INFORMATION SECURITY IMPLICATIONS OF AUTONOMOUS SYSTEMS

Zia Hayat[1], Jeff Reeve
University of Southampton, UK

and

Chris Boutle, Martin Field
BAE Systems, UK

Abstract

A key challenge within autonomous systems, which consist of autonomous entities (human or machine), is to automatically evaluate the relevance of information thus enabling individual entities to dynamically control access (classify) to any information they possess. This is particularly important when decentralized information sharing is required, ensuring that only authorized entities have access to potentially sensitive information. A high-level model for representing information relevance (and classification) based upon context-aware computing and role based access control concepts is described in this introductory paper, with potential future work detailed.

Key Words: Information Security, Autonomous Systems, Information Classification, Information Sharing & Network Centric Warfare

Introduction

Since the famous early theory on cybernetics in [1] the chimera of developing autonomous systems has been sought. In reality only automation has been achieved with high-levels of autonomy still an esoteric aspiration. In recent years however significant strides have been taken in the development of truly autonomous systems, most of this work has been in the software domain, in the form of autonomous agents. Autonomous agency as described in [2] is a concept intending to shift software development from that of object-orientation and latterly component-ware to richer and more natural techniques.

Much work has been carried out in developing software architectures [3], [4], [5] to model autonomous agents capable of monitoring and controlling complex and distributed processes [6]. In [7] Jennings describes an implementation of a multi-agent system (MAS) in which autonomous software agents collaborate to achieve desired objectives which may be set by any controlling authority (human or machine). However the environments in which autonomous agents have so far been considered have been relatively constrained resulting in little emphasis on security issues, therefore this work aims to build on previous research by taking information security issues into consideration, which is essential for the military application of autonomous systems. It must be noted that communications and storage as well as system state trust are also major electronic security issues; however they are not explicitly considered in this paper.

So far autonomy has been defined and studied using two distinctive approaches in the autonomous systems and agents research communities. Firstly it has been defined as the degree of self-control an entity has over its own decisions, which is assigned by a higher-level authority, e.g. a supervisor (human or machine), we call this *decision autonomy*. Decision autonomy is a satisfactory concept when considering autonomous entities in closed environments. However as the operational environment becomes more uncertain (real-life) then an additional understanding of autonomy with respect to its self-capability is required as described in [8], we call this *self-capability autonomy*. This latter approach is an assessment of an autonomous entities ability, to accomplish its assigned mission objectives with minimal external co-operative intervention.

Evolving Organizational Structure

The deployment of machine autonomous entities in various military scenarios could be to supplement current

[1] Sponsored by BAE Systems Integrated System Technologies and EPSRC, UK.

human activities for specialist or broader task completion. In the near term it is envisaged that machine autonomous entities will operate under the direct control of human commanders, completing specific tasks in a well constrained manner. However as the machine autonomous entity becomes more advanced it may be used at different levels of the command hierarchy giving rise to a military structure which consists of fully interoperable and cooperating human and machine entities capable of fulfilling the required goal(s) as an effective team; this is illustrated in figure 1. Such an architecture will encompass: machine to machine (M2M), human to machine (H2M) and human to human (H2H) interactions, where machines are capable of dynamically adapting according to the specifics of the current and projected environment. This capability is known as a *self-healing* architecture [9], enabling machines to adjust their role(s) and task execution analogous to human operatives, in the event of a breakdown in the current command structure.

Due to the remote-controlled nature of current autonomous entities such as unmanned air vehicles (UAVs) the requirement to exchange information with entities other than the ground control station (centralized) is not prevalent, however as autonomous entities are assigned more decision autonomy then the communications will become less constrained giving rise to complex (decentralized) interactions.

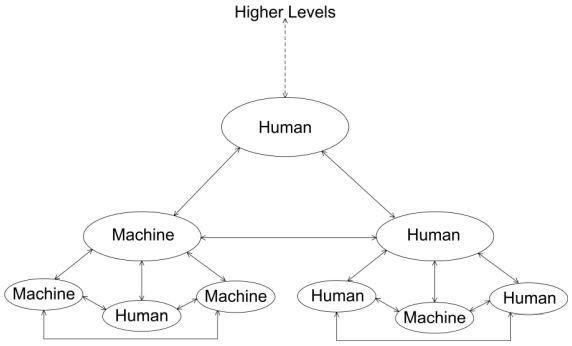

Figure 1. Envisaged human-machine command structure

In current military systems information is protected at a blanket *system-high* security mode, due to the use of a simplistic process for deriving the significance of information. Although it is not currently a major issue in a centralized communications system, decentralized communications will mean that a system-high security mode is likely to introduce inefficiencies in information sharing as many entities will not possess sufficient clearances in order to preserve the security principle of least-privileges (need-to-know). It is this aspect of information security, which is addressed in this paper aiming to reduce limitations in information sharing due to system-high classifications. We investigate the current method of information classification and offer an alternative theory which controls access to information services from a relative (unique to each entity) point-of-view using concepts from context-aware computing and role based access control (RBAC), which can be used to assess the authorization of an entity to access some information service.

Information Security

For the purposes of this paper we assume the use of a standard subject-privilege-object security model, where a subject (autonomous entity fulfilling a role) can access an information service (object) in a constrained manner according to the authorizations (privilege) it possesses. It is believed this will enable machines to be used analogous to humans (i.e. replace and supplement human activities) allowing the necessary interactions (both H2M and M2M) to take place, whilst preserving the security of information, ensuring that only authorized entities operate (read, write etc.) on sensitive information according to its perceived relevance.

It is imperative to identify the relevance of information eliminating any unnecessary overheads introduced by the current system-high security approach. System-high security associates a static and monolithic relevance to all information in-line with the overall mission. However this can result in: Inappropriate information exchanges, where all entities within an operation may exchange either all or no information, violating the security principle of least-privileges or the key requirement for efficient information sharing.

Information Sharing

Currently in the military organization information sharing is achieved through a mixture of human and machine elements. In this arrangement machines are used to store and transmit raw data, whilst authorized human

entities subjectively filter (process) and disseminate such data to other human entities providing situational awareness (SA) based upon their current operational context. Such a human filtering process preserves the information security principle of least-privileges, as well as ensuring that recipient entities are not overloaded with information, which could degrade their performance.

The UK MoD concept of network enabled capability (NEC) which is also endorsed by the wider NATO community as network centric operations[2] (NCO) advocates the use of ubiquitous computing and communications throughout the military organization, this is aimed at improving operational effectiveness through efficient information sharing (horizontally and vertically). In order to maintain efficiency in information sharing one must devise information processing techniques capable of providing the right information to the right entity at the right time, which has also been described by the US DoD [10] as providing only relevant information to entities in the battlespace. Such a requirement is unlikely to be served through the use of manual processes for information processing which would introduce significant bottlenecks (time and accuracy) into the information sharing process ultimately reducing the potential effectiveness of operations. The use of machine autonomous entities at various levels within the command chain will further exacerbate the need for an automated information filtering technique enabling machine-real-time operations.

Any automated filtering technique must be capable of taking into consideration the current operational context of an entity in order to identify the subset of information it requires. This is equivalent to the human information filtering and dissemination process described previously. In order to achieve such intelligent machine processing semantic information is required to allow for a machine to consistently differentiate between information sets according to context as described by Kimber [11]. The concept of context-aware computing (or location-based services) [12], [13], [14] can be seen to tackle similar issues.

The fundamental aim of context-aware computing is to push/pull relevant information services to entities based upon their perceived current context, this aids in lowering overheads in communications, CPU usage, battery power and most importantly information overload. From a security point-of-view one may extend traditional access control mechanisms such as RBAC to constrain the flow of information to individual entities based upon the current context in which a specific entity is operating thus preserving the principle of least-privileges, this was first considered by Zhang and Parashar in [15]. The majority of location-based services use location and time as the two principal attributes to evaluate the relevance of specific services to a given entity. Both of the aforementioned attributes may be used in conjunction with a third dimension of operational risk to identify the authorization a given entity may have in relation to some information service in a given operational threat environment.

Relative Information Classification

The model depicted in figure 2 highlights the principal metrics, which are proposed to distinguish relevant military SA information for a given subject. The *time* axis relates to the period over which a given piece of information is valid – therefore although a particular service may be capable of providing the planned location of logistics drop off points for the next two days, only such information for the next three hours may currently be of relevance; the *space* axis relates to the geographical zone over which a given piece of information is valid – therefore although a particular service may be capable of providing the location of all friendly and foe entities in the complete battlespace, only such information for a $500m^3$ zone may currently be of relevance. It is proposed to use a third attribute of *risk* to determine the perceived threat level associated with executing a given service and ultimately control access to information sharing, where risk is a function of the temporal and spatial validity of information provided by a given service as described previously as well as known friendly and adversary capabilities. This model for preserving the security of

[2] In this context Network Centric Operations is a generic term also covering similar topics like Network Centric Warfare, Network Based Defence, etc.

information services is called 'relative information classification', as information is no longer perceived to have a static and universal relevance for all entities; instead the same subject may have a differing need-to-know (or access rights) over the same piece of information according to its current role and location in temporal and geographical space.

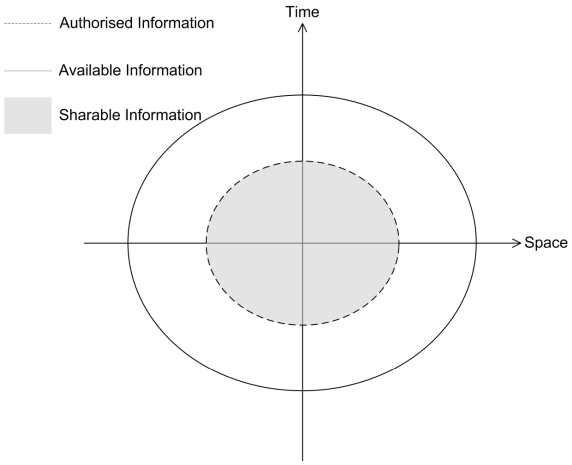

Figure 2. Information relevance metrics

In order to achieve a relative classification system, entities will have temporal and geographical location (origin in figure 2) attributes from which information relevance and therefore access control decisions will be derived. These attributes will be based upon the current time and location of the entity or some other pre-authorized time and location within which a given entity is to carry out a mission. Relevant information will then be identified based upon the role an entity is currently fulfilling, as each role has associated maximum time (past and/or future) and space windows for which a given information service may be accessed. A maximum tolerated residual risk value will be associated with an individual or groups of service, similar to that currently calculated for UK government information technology (IT) systems using HMG infosec standard 1 [16]. The attributes which will be used to evaluate the level of risk associated with a specific information exchange are currently under investigation, they will incorporate the significance of the information exchange to friendly and foe entities, and the ability of foe entities to successfully exploit that information according to the communications environment e.g. high-grade/standard crypto, known adversary passive/active eavesdropping ability.

A privilege derived according to the three metrics of time, space and risk for a particular service is called an *authorization state* this enables the exchange of relevant (context-aware) information. Based upon the movement of a given entity its authorization state(s) may also be automatically updated adding and revoking rights automatically. A trusted mechanism is also required to provide authentic context attributes of time, space and risk. One such mechanism is the toolkit developed by Dey and Abowd [17] which was subsequently used in [15].

A criticism of using an approach such as authorization states may be that missions do not often go to plan, therefore providing a single set of authorization states may result in a lack of relevant information. However a method to overcome this could be to enable fallback (redundant) states which are authorized in the event of a specific occurrence, such as:

- Vehicular problems (e.g. loss of fuel).
- Force transformation (e.g. loss of another friendly entity for which the entity in question is a substitute).
- Operation evolution (e.g. change in enemy tactics requires rapid change in mission and therefore time and location).

A redundant authorization state(s) mechanism may sufficiently overcome the first two of the listed issues. The potential for an autonomous entity to adapt its behaviour and therefore task execution with increasing degrees of flexibility will to some extent be bounded to satisfy safety requirements therefore it is practically achievable to identify a pre-defined set of redundant authorization states.

An example use of authorization states may be where an autonomous air vehicle (A_1) is flying back to base having carried out some sensitive reconnaissance mission, however on its way back A_1 encounters another autonomous air vehicle (A_2), which requests information on the location (service – *foe_loc*) and types (service – *foe_type*) of foe entities encountered by A_1 on the mission, as well as other information A_1 may possess regarding known

future friendly plans (service – *fut_friend_plan*) for a particular location in space. A_1 then requests authentication credentials (e.g. digital certificates signed by a mutually trusted higher-level commander) from A_2 to prove its identity and role, after A_2 has proved possession of the necessary credentials both A_1 and A_2 assess the risk level associated with the current environment for each service, and if the risk level is below the maximum threshold for each service A_1 will provide A_2 with the information it is deemed to be authorized to access. An example overview of how authorization states may be used to restrict the exchange of information to only that which is relevant and authorized is illustrated in figure 2 as the 'sharable information' (available information∩ authorized information) region which is the intersect of the 'available information' and 'authorized information'. It must be noted that if the risk level calculated by the communicating parties (e.g. A_1 and A_2) is greater than the maximum tolerated for the service in question (e.g. *foe_loc*) then execution of that service is prohibited. Figure 3 further highlights the case where an extended situational picture (high-level of sensitivity) is available as part of some service, however an entity may posses an authorization state[3] which allows access to only a subset of this information as indicated by the dotted line.

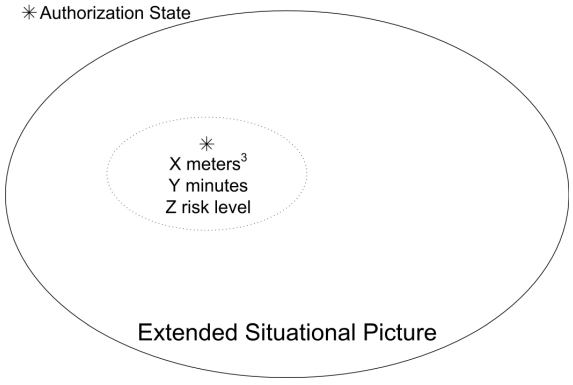

Figure 3. Illustration of an authorization state in the wider SA context

As mentioned earlier it is proposed to restrict the relevant time and space windows for each role/service pair. The tolerated risk level will be unique to each service regardless of role. Therefore entities with roles requiring greater context (i.e. section commander as opposed to infantry soldier) would be authorized to access information with a greater time and space validity giving a broader situational picture. One must therefore associate appropriate values for time and space to role/service pairs as well as maximum tolerated risk levels to each and every service.

It is believed a relative information classification technique would:

- Reduce limitations associated with a system-high security approach.
- Allow fine-grained access control over information through its dynamic classification relative to the recipient entity.
- Enable both security and cooperation through centralized and decentralized information sharing in machine-real-time.

Key Research Issues

The level of decision and self-capability autonomy achieved by entities in a given environment is significantly influenced by the ability of those entities to appropriately classify and therefore share information with other authorized entities. Thus an automated method incorporated into each distributed entity is required for the classification of information; otherwise autonomous entities may share either all (optimistic and high-autonomy) or none (pessimistic and low-autonomy) of their information. This form of binary security is potentially inappropriate, resulting in information security compromises or non-effective operations. It is important to understand that automatic information classification is not a unique issue for autonomous entities. However the need to introduce an automatic technique for information classification becomes pressing when using autonomous entities if the information they gather, possess and share is sensitive and must therefore be secured.

Before sharing information one must identify its relevance as described in the 'Information Security' section previously, however the relative information classification model must be further refined. In particular the types of service available to different roles

[3] For simplicity entity A_2 is shown to possess only one authorization state, however in reality it may have more than one such state.

must be identified for different scenarios (air, land and sea). At a simple level this can be seen as identifying consistent differences between situational pictures at various levels of the command chain from strategic down to tactical. A qualitative assessment of such differences will therefore be undertaken using subject matter experts (SMEs) for air, land and sea domains and used to enable subsequent automatic information filtering.

It is proposed to investigate the effect of restricting the flow of information (level of security applied) between autonomous entities on an autonomous systems' operational capability (quality of service). By increasing and decreasing time and space windows for role/service pairs as well as risk levels for individual services one may identify optimum or pragmatic values for each one in different scenarios. We aim to develop two simple scenarios (land and air) and assess the effect on an autonomous systems performance. We will measure the performance against objectives using a number of parameters, such as:

- Time taken to execute task.
- Percentage of foe's executed.
- Percentage of fratricide cases.
- Number misidentifications (i.e. friend as foe and vice versa).
- Rounds of ammunition utilized.
- Casualties taken.
- Time period friend was identified as a foe (and vice versa) or no identification.

Two methods of testing the aforementioned scenarios are to be investigated in order to identify appropriate levels of information flow constraints for security purposes. Firstly Gardener and Moffat's [18] technique for quantifying the benefits of information sharing may be used by restricting the flow of information in a combat model (e.g. HiLOCA – High level Operations with Command Agents), and analysing the effect of such restrictions on the decision making of autonomous systems through the overall network performance (ONP) concept they introduce. Alternatively off the shelf synthetic environment (SE) technology may be utilized in constructive mode, where all actors are machine autonomous entities.

Rules must be defined to specify when a given entity may utilize a fallback authorization state. Model checking techniques such as [19] may then be used to evaluate the integrity of such a model, where known and unknown sequences of events may be tested to check if fallback authorization states can be activated in an unauthorized manner. A mechanism must also be derived to overcome potential differences in risk levels calculated by communicating parties.

Conclusions

We have identified and described the information security implications of autonomous systems, which it is need to be addressed. A clear distinction has been made between two forms of autonomy: decision and self-capability, which have previously been used interchangeably throughout the literature. Information security issues surrounding the need for automated and dynamic classification of information hosted by machine autonomous entities have also been identified as a key requirement. It is believed such a capability would enable efficient information sharing whilst preserving the security principle of least-privileges (or need-to-know) in a decentralized communications architecture.

References

[1] N. Wiener, "Cybernetics or Control and Communication in the Animal and the Machine", MIT Press, Cambridge, Massachusetts, USA, 1948.

[2] N.R. Jennings, "An agent-based approach for building complex software systems", *Communications of the ACM* 44(4):pp. 35-41, 2001.

[3] JACK User manuals, Available Online: http://www.agent-software.com, last accessed on 28th February 2006.

[4] SOAR User Manuals, Available Online: http://sitemaker.umich.edu/soar, last accessed on 28th February 2006.

[5] AgentBuilder User Manuals, Available Online: http://www.agentbuilder.com/AgentTools/index.html, last accessed on 28th February 2006.

[6] N.R. Jennings and S. Bussmann, "Agent-based control systems", *IEEE Control Systems Magazine*, pp. 61-74, 2003.

[7] N.R. Jennings, "Decentralised co-ordination of agents to achieve operational goals", *In BAE Systems Platforms & Systems Conference Proceedings* CD, Data and Information Systems/Nick Jennings, Loughborough, UK, 2005.

[8] A. Yavnai, "An Information-Based Approach for System Autonomy Metrics Part I: Metrics Definition", *In Proceedings of Performance Metrics for Intelligent Systems Workshop*, Maryland, USA, 2003.

[9] I. Georgiadis, J. Magee and J. Kramer, "Self-Organising Software Architectures for Distributed Systems", *In Proceedings of 1st Workshop on Self-Healing Systems*, ACM Press, pp. 33-38, South Carolina, USA, 2002.

[10] D. S. Alberts, J. J. Garstka and F. P. Stein, "Network Centric Warfare: Developing and Leveraging Information Superiority", C4ISR Cooperative Research Program, US DOD, 1999.

[11] P. Kimber, "Information Exchange for Autonomous Systems: Addressing the Challenges of Machine Understandable Representation", *In Proceedings of the IEE Conference on Autonomous Systems*, London, UK, 2005.

[12] G. Chen and D. Kotz, "A Survey of Context-Aware Mobile Computing Research", Technical Report TR2000-381, Computer Science Department, Dartmouth College, New Hampshire, USA, 2000.

[13] P. Coppola, V. Della Mea, L. Di Gaspero, S. Mizzaro, I. Scagnetto, A. Selva, L. Vassena and P. Z. Riziò, "Information Filtering and Retrieving of Context-Aware Applications Within the MoBe Framework", *In proceedings of the Workshop on Context-Based Information Retrieval*, Paris, France, 2005.

[14] A. Smailagic, D. P. Siewiorek, J. Anhalt, F. Gemperle, D. Salber and S. Weber, "Towards Context Aware Computing: Experiences and Lessons", *IEEE Journal on Intelligent Systems*, Vol. 16, No. 3, pp 38-46, 2001.

[15] G. Zhang and M. Parashar, "Dynamic Context-aware Access Control for Grid Applications", In IEEE Computer Society Press, *4th International Workshop on Grid Computing*, pp. 101-108, Phoenix, Arizona, USA.

[16] Ministry of Defence, Her Majesty's Government Infosec Standard No. 1, "*Assurance Requirements for IT Systems*", London, UK, 2001.

[17] A. K. Dey and G. D. Abowd, "The Context Toolkit: Aiding in the Development of Context-Aware Applications', *In Proceedings of Human Factors in Computing Systems*, Pittsburgh, Pennsylvania, USA, 1999.

[18] T. Gardener and J. Moffat , "Quantifying the Benefit of Shared Information", *In Proceedings of Institute of Mathematics and its Applications Conference on Analysing Conflict and Its Resolution*, Oxford, UK, 2004.

[19] LTSA, Available Online: http://www.doc.ic.ac.uk/~jnm/book/ltsa/LTSA.html, last accessed on 28th February 2006.

Acknowledgements

The work reported in this paper was funded by the Systems Engineering for Autonomous Systems (SEAS) Defence Technology Centre established by the UK Ministry of Defence, under contract number DTC/RAO/WPE/N03751/SEAS/IF006. The authors would also like to thank David Charles, Chris Coles, Richard Elder, Robert Johnston, Paul Kimber, Chris Wood and Andrew Wright for their invaluable contribution to this work.