Securing Autonomous Systems

Zia Hayat[1], Jeff Reeve
University of Southampton
Communications Laboratory
School of Electronics & Computer Science
Faculty of Engineering
Hampshire S017 1BJ
UK
zia.hayat@baesystems.com
jsr@ecs.soton.ac.uk

Chris Boutle, Martin Field
BAE Systems
Integrated System Technologies
Frimley
Camberley
Surrey GU16 7EX
UK
chris.boutle@baesystems.com
martin.field@baesystems.com

Abstract

The security of an autonomous system, which consists of autonomous entities,[2] *could be compromised as a result of a physical or electronic attack. The compromise of information assets such as a system's processing base or its communications links could potentially place an enemy ahead in terms of the OODA (Observe, Orient, Decide & Act) loop; it is this aspect of electronic warfare, which is considered in this introductory paper, ensuring that the correct information is provided to the correct entities at the correct time.*

In this paper particular emphasis is placed on information security and system state trust as much work has already been done to preserve the confidentiality, integrity and availability of data whilst in storage and transit. A key challenge within autonomous systems is to achieve automatic information valuation enabling individual systems to dynamically derive the classification of any information they posses. This is particularly important when decentralised information sharing is required, ensuring that only authorised entities have access to potentially sensitive information. A high-level model for representing information value (and classification) based upon context-aware computing concepts is described. If decision autonomy can be described as delegation of authority then rationally one would only prescribe higher levels of decision autonomy to a given entity if sufficient trustworthiness can be associated with it. However how is trustworthiness ascertained? A number of techniques for achieving this are described and surveyed.

Keywords : Autonomous Systems, Information Security, Communications Security, System Trust, Information Sharing and Network Enabled Capability.

1. Introduction

The research described in this paper is work in progress (IF006) being carried out as part of the Systems Engineering for Autonomous Systems (SEAS) Defence Technology Centre (DTC), which is a UK MoD initiative looking into the requirements of autonomous systems capable of operating in various military scenarios. The aim of this study is to identify the key electronic security issues in autonomous systems, with the potential for more detailed downstream work on tackling specific issues.

Since the famous early theory on robotics in [1] the chimera of developing autonomous systems has been sought. In reality only automation has been achieved with 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. With the increasing requirement to engineer more complex software systems, the autonomous agent is seen as an ideal self-contained (abstraction) tool enabling the engineer to focus on their functional requirements (problem domain) and not underlying semantics of software programming.

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] the author 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 a key challenge remains as to how one would effectively integrate humans into such architectures enabling human and machine entities to cooperate and collaborate in an efficient manner achieving common goals, this issue is being looked into in [8]. 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 security issues into consideration.

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 [9], 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.

The deployment of machine autonomous entities in various military scenarios could be to supplement current 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 bounded (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 commonly known as a *self-healing* architecture [10], 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 (centralised) is not apparent, however as autonomous entities are assigned more decision autonomy then the communications will become less constrained giving rise to complex (decentralised) interactions.

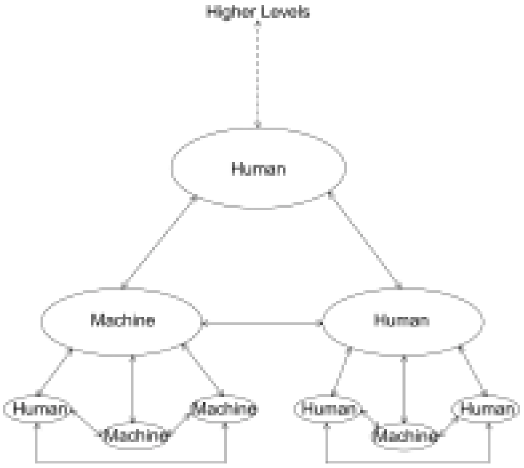

Figure 1: Envisaged human-machine command structure

Three broad areas of electronic security have been identified in which key challenges must be tackled to successfully accredit and therefore deploy autonomous systems for military purposes. These are:

- Communications & storage security
- Information Security
- System state trust

Communications & storage security focuses on specific types of service or technique, which can be used to protect ones information assets such as encryption, authentication and tamper resistant hardware. In current military systems information in transit or at rest is protected to a blanket *system-high* mode, due to the use of an overly simplistic process for deriving the value of information. Although it is not currently a major issue in a centralised communications system, decentralised communications will mean that a system-high security mode is likely to introduce inefficiencies in information sharing as many entities will not posses 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 classifies information from a relative (unique to each subject) point-of-view using concepts from context-aware computing, the result is a triangular surface model, which may be classified into one of a finite set of pre-defined surfaces. The third and final aspect, which is considered in this paper, is system state trust, which relates to the trust one autonomous entity places in another when using its service(s) to perform a task. Much work in this area has already been carried out for MAS [11], [12] based environments (machine only) as well as for HMI (Human Machine Interaction) environments [13], [14], [15]. We analyse these techniques and identify possible synergies between the different approaches as well as potential areas for future work.

This paper is organised as follows, section 2 describes the potential impact on communications & storage security techniques of deploying autonomous systems, section 3 identifies deficiencies in current information security classification techniques and how these may be improved to successfully incorporate fully data-enabled autonomous entities improving the information exchange between such entities at all levels of the military command hierarchy. In section 4 we introduce the problem of system state trust due to autonomous entities and how this may be tackled using currently available techniques. The key research issues to be immediately tackled are then detailed in section 5, with a brief summary given in section 6.

2. Communications & Storage Security

Communications and storage security focuses on maintaining the confidentiality, integrity and availability of data. This is to ensure that only authorised entities can intercept, interpret and modify data whether at rest or in transit. Such data may be for Command & Control (C2), reconnaissance or health status purposes, all of which can be critical to the correct operation of any autonomous entity.

As autonomous entities gain higher levels of self-capability autonomy their reliance upon information from and communications (particularly for C2) with external sources should intrinsically decrease according to [9] and [16]. Thus increasing self- capability autonomy can be seen to improve the stealth of a system by reducing communications and therefore the likelihood of traffic intercept and analysis by an adversary. The ability to continue operating in the presence of reduced communications with external entities also enhances its resilience. Although the volume of communications may decrease with increasing self-capability autonomy, it is suggested that any remaining communications could be highly critical providing invaluable intelligence, which may be for Battle Damage Assessment (BDA) or other potentially emergency reasons, such as warning friendly entities of previously unknown but crucial information such as enemy air defence deployments.

If vast numbers of autonomous entities are to be deployed in swarm like architectures analogous to humans then the mechanisms used to provide security services must be efficient and effective. Therefore solutions need to be engineered which will enable information provenance, authentication, authorisation, access control and revocation of rights in a vast and decentralised system. For all of the aforementioned requirements the suitability of third party key management techniques such as a Certification Authority (CA) as well as asymmetric cryptography need to evaluated. Issues such as:

- What is the availability of CAs likely to be for autonomous entities?
- At what rate are CAs likely to be updated to take into account updated rights[3] information?

In a scenario whereby autonomous entities are capable of adapting roles in a dynamic fashion it is likely that rights (privilege) information associated with an entities role will have a high turnover, this will result in the need for high levels of connectivity between autonomous entities and servers managing such information. However, as the level of self-capability autonomy associated with a given entity increases then the viability of utilising security solutions which require high levels of connectivity with external sources must be questioned, due to the requirement for entities with increasing self-capability autonomy to have decreasing reliance on communications with external sources as described in [9] and [16].

The drive towards designing security solutions which are less computationally intensive in fields such as sensor networks has seen the advent of techniques such as μTESLA [17], Multi-Level μTESLA [18] and pragmatic trust establishment [19] enabling efficient broadcast authentication and key management. Although this work aims primarily to reduce overheads due to battery and processor power it may be further extended to explicitly consider the requirement for improving efficiency from a communications overhead point-of-view. For example a machine may be loaded with information encrypted with differing keys. Therefore during the course of time it may then access information if it has a need-to-know with only the corresponding encryption key communicated by external entities. A similar method is employed for broadcast communications in the μTESLA strategy. Other work on reducing communications overheads should concentrate on identifying efficient architectures for distributing security credential information and optimum levels for refreshing security credential information for individual entities and the servers they rely upon.

When storing sensitive data appropriate strategies must be identified to protect it from unauthorised access. If a machine falls into enemy hands then appropriate measures need to be taken to ensure that the confidentiality of any information stored in memory is not compromised. This may be achieved by purging the contents of memory in specific circumstances such as unauthorised tampering. Purging the contents of memory to military standards can take long periods of time and consume excessive power, either of which may not be sufficiently available in certain emergency scenarios. Therefore the alternative of always encrypting resident memory contents would overcome such issues, where the cryptographic keys (likely to be much smaller) may be purged instead. However one must assess the risk of being unable to perform such an action in the event of an emergency.

3. 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) can access information (object) in a constrained manner according to the authorisations (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 authorised entities operate (read, write etc.) on sensitive information according to its perceived value.

It is imperative to identify the value of information eliminating any unnecessary overheads introduced by the current system-high approach. System-high security assumes a static and monolithic valuation of all data in-line with the overall mission. However this can result in:

- Over and even under provisioning of security services when considering communications

& storage aspects as described in section 2

- 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

One must therefore identify the overheads associated with applying differing levels of security for storage and communications security purposes. It is assumed that for the majority of unmanned vehicular systems all data in memory may be protected to the same level, without compromising its security and incurring unnecessary overheads. A similar technique may be used when communicating information to external entities, however the overheads associated with utilising different levels of security (e.g. cryptographic key sizes) may be significantly low during the course of mutual authentication and other initial handshaking protocols. This would enable distinct techniques to be employed based upon the perceived value of the information as well as the techniques available to the communicating end points.

Currently in the military organisation 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 authorised 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[4] (NCO) advocates the use of ubiquitous computing and communications throughout the military organisation, 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. 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. This is highlighted in currently proposed UK army systems such as FRES and FIST, where it is envisaged that data as well as voice information will be relayed to the individual soldier. The use of machine autonomous entities at various levels within the command chain will further exacerbate the need for an automated 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. The concept of context-aware computing (or location-based services) [20], [21], [22] can be seen to tackle similar issues. The fundamental aim of context-aware computing is to push/pull relevant information and 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 utilise a similar concept 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. The majority of location-based services use location and time as the two principal attributes to evaluate the relevance (value) of specific services to a given entity. Both of the aforementioned attributes may be used in conjunction with a third dimension of information context to identify the authorisation a given entity may have in relation to some piece of information.

The 3D model depicted in figure 2 highlights the principal metrics, which are proposed to identify the value of military SA information. The *time* axis relates to the period over which a given piece of information is valid, therefore as the valid time period increases so does the perceived value of the information; the *space* axis relates to the physical area over which a given piece of information is valid, therefore as the area increases so does the perceived value of the information; and the *contextual richness* axis is a proxy relating directly to the role of the recipient of a given piece of information, hence as the recipients role increases in importance so does the perceived contextual richness and hence value of the information. We call this *relative information classification* as information is no longer perceived to have a static and universal value for all entities, instead the same piece of information may have differing values based upon the location (in time and space) and role of the subject (entity) in question.

From discussions with key military information system security researchers in the UK it has been found that concepts similar to the relative information classification theory are being worked upon in a number of military projects in the USA. In particular information classification for the Global Information Grid (GIG) architecture, which is the future unified information storage and transportation system for the US military is believed to be propose a similar technique.

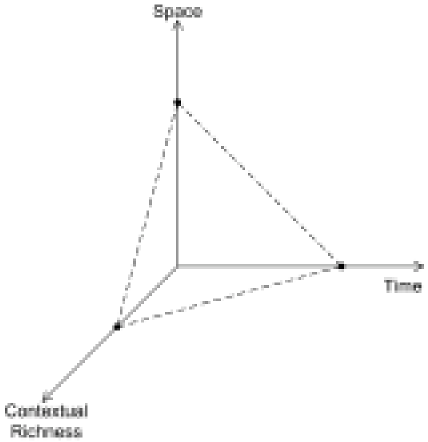

Figure 2: Information value metrics

Based upon an entities current role it will have an operational need-to-know for specific information elements. Such requirements may be identified by examining information exchange in real-life scenarios or synthetic environment simulations.

In order to achieve a relative classification system as discussed previously entities will require in real terms, time and space attributes from which security classifications may be derived. Therefore the time, space and contextual richness attributes for which a given entity is authorised must be identified we call such values *authorisation state(s),* this may be based upon the current time, location and role of an entity or it may be some other pre-authorised time and location within which a given entity is to carry out a mission. The value[5] of some piece of information can therefore be derived by assessing its time, space and contextual richness validity with reference to an authorisation state (origin). Some form of proof of authentication (e.g. digital certificates signed by a higher-level commander) will be required in order to enable such a technique. It must be noted that a single entity may posses numerous such authorisation states at any one point in time. Such a model would enable secure and efficient sharing of information, which is required in real-time for both centralised and de-centralised exchanges.

A criticism of using an approach such as authorisation states may be that missions often do not go to plan, therefore providing a single set of authorisation states in time, space and role may result in a lack of relevant information. However a method to overcome this could be to enable fallback (redundant) states which are authorised 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 authorisation state(s) mechanism may sufficiently overcome the first two of the listed issues, however the potential for an autonomous entity to adapt its behaviour and therefore task execution with increasing degrees of flexibility would require an undefined set of redundant authorisation states[6], which is unlikely to be practically achievable.

An example use of authorisation 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 from A_1 regarding the information it has just gathered, as well as other information A_1 may possess regarding known future plans for a particular location in space. A_1 then requests authentication credentials from A_2 to prove its identity and its authorisation state(s), subsequently A_1 will provide A_2 with the information it is deemed to be authorised to access based upon time, space and contextual richness parameters as discussed previously. This is illustrated in figure 3, where entity A_1 possesses a significant (wider) situational picture (classified SECRET) and entity A_2 has an authorisation state[7] for a particular location in space and time for which it may access information. Based upon the proof-of-possession of an authorisation state entity A_1 may then share information with entity A_2 for the authorised state zone (in both time and space) indicated by a dotted line in the diagram. This information may also need to be sanitised to ensure that only information required for the role (context richness) of entity A_2 is shared, such as location and or type of friendly and or foe entities etc. Based upon the velocity of a given entity its authorisation state(s) may also be automatically updated adding and revoking rights automatically.

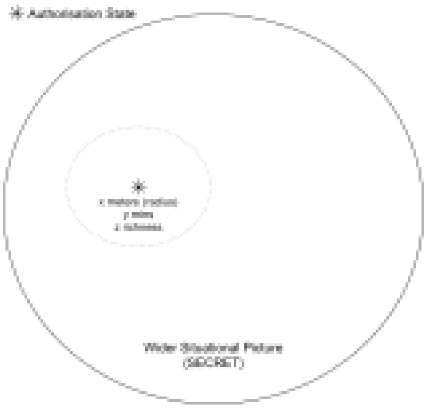

Figure 3: Illustration of an authorisation state in the wider SA context

It is proposed to restrict the time and space size windows as well as contextual richness of information for an entity based upon its role. Therefore entities with roles requiring greater context (i.e. section commander as opposed to infantry soldier) would be authorised to access information with a greater time and space validity as well as contextual richness giving a broader situational picture. One must therefore associate appropriate values for time, space and informational parameters for each and every role.

A key question is: when to evaluate information for its perceived value? A couple of strategies exist for this, firstly all information hosted by a machine may be classified as it is downloaded to the entity, this would require complex processing and each file would have to be classified based upon the information hosted. This may be achieved by applying a classification technique (rule or non-rule based) based upon the files meta-data structure as described by Kimber [23], where the meta-data is used as a contextual provider. As well as complexity another criticism of using such a technique is that the classification of information would remain static, which in a dynamic operational scenario may limit the information sharing between entities ultimately prohibiting operational effectiveness. Hence it is proposed to classify information once it is to be released by one entity to another, with all data stored to the same security level whilst at rest on the machines memory. This would:

- Reduce limitations associated with a system-high approach or overheads due to continuous classification

- Allow the dynamic classification of information relative to the recipient entity

- Enable both security and cooperation through centralised and de-centralised information sharing in machine-real-time

4. Trust in System State

Traditionally computer security has focused on communications security, ensuring that what is sent by one entity is received only by authorised entities in an integral state. However this assumes that all information sent by the transmitting party is faithful and accurate. If autonomous systems are to be developed using Commercial off The Shelf (COTS) components from developers of an unknown pedigree; consist of entities from coalition partner(s) in whom one may have varying degrees of trust as well as exhibit emergent behaviour due to learning abilities, then the aforementioned assumption must be re-evaluated, as the transmitting system (information source) may itself be compromised.

The trust in autonomous systems must be assessed as well as the trust in the communications systems they utilise. If compromised either one could lead to misinformed entities. However the compromise of an entities state is potentially more damaging than any compromise of its communication link(s). This is due to the fact if an entity is itself compromised then all of its communications are also deemed to be compromised as it may intentionally provide false information as well as leak information to unauthorised entities (enemy). However if a communications link is compromised then the information delivered over any such link may be subject to denial of service, eavesdropping and even spoofing. Depending on the remote control capability of an entity, it may be compromised to differing degrees through a compromised communications link, which enables telecommand.

System state trust can be seen as a relatively new dimension when considering electronic systems security aspects. It relates to the trustworthiness one associates with an entity and therefore the information it provides and actions it executes. From an abstract point-of-view, ones trust in any such system can be summarised as the perceived: competence, benevolence and compliance this was originally described by Bradshaw et al. in [12], where:

- Competence is the ability of a given entity to reliably perform a task in a manner, which is good for achieving the overall goal

- Benevolence is the confidence of protection from malicious intent (compromise)

- Compliance is a unifying factor, which relies on an entities conformance with supervisory controls, aiming to make up for gaps in competence and limit damage from malicious intent

Together benevolence and competence can be used to derive the trust one entity associates with another, where benevolence is as a result of confidence gained over time for a given entity. Neither benevolence nor competence alone can be used to derive trust, for example if you have faith in ones character but not competence you would still not trust them in certain scenarios. Therefore both benevolence and competence are required in tandem in order to trust an entity and therefore ultimately delegate authority.

Due to the relative difficulty in ascertaining the benevolence or motives of any given entity at any point in time it is recommended that a strategy of *train and test* be applied relating purely to the competence one associates with a given entity. This in effect accepts that an entity may behave maliciously as no clear scientific or procedural mechanisms exist to prevent such an issue. Therefore the provenance of information is also very important enabling one to trace the origin of information, which may have emanated from a known un-trusted source. Information provenance is a complete research area in its own right and is being tackled in the EU Provenance project [24], which aims to build trust and validation into distributed computer networks. Although benevolence is an inherently difficult aspect to measure and control, it is argued that if one strictly obeys the computer security principle of least-privileges, then the level of damage any given entity can cause in terms of providing false information and executing prohibited actions is limited.

Johnson [6] describes in some detail his understanding and definitions of trust in different circumstances. It is shown how trust can be correctly placed; misplaced and even displaced, however the relationship between trust and autonomy (decision or self-capability) is not explicitly identified. Ming et al.'s ALFUS[8] (Autonomy Levels For Unmanned Systems) [25] model describes the level of self-capability autonomy associated with a given entity, however in order to analyse the competence of any given entity at any point in time relevant attributes must be identified and evaluated. Therefore the identification of generically relevant attributes is a potential area of future work.

The aim of deriving trust from a security point-of-view is to ultimately evaluate the level of decision autonomy one entity delegates to another. Based upon Bradshaw et al.'s [13] notion of achieving trustworthiness by associating competence and benevolence with an entity, one may derive a unified model for delegating decision autonomy. Therefore the concepts of trust, mistrust and distrust as described by Johnson may be used to identify the benevolence of a given entity based upon past experiences. Similarly the ALFUS model described by Ming et al. may be used to derive the competence one entity (e.g. supervisor) associates with another (e.g. assistant), with an underlying security policy (static or dynamic) providing a bounding model.

It is important to note that given precisely the same levels of competence and benevolence as well as compliance model one may decide to trust an entity at one point in time and not another due to other contextual reasoning. Therefore a fuzzy logic based approach may be adopted to evaluate the level of trust and therefore decision autonomy one entity would place in another. This is similar to the use of a confidence model based upon fuzzy sets by Ramchurn et al. [26] to derive the trust one entity places in another (software agent), based upon assertions made by other members within a MAS society, who may have interacted (directly or indirectly) with the entity in question.

5. Key Research Issues

It is argued that 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 authorised entities. Therefore 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 inappropriate, resulting in information security compromises or non-effective business 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 classifying information one must identify its value as described in section 3, the information valuation system based upon time space and contextual sensitivity must therefore be further refined. In particular metrics capable of distinguishing between various levels of 'contextual richness' 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. It is hoped sufficiently consistent differences can be identified to then quantify and subsequently automatically filter information.

It is proposed to investigate the effect of: restricting information flow (level of security applied) on the quality of information sharing between autonomous entities and subsequently the effect on an autonomous systems operational capability (quality of service). By expanding and contracting time and space validity as well as contextual richness of information parameters one may identify optimum or pragmatic values for each one in different roles and scenarios. We aim to develop two simple scenarios (land and air) and assess the affect of an autonomous systems performance. Firstly a land mine sweeping objective (from point A to point B) given to a group of autonomous Unmanned Ground Vehicles (UMVs) operating under a single commander in a hostile environment. The second scenario is for air surveillance and shoot to kill (over a pre-defined space) where a task given to a group of autonomous unmanned air vehicle's operating under a single commander in a hostile environment, where the enemy has deployed air defence systems. We will measure the performance of completing these objectives using the following parameters:

- 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 utilised
- 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 [27] technique for quantifying the benefits of information sharing may be used by restricting the flow of information in their 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 utilised in constructive mode, where all actors are machine autonomous entities. In [28] the author describes a knowledge based decision support system for autonomous systems, it is believed that an information flow restriction technique for security purposes may also be incorporated into such a model to assess the effects on shared situational awareness, decision making and ultimately operational effectiveness.

Rules must be defined to specify when a given entity may utilise a fallback authorisation state. Model checking techniques such as [29] 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 authorisation states can be activated in an unauthorised manner.

6. Conclusions

We have identified and described the security implications of autonomous systems, which it is believed must be addressed before any such system may be accredited and procured for military purposes. A clear distinction has been made between two forms of autonomy decision and self-capability, which have previously been used interchangeably throughout the literature. In order to achieve increasing levels of self-capability autonomy requirements for reduced communications overheads have been described, related work in sensor networks may be leveraged to realise these requirements. Information security issues surrounding the need for automated and dynamic valuation of information hosted by autonomous machines have also been identified as a key requirement. It is believed such a capability would enable efficient information sharing as well as preserving the security principle of least-privileges (or need-to-know). Synergies in relevant work on trust issues in autonomous systems have been described, the relationship between trust and decision autonomy has also been explicitly defined. This research project now aims to concentrate on tackling issues principally in the information security domain as described in section 5.

7. References

[1] I. Asimov, *"I Robot"*, 1950.

[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 Platforms & Systems Conference Proceedings CD, Data and Information Systems/Nick Jennings, Loughborough, UK, 2005.

[8] J. Thoms, *"Autonomy: Beyond OODA"*, In Proceedings of the IEE Conference on Autonomous Systems, London, UK, 2005.

[9] 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.

[10] I. Georgiadis, J. Magee and J. Kramer, *"Self-Organising Software Architectures for Distributed Systems"*, In Proceedings of 1st Workshop on Self-Healing Systems, pp. 33-38,

ACM Press, South Carolina, USA, 2002.

[11] K.P. Sycara, *"Multiagent Systems"*, In AI magazine Volume 19, No.2 Intelligent Agents, 1998.

[12] S. Bussmann, N. R. Jennings and M. Wooldridge, *"Multiagent systems for manufacturing control: A design methodology"*, Series on Agent Technology, Springer Verlag, 2004.

[13] J. Bradshaw, H. Jung, S. Kulkarni, J. Allen, L. Bunch, N. Chambers, P. Feltovich, L. Galescu, R. Jeffers, M. Johnson, W. Taysom and A. Uszok, *"Toward Trustworthy Adjustable Autonomy and Mixed-Initiative Interaction in KAoS"*, In Proceedings of 7th International Workshop on Trust in Agent Societies, New York, USA, 2004.

[14] R. Falcone and C. Castelfranchi, *"The Human in the Loop of a Delegated Agent: The Theory of Adjustable Social Autonomy"*, IEEE Transaction on Systems, Man and Cybernetics—Part A: Systsems And Humans, Vol. 31, No. 5, 2001.

[15] P. Johnson, *"Interactive to Autonomous Systems"*, Proceedings of the IEE Conference on Autonomous Systems, London, UK, November 2005.

[16] F.M. Watkins, *"UCAV Autonomy"*, MSc Thesis in Systems Engineering, Royal Military College of Science Engineering Systems Department, Cranfield University, UK, 2004.

[17] A. Perrig, R. Szewczyk, J.D. Tygar, V. Wen and D.E. Culler, *"SPINS: Security Protocols for Sensor Networks"*, In ACM Journal of Wireless Networks, pp. 521-534, Netherlands, 2002.

[18] D. Liu and P. Ning, *"Multi-Level µTESLA: Broadcast Authentication for Distributed Sensor Networks"*, ACM Transactions in Embedded Computing Systems, Vol. 3, No. 4, pp. 800-836, 2004.

[19] R.J. Anderson, H. Chan, and A. Perrig, *"Key infection: Smart trust for smart dust"*, In Proceedings of 12th IEEE International Conference on Network Protocols, pp. 206–215, Berlin, Germany, 2004.

[20] 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.

[21] 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 Proceedings of the Workshop on Context-Based Information Retrieval, Paris, France, 2005.

[22] 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.

[23] 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.

[24] EU Provenance project, Available Online: http://twiki.gridprovenance.org/bin/view/Provenance/, last accessed on 28th February 2006.

[25] H. Huang, E. Messina, and J. Albus, "*Toward a Generic Model for Autonomy Levels for Unmanned Systems (ALFUS)*", In Proceedings of Workshop on Performance Metrics for Intelligent Systems, Maryland, USA, 2003.

[26] S.D. Ramchurn, N.R. Jennings, C. Sierra and L. Godo, “*Devising A Trust Model forMulti-Agent Interactions using Confidence and Reputation*”, Applied Artificial Intelligence,18:833-852, 2004.

[27] 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.

[28] J. Harding, “*Knowledge Based Strategies*”, Workshop on Systems Engineering, SEAS DTC, Loughborough, UK, 2005.

[29] 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. The authors would also like to thank Robert Johnston, Paul Kimber and Andrew Wright for their invaluable contribution to this work.

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

[2] An autonomous entity as referred to throughout this paper may represent either a human or non-human entity such as an unmanned vehicle or autonomous software agent.

[3] A right may be added or revoked with reference to any subject-object combination.

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

[5] To asses the 'value' of a file meta-data elements will be required to describe time, space and contextual richness attributes.

[6] Introducing overheads due to proof of possession etc.

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

[8] Although the ALFUS model of autonomy has been used here, this research is not constrained to the use of this or any other specific model.