$

where q is some escape condition.

Proof: Given that the antecedent is true, we want to show that the consequent follows from the AAA mission statement.

Expanding the mission statement using the definition of TI, we get the following TPG

$\xi = [\textbf{TPG}\ T\ \{\textbf{DONE}\ \gamma?;\ a(y)\}\ (\text{registered } y\ T)]$ where,

$\gamma = [\textbf{GMB}\ T\ (\textbf{HAPPENS}\ T\ a(y))]$, and

$a(y) = (\textbf{WHILE}\ (\text{registered } y\ T)\ \ \textbf{DO}\ [\text{if } \neg(\text{connected } y\ T)\ \textbf{THEN}\ (\text{connect } y\ T)])$

$= d(y);\ (\text{registered } y\ T)?;\ [\neg(\text{connected } y\ T)?;\ (\text{connect } y\ T)\ \mid\ (\text{connected } y\ T)?];\ e(y)$

where, *d*(y) denotes the previous iterations that have already been done

e(y) denotes the remaining iterations

Substituting for the action expression, we get

$$\xi = [\textbf{TPG}\ T\ \{\textbf{DONE}\ \gamma?; d(y); (\text{registered } y\ T)?; \neg(\text{connected } y\ T)?;\ (\text{connect } y\ T);\ e(y)\}(\text{registered } y\ T)]$$

We are interested in the iteration in which the antecedent becomes true.

$\therefore$ The action subsequence $\gamma?;\ d(y);\ (\text{registered } y\ T)?;\ \neg(\text{connected } y\ T)?$ has just been done.

On similar lines of reasoning as in [8] theorem 2, we can establish the following theorem,

$$\models (\textbf{TPG}\ T\ (\textbf{DONE}\ a;b)) \wedge (\textbf{GMB}\ T\ (\textbf{DONE}\ a) \wedge (\textbf{GMB}\ T\ \neg(\textbf{DONE}\ b))$$
$$\supset (\textbf{TPG}\ T\ (\textbf{DONE}\ b)\ [\textbf{TPG}\ T\ (\textbf{DONE}\ a;b)])$$

Note that this theorem will hold even if the team membership has changed after the initial action subsequence was done.

$\therefore$ Using this theorem on action sequences, we get

$\xi \supset \chi$

where, $\chi = [\textbf{TPG}\ T\ \{\textbf{DONE}\ (\text{connect } y\ T);\ e(y)\}\ \xi]$

Now, assume that the brokers have successfully acted rationally on the commitment that follows from theorem 3.3. Therefore, it follows that we can assume the present members of the broker team to know when the action ¬(connected y T)? has been done and so they know when the next action (connect y T) is to start. Hence, the brokers in the team now have a team commitment to do (connect y T) with respect to the team commitment to do (connect y T) followed by *e*(y).

$\therefore \chi \supset (\textbf{TPG}\ T\ [\textbf{DONE}\ (\text{connect y T})]\ \chi)$

This proves the desired result that *when an agent registered with the broker team gets disconnected, the brokers in the team have a team commitment to connect to that agent*.

The remaining theorems from section 3.2 have been established using a similar approach. Theorem 3.8 and Theorem 3.9 follow from proposition 3.1, and all the other theorems are a consequence of the AAA mission statement.

In summary, we showed that teams formed using joint commitment (JPG) cease to exist when team members change. Therefore, we introduced the notion of team commitment (TPG) such that the teams formed using TPG exist independent of the team members. Thereafter, we showed how one of the main theorems about team commitments follows from the AAA mission statement. The other theorems can be proved similarly, thus establishing that the fault-tolerance in AAA is a consequence of the teamwork specification that AAA was built to satisfy.

5 RELATED WORK, DISCUSSION, AND FUTURE WORK

Jennings has used teamwork for cooperative problem solving and shown that teams as a whole waste fewer resources and are more robust than self-interested agents [13]. This work does not explicitly address the problem of recovery from failures that we are concerned with in this paper. Tambe and Kaminka have used teamwork model for failure diagnosis and recovery and for achieving cooperative behavior between teams of helicopters in dynamic, unpredictable environments [14, 20]. However, the helicopter teams are not persistent in the sense that if one of the helicopters gets shot down, the original joint persistent goal ceases to exist. This situation may lead to undesired behavior as the surviving team members can now drop their individual goals because of the impossibility of establishing mutual belief with the helicopter that was shot down.

Hägg [11] and Klein [15] use fault-tolerance approaches that do not make use of the teamwork theory. These work use external sentinels to monitor the progress of a multi-agent system, diagnose problematic situations, and initiate recovery actions. In teamwork-based approaches, the problem solving agents themselves participate in the fault-tolerance process. Moreover, these approaches are essentially centralized approaches as whereas those based on the teamwork theory are decentralized.

Tambe and Zhang have used the notion of persistent teams in [21] wherein they consider the problem of optimizing between the long-term and short-term goals of persistent teams. They also point out the need for a formalism that explains the maintenance of a team identity in the face of changing membership. We believe that the concept of team persistent goal introduced in this paper provides this much needed formalism. Tambe and Zhang also point out the issue of persistence of a team beyond specific temporary objectives. We believe that maintenance goals formalize this idea and teams based on maintenance goals will continue to exist as long as they keep the maintenance goal. We have briefly touched upon maintenance goals in this paper [section 3.2] for implementing and explaining the self-healing process in an AAA-based multi-agent system. A formulation of maintenance goals for individuals and for teams will be discussed in another paper.

The recovery schemes presented in this paper guarantee recovery of the system connectivity but it is left to an agent to guarantee idempotent behavior when a message that was presumably lost due to broker failure is resent by the AAA agent shell. Moreover, recovery of the ongoing conversations is required for a complete recovery. We are also investigating general plan based approaches for implementing teamwork wherein generic BDI agents can establish and execute team commitments specified at run time.

6 SUMMARY

We introduced the Adaptive Agent Architecture (*AAA*), which is a brokered multi-agent infrastructure for building robust multi-agent systems. An AAA-based multi-agent system can recover from broker failures that arise from problems such as machine crashes, communication breakdown, and death of a broker process. We presented the specification of teamwork that the AAA implements and argued that the fault-tolerant behavior of AAA-based multi-agent systems is a consequence of this teamwork specification. This specification says that whenever an agent registers with an AAA broker team, the brokers have a team intention of connecting with that agent, if it ever disconnects, as long as it remains registered with the broker team. We showed that the required team and individual commitments could be derived from this specification. We conclude that teamwork can be used to design robust brokered multi-agent system architectures that can recover from broker failures.

We also formalized the notion of persistent teams that can continue to exist even when the team members change. Such teams have a team persistent goal wherein the team members are committed to the team as an entity rather than to the individuals that constitute the team.

The AAA is currently used in the Quickset [9] multi-agent system and in the DARPA CoABS program [10] where it is used to glue together three different technology integration experiments.

7 REFERENCES

[1] P. A. Bernstein and E. Newcomer, editors. *Principles of Transaction Processing*. High Availability, chapter 7, 1997.

[2] K. P. Birman, editor. *Building Secure and Reliable Network Applications*. Part III, Reliable Distributed Computing, chapters 12-26, 1996.

[3] L. Chen and A. Avizienis. N-version programming: A fault-Tolerance Approach to Reliability of Software Operation. *In Digest of Papers of the 8th Annual International Conference on Fault-Tolerant Computing,* Toulouse, France, 1978. As referred in [11].

[4] P. R. Cohen. On Knowing What to Say: Planning Speech Acts. *Ph.D. Thesis*, Department of Computer Science, University of Toronto, 1978.

[5] P. R. Cohen, A. Cheyer, M. Wang, and S. C. Baeg. An Open Agent Architecture. *AAAI Spring Symposium*, pages 1-8. March 1994

[6] P. R. Cohen and H. J. Levesque. Intention is Choice with Commitment. *Artificial Intelligence*, Volume 42, pages 213-261, 1990.

[7] P. R. Cohen and H. J. Levesque. Rational Interaction as the Basis for Communication. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors. *Intentions in Communication,* chapter 12, System Development Foundation Benchmark Series, Bradford Books, MIT Press, 1990.

[8] P. R. Cohen and H. J. Levesque. Confirmations and Joint Action. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, Morgan Kaufman Publishers, Inc., San Mateo, California, August 1991, pages 951-957.

[9] P. R. Cohen, M. Johnston, D. McGee, S. Oviatt, J. Pittman, I. Smith, L. Chen, and J. Clow. QuickSet: Multimodal Interaction for Distributed Applications, *Proceedings of the Fifth International Multimedia Conference (Multimedia '97),* ACM Press, pages 31-40.

[10] Control of Agent Based Systems. http://coabs.globalinfotek.com

[11] S. Hägg. A Sentinel Approach to Fault Handling in Multi-Agent Systems. In *Proceedings of the 2nd Australian Workshop on Distributed AI,* Cairns, Australia, 1997.

[12] C. Heckman and A. Roetter. Designing Government Agents for Constitutional Compliance. *Autonomous Agents '99,* Seattle, 1999.

[13] N. R. Jennings. Controlling Cooperative Problem Solving in Industrial Multi-Agent Systems using Joint Intentions. *Artificial Intelligence.* 75(2), pages 195-240, 1995.

[14] G. A. Kaminka and M. Tambe. What is Wrong With Us? Improving Robustness Through Social Diagnosis. *In Proceedings of the 15th National Conference on Artificial Intelligence,* 1998.

[15] M. Klein and C. Dellarocas. Exception Handling in Agent Systems. *Autonomous Agents '99,* Seattle, 1999.

[16] S. Kumar and P. R. Cohen. Towards a Fault-Tolerant Multi-Agent System Architecture. *In submission,* 1999.

[17] H. J. Levesque, P. R. Cohen, and J. Nunes. On Acting Together. In *Proceedings of the National Conference on Artificial Intelligence,* Morgan Kaufmann Publishers, Inc., California, 1990.

[18] D. L. Martin, A. J. Cheyer, and D. B. Moran, "The open agent architecture: A framework for building distributed software systems," *Applied Artificial Intelligence*, Vol.13, pp. 91-128, 1999.

[19] K. Sycara, K. Decker, and M. Williamson. Middle-Agents for the Internet. *Proceedings of IJCAI-97*, 1997.

[20] M. Tambe. Teamwork in Real-world, Dynamic Environments. *In Proceedings of the 2nd International Conference on Multi-Agent Systems*, AAAI Press, Menlo Park, California, 1996.

[21] Tambe, M. and Zhang, W. 1998 Towards Flexible Teamwork in Persistent Teams. *In Proceedings of the 3rd International conference on multi-agent systems*, 1998.