What is an Agent?

Technical Report #6/99

(Originally written in 1997)

James Ingham

Centre for Software Maintenance

University of Durham

Durham

Introduction

This document aims to define the term agent with respect to the field of software agency. Agents will be defined in terms of other work already carried out in this field. This is a highly heterogeneous field with many conflicting viewpoints on what actually constitutes an agent. It should be of no surprise if the definition in this report also directly conflicts with other agent definitions. This point is made well by Krogh [RightsOfAgents, 1996] :-

> *"... What's an agent anyway?*
> *Agents may be many things.Attempts to find one central common denominator of operative or theoretical conceptions of agents in recent publications on the topic ... will probably fail"*

This is a working document so therefore many of these definitions may change if new ideas becomes known either from other sources or from the furthering experience in this field.

Agent Definitions

This has been included to help a reader who is new to the field of agents. However, as previously stated, the field of agents has many conflicting definitions so please do not be surprised if you have read different or conflicting ones.

Agent definition 1. [AgentSurvey, 1994]

> *"**AGENT** - We distinguish two usages. The most general usage is to mean an **AUTONOMOUS**, self-contained, **REACTIVE**, **PRO-ACTIVE**, computer system, typically with a central focus of control, that is able to communicate with other **AGENTS** via some **ACL**. A more specific usage is to mean a computer system that is either conceptualised or implemented in term of concepts more usually applied to humans (such as beliefs, desires and intentions*[**BDI ARCHITECTURE**]*)...* [also connected to] ***INTENTIONAL SYSTEM***[s]*"* [AgentSurvey, 1994]

This paper defines autonomy as

> *"...The assumption that, although we generally intend **AGENTS** to act on our behalf, they nevertheless act without direct human or other intervention, and have some kind of control over their internal state..."* [AgentSurvey, 1994]

Reactive as

> *"...Capable of maintaining an ongoing interaction with the environment, and responding in a timely fashion to changes that occur in it. Note that the term is now widely used to mean a system that includes no symbolic representation or reasoning: such an AGENT does not reflect on the long-term effects of its actions, and does not consider the co-ordination of activity with other agents. Thus, a **REACTIVE AGENT** will always respond in a timely fashion to external stimulus..."* [AgentSurvey, 1994]

Pro-active as

> *"...Capable of taking the initiative; not driven solely by events, but capable of generating goals and acting **RATIONALLY** to achieve them."* [AgentSurvey, 1994]

ACL as

> *"Agent Communication Language - a common language used by AGENTS for communicating with one another .."* [AgentSurvey, 1994]. An example of **ACLs** are **KQML** combined with **KIF** and a set of **Ontologies.**

BDI ARCHITECTURE as

> *"An architecture containing explicit representations of **beliefs**, **desires** and **intentions**... **Beliefs** are generally regarded as the information an **agent** has about its environment, which may be false; **desires** are those things that the **AGENT** would like to see achieved - **desires** need not be consistent, and we do not expect an **AGENT** to act on all its desires - and finally, **intentions** are those things the **AGENT** is either committed to doing (intending to) or committed to bringing about (intending that)."* [AgentSurvey, 1994]. Some different examples of agent architectures can be found in many books including [IntelligentAgents, 1995], [AgentsBreakingAway, 1996] , [IAgentsConferenceI, 1994] and [IAgentsConferenceII, 1995]

Agent definition 2 [Agent FAQ, 1995]

> *"Autonomous:*
> *An agent is able to take initiative and exercise a non-trivial degree of control over its' own actions.*
>
> *Goal-Oriented:*
> *An agent accepts high-level requests indicating what a human wants and is responsible for deciding how and where to satisfy the request.*
>
> *Collaborative:*
> *an agent does not blindly obey commands, but has the ability to modify requests, ask clarification questions, or even refuse to satisfy certain requests*

Flexible:
the agent's actions are not scripted; it is able to dynamically choose which actions to invoke, and in what sequence, in response to the state of its external environment.

Self-starting:
unlike standard programs which are directly invoked by the user, an agent can sense changes to its environment and decide when to act.

Temporal continuity:
agents are continuously running processes, not "one-shot" computations that map a single input to a single output, then terminate.

Character:
an agent has a well-defined believable "personality and emotional state"

Communicative:
the agent is able to engage in complex communication with other agents, including people, in order to obtain information or enlist their help in accomplishing its goals.

Adaptive:
the agent automatically customises itself to the preferences of its user based on previous experience. The agent also automatically adapts to changes in its environment.

Mobile:
an agent is able to transport itself from one machine to another and across different system architectures and platforms."

As can be seen, these are not all applicable to every type of program. To re-iterate, an agent does not have to be all of these categories, rather these are recognised agent categories. For the purpose of this research the definition of **character** is not really applicable, although there are systems such as one where an agent sits in a MUD (Multi User Dungeon) and offers advice and also interacts with other players. However it is debatable whether it has any more personality than the average C-shell script. No further argument is offered upon these points as this quote was intended as an introduction only and the question is more in the realms of philosophy than Computer Science.

The following properties of agents are often said to be defining properties of agents.

- Autonomous
- Pro-Active and/or Reactive
- Temporally continuous

The purpose here is to define exactly what is meant, where possible in a more detailed manner than the original terms used. The semantics of these definitions may be similar to other papers.

Conflict 1 - Agents are autonomous.

Using the definitions of autonomy highlighted above, one can have a basic concept of what an autonomous agent is. However the extent to which an agent is autonomous is unclear.

[AgentOrientedProgramming, NoDate] states ...

> *"... Most often, when people in AI use the term "agent", they refer to an entity that functions continuously and autonomously in an environment in which other processes take places and other agents exist. This is perhaps the only property that is assumed uniformly by those in AI who use the term. The sense of "autonomy" is not precise, but the term is taken to mean that the agent's activities do not require constant human guidance or intervention."*

There exists debate to what extent agents have to be autonomous, and what exactly is meant by autonomy anyway. At one extreme are agents who will simply perform their tasks with no interaction of any type between agents (human or otherwise). These could be said to be highly autonomous because all decisions are made by that agent only. On the other extreme is an agent which asks for advice when reaching any form of decision. Neither of these are of general practical use.

Definition of an Agent with Respect to Autonomy.

> An entity having the property of controlling its' own action independent of other entities unless it desires to communicate with other entities. It typically does not refer to a second party unless it possesses insufficient knowledge to perform that particular action itself.

To reiterate, an agent controls itself unless it chooses to refer to another entity, be it an agent or otherwise. This is analogous to a human being making its' own decisions unless it decides that someone else knows better and asks their opinion instead. The other point made is that there is little practical value of the agent referring to another source of information if the agent knows enough without asking it. Of course this depends on whether the agent knows that it knows enough to perform the action which depends how it models itself and also its' environment.

Conflict 2 - Agents must be Pro-Active and Reactive

There are a number of supporters for the argument that agents must not just **React** to the environment. Instead at some point in the agent's "life" it must take control and behave **Pro-Actively** (e.g. as in [AgentOrProgram, NoDate] and also as the "weak notion of agency" in [IntelligentAgents, 1995, pp4-5]). This implies a combination of an event driven (**Reactive**) and a goal driven (**Pro-Active**) control system rather than simply an event driven one. However other authors claim that **Pro-Active** agents are just one of many different agent system types. For example [Paper12-PAAM96] classifies agents as follows :-

1 Basic software agents

These are black box with no intelligence/adaptability/communication and are somewhat similar to a background process.

2 Reactive Agents

These are Data/response driven and not pro-active. Such an agent may reason (to an extent) and be adaptive. They generally have little or no social organisation.

3 Deliberate

These agents are Goal driven only (if they are purely deliberate). They must be able to reason and may be able to plan. This type of agent is described as Pro-active.

4 Hybrid Agents

This type of agent can have a mixture of types (e.g. deliberate and reactive is the usual combination)

As a slight aside [Minsky, 1985] does not agree that **2) Reactive agents** can only reason to an extent. In fact some AI researchers, including Minsky, hypothesise that an intelligent system is simply an amalgamation of many smaller simple systems and any intelligence is **Emergent behaviour** of the complex system. This is another question which is probably discussed more frequently (and relevantly) by philosophers than Software Engineers.

In [AgentOrProgram, NoDate], which "*propose*[s] *a formal definition of an autonomous agent which clearly distinguishes a software agent from just any program*" they define the following.

> *"An **autonomous agent** is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future"*

This implies some degree of **Pro-activeness** (e.g. it pursues its' own agenda). For the purposes of this research this is not a defining attribute of an agent. This paper has made artificial criteria for "agenthood" to give a clear distinction between agents and concurrent processes that react to input e.g. device drivers) where perhaps no such "clear" distinction can be made. The distinction between the agents and non-agent systems will be discussed later.

Definition of an Agent with respect to Reactive/Pro-Activeness.

> The agent may **Reactive** and **ProActive**, **Reactive** only or **ProActive** only. That is to say that an agent could have a purely event driven system, such as if environment constraint x holds do y, or purely goal driven system such as "I desire to do x, which requires a plan with stages y and z in it" but more typically will require a combination of the two approaches.

The other combination (not either **Reactive** or **ProActive**) was omitted since a useful system with such attributes could not be found or percieved. Such a system would not do anything at all!

3) Agents are Temporally Continuous

The attribute of temporal continuity with respect to an agent is defined in [AgentOrProgram, NoDate] as being

> "[a temporally continuous agent is]... *a continuously running process"*

There is no reason why an agent could not go to sleep for a specified amount of time or terminate. If the system uses mobile agents then there is definitely a time when the agent is actually in transit and not actually present on any machine. The [AgentOrProgram, NoDate] definition of a temporally continuous agent seems to preclude mobile agents.

The concept that the agent must be able to maintain its' execution state is prefered here. As has been pointed out in [AgentsMailingList, 1996, 1997] it may be unnecessary to be able to suspend execution at every state, rather that the state may be stored at certain points in an agents execution and the agent re-assembled at another point with the same state. To re-iterate the point that temporal continuity concerns more of continuity of the program execution state rather than the actual continuity of the actual program which of course is an illusion in a uni-processor system anyway.

Definition of an Agent with respect to Temporal continuity

The agent may suspend, terminate or continue execution but the state of the agent if suspended may be recorded in some way and recreated when the agent is re-animated.

Other attributes of agents

The other attributes of agents are generally agreed to be optional. This is a non-exhaustive list of common agent types. Certain agent attributes may be known under different names.

Adaptation

Agents may in some cases attempt to adapt themselves to better suit their new or changing environment or to deal with new or changing goals. Considerations of agent adaptation include what algorithm can be used to adapt the agent, what is the maximum amount of change forseen in the system, will the adaptation algorithm change as well, how is the system to stop evolution going out of control, and how to detect and deal with an adaptation which does not create the desired effect.

Communication

Agents may in some cases communicate to achieve some goal or due to some event. Considerations of inter-agent communication include what protocol to use, how to describe a domain in terms an agent from another domain can understand, and how efficient the communication method is.

Co-operation

Agents may in some circumstances work together to achieve some goal or due to some event. Considerations of co-operative systems include whether each agent will need to benefit from the situation in order to cooperate, whether agents will cooperate without necessarily obtaining any gain for such interaction, what to do if an agent is taking advantage of other agents generosity and whether the agent system can cooperate in some way with non-agent specific systems.

Learning

Agents may in some circumstances learn from past occurrences in the environment to predict the future and in some cases (pro)actively affect the environment. Considerations of learning systems include what algorithm will the agent use to learn (will the learning algorithm learn as well) and what will be measured / modelled and is this sufficient to understand the environment.

Mobile

Agents may in some circumstances move from one system to another. Considerations of mobile agent systems include how will the agent ensure that it can execute on the new system (common environment, common language, or standards), and whether it can suspend execution at every possible state or only for certain states [AgentsMailingList, 1996, 1997].

Negotiating

Agents may in some cases negotiate, usually concerning some form of co-operation. Considerations of negotiation include what model will the negotiation process use, what is done if there is no satisfactory answer and how will arbitration be performed when two agents want different but potentially reconcilable things.

Simulating

Agents may in some cases simulate a system or element of a system. This may be a simulated character in a computer game or an agent wrapper for a database, in which case the agent can be viewed as simulating a database. Considerations of simulation include levels of realism (i.e. exact / identical functionality, e.g. database wrapper, in comparison to a rough approximation of behaviour, e.g. "intelligent" character in a computer game) and how applicable is the agent paradigm in a system of this type.

Definition of an agent.

This section gives the definition of an agent that is suitable for this research.

Working Definition

Using what has been constructed so far an agent is the following.

> An entity having the property of controlling its' own action independent of other entities unless it desires to communicate with other entities. It typically does not refer to a second party unless it possesses insufficient knowledge to perform that particular action itself. The agent may **Reactive** and **ProActive**, **Reactive** only or **ProActive** only. That is to say that an agent could have a purely event driven system, such as if environment constraint x holds do y, or purely goal driven system such as "I desire to do x, which requires a plan with stages y and z in it" but more typically will require a combination of the two approaches. The agent may suspend, terminate or continue execution but the state of the agent if suspended may be recorded in some way and recreated when the agent is re-animated.

The agent may also be any or all or none of the following attributes

- Communicative
- Co-operative
- Learning

- Mobile
- Negotiating
- Simulating

Some common assumptions of intelligent agents are

> *"...**Veracity** ... - the assumption that an agent will not knowingly communicate false information [Galliers, 1988b, pp159-164].*
>
> ***Benevolence**... - the assunption that agents do not have conflicting goals, and that every agent will therefore always try to do what is asked of it [Rosenschein and Genesereth, 1985, p91].*
>
> ***Rationality** ... - (crudely) the assumption that an agent wiill act in order to achieve its' goals and will not act in such a way as to prevent its' goals being achieved - at least insofar as its' beliefs permit [Galliers,1988b. pp49-54].*

Whether these are useful general assumptions is a contentious issue (i.e. as in [AgentMailingList, 1996-1997], especially concerning security and the **Benevolent** agent assumption.

Problem with the definition

The problem is that by this definition every program that was ever written is an agent, including "Hello World". This is where the meaning of an agent becomes clear. It cannot specifically be defined upon functionality. It has been stated more the once in the Agents mailing list [AgentsMailingList, 1996, 1997] that whatever is done with agent programming could have been done by a programmer 10 years ago. Therefore looking for the difference between agent systems and non-agent systems in the functionality seems to be a futile course. The only difference between Agent Orientated programming and any previous form of programming seems to be the paradigm and the structure of the resulting system.

The following reasons were found by [Hall&Shahmehri, 1996] for using Agent systems

- *"It reduces the complexity of the implementation, where each part can be constructed separately.*

- *It gives natural modularisation* [or can seem to be the most natural way of modelling certain classes of problems]
- *Several similar agents can operate simultaneously on the same piece of information. This supports a multi-user environment.*
- *It supports interoperability between different service functions* [assuming there is some organisation, different service functions should be relatively interchangable.]
- *It makes it possible, if needed, to split the workload between several agents*
- *It makes it easy to add new services as new agents."*[Hall&Shahmehri, 1996]

Another advantage with agents is that they are suitable for working in systems where parts of the system may be isolated from other parts of the system, hence justifying the decentralised control structure. (also where a centralised system would be too over-loaded / vulnerable / fragile)

Many of the above claims have not yet been proved (i.e. agents reduce complexity and make it easy to add new functionality) but perhaps in the course of this research some may be tested.

Conclusions

The main aim of this report is to define the term agent for the purposes of establishing a vocabulary on which to discuss further work in the field of Self Fixing Software which may be conducted using agents. An attempt has been made to explain the present confusion due to multiple (and inevitably conflicting) definitions of agents. It has also been decided and argued for a number of points of contention in the field of agent technology which has resulted in the view of what an agent is for the purposes of the **Self Fixing Software** project.

Thus the working definition of an agent is:-

> An entity having the property of controlling its' own action independent of other entities unless it desires to communicate with other entities. It typically does not refer to a second party unless it possesses insufficient knowledge to perform that particular action itself. The agent may **Reactive** and **ProActive**, **Reactive** only or **ProActive** only. The agent may suspend, terminate or continue execution but the state of the agent if suspended may be recorded in some way and recreated when the agent is re-animated.

Other possible characteristics that agents could possess include:-

- Communicative agents
- Co-operative agents
- Learning agents
- Mobile agents
- Negotiating agents
- Simulating agents

It has been claimed here that it is futile to define an agent in terms of functionality as one could implement them in an object orientated, modular or plainly unstructured system. Instead it has been claimed that the defining quality of an agent is the design considerations and structure of the resulting system. It has also been claimed that for certain applications this paradigm and resulting system structure is more easily applicable than a different methodology.

However, an important point is that not every application should use multi-agent / agent technology. If the system to be created does not have at least some characteristics which would be appropriate to an agent system, then there will probably be a much better defined existing solution. To re-iterate, agents are not applicable (or at least any more applicable than a conventional solution) to all situations.

References

[AgentsBreakingAway, 1996]	“Agents breaking Away”, Walter Van de Velde and John W. Perram (Eds.) 7th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, MAAMAW ‘96, Eindhoven, The Netherlands, Jan 1996 Proceedings. Verlag Springer ISBN 3-540-60852-4
[AgentFAQ, 1995]	Agent FAQ at the address http://www.ee.mcgill.ca/~belmarc/agent_faq.html
[AgentsMailingList, 1996, 1997]	A common interest mailing list from the internet.
[AgentOrProgram, NoDate]	"Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents" Stan Franklin and Art Graesser, Institute for Intelligent Systems, University Of Memphis.
[AgentSurvey, 1994]	Wooldridge, Michael and Nicholas R. Jennings "Agent Theories, Architectures and Languages: A Survey" in Wooldridge and Jennings Eds., Intelligent Agents, Berlin:Springer-Verlag, 1-22
[Galliers, 1988b]	“A Theoretical Framework for Computer Models of Cooperative Dialogue, Acknowledging Multi-Agent Conflict.” PhD thesis, Open university, as referenced in [IntelligentAgents, 1995]
[IAgentsConferenceI, 1994]	“Intelligent Agents”, ECAI-94 Workshop on Agent Theories, Architectures, and Languages. Amsterdam, The Netherlands, August 1994 Proceedings. Springer Verlag ISBN 3-540-58855-8
[IAgentsConferenceII, 1995]	“Intelligent Agents II: Agent Theories, Architectures and Languages”. Michael Wooldridge, Jorg P. Muller and Milind Tambe (Eds.) IJCAI ‘95 Workshop (ATAL), Montral, Canada, August 1995 Proceedings Springer-Verlag ISBN 3-540-60805-2
[Hall&Shahmehri, 1996]	“An Intelligent Multi-Agent Architecture fpr Support of Process Re-use in a Workflow Management System”, by Thomas Hall and Nahid Shahmehri, in The First internation Conference and Exhibition on The Practical Application of Intelligent Agents and Multi-Agent Technology, London, 22-24 April 1996 Proceedings ISBN 0-9525554-3-3
[IntelligentAgents, 1995]	“Intelligent Agents:Theory and Practice”, by Wooldridge, Michael and Nicholas R. Jennings Submitted to Knowledge Engineering Review, October 1994, revised January 1995.

[Minsky, 1985]	"The Society of the Mind" by Marvin Minsky, New York:Simon & Schuster, 1985
[Paper12-PAAM96]	"A Practical Approach to designing and building intelligent software agents." PAAM 1996, The First International Conference and Exhibition of The Practical Application of Intelligent Agents and Multi-Agent Technology, The Westminster Central Hall, London, UK, 22nd - 24th April 1996
[RightsOfAgents, 1996]	“The Rights of Agents” , by Christen Krogh, in Intellgent Agents II: Agent Theories, Architectures and Languages, eds M. Wooldridge, J.P.Muller and M. Tambe , printed by Springer. ISBN 3-540-60805-2
[Rosenschein and Genesereth, 1985]	“Deals among rational agents” by Rosenschein, J.S. and Genesereth M.R in Proceedings of the Ninth International Joint Conference on Artificial Intelligence (IJCAI-85), p91-99, Los Angeles, CA as referenced in [Intelligent Agents, 1995]