2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Command and Control of Cyber Weapons

Enn Tyugu
Institute of Cybernetics
Tallinn University of Technology
Tallinn, Estonia
tyugu@ieee.org

Abstract: With the development of autonomous malware and autonomous anti-malware, command and control of cyber weapons is becoming an important part of cyber defence. In the present paper we discuss the dangers of deploying and controlling intelligent cyber weapons in a unified setting, considering these weapons as intelligent agents. Command and control of intelligent agents causes new threats that are difficult to avoid due to the complexity of behaviour of agents. Situation awareness of agents must be improved and verified, or at least carefully tested with respect to safety of their behaviour. Several possible dangerous behaviours of cyber weapons are discussed in the talk: misunderstanding of a situation, misinterpretation of commands, and loss of contact and formation of unwanted coalitions. A specific threat is the formation of unwanted coalitions by proactive weapons. This can happen if they get too much autonomy in decision making. A scenario of insubordination of agents is presented, considering a longer time perspective. General conclusions are the following: the more intelligent software becomes the more difficult it will be to control it; when designing and developing new cyber weapons, one has to guarantee the appropriate control over these weapons under any circumstances. It is practically impossible to use formal methods for verifying the safety of intelligent cyber weapons for their users. Setting strict constraints on the behaviour of cyber weapons and their careful testing are necessary.

Keywords: *command and control, intelligent cyber weapons, situation awareness, autonomous agents, proactiveness and adaptability in cyber defence*

1. INTRODUCTION

Command and control (C2) is a key aspect of any military activity, and according to a common understanding it concerns only human actors. With the development of autonomous malware and autonomous anti-malware, command and control of cyber weapons is becoming an important part of cyber defence. This is especially true for intelligent cyber weapons that can make decisions and autonomously plan actions. Hence, command and control must be extended to autonomous cyber weapons. An existing command and control application of this kind is known for botnets. However, it is still a simple case, because the botnets of today still have a

rather straightforward and simple way of operation. However, the situation changes when bots become more intelligent and get more freedom of action. Already in the foreseeable future we can expect much more proactive and intelligent cyber weapons both for offence and defence. One can classify them as intelligent agents and apply respective command and control. Special attention has to be paid to the cooperative behaviour of agents. In the long run, there exists the danger that intelligent agents may become too independent and they will perform unexpected and unwanted (harmful) actions. Avoiding this requires at least thorough verification of the possible behaviours of intelligent cyber weapons, and this is not a trivial task. For instance, on the phenomenological level one can easily postulate Asimov's laws of robotics, but to implement these laws requires more effort than one may expect.

A report from research firm Visiongain predicts that by the end of this year the cyber warfare market will be worth about sixteen billion dollars, as governments around the world invest further resources, creating new systems and protective measures to combat cyber criminals and hostile state hackers [1]. The Japanese newspaper Yomiuri Shimbun reported that the Defence Ministry's Technical Research and Development Institute began developing the anti-viral virus in 2008. Japan has reportedly requested for $2.3 million from Fujitsu to build a self-replicating assassin squad – a computer virus it can set loose in the network to track down and eliminate other viruses [2].

These are just examples, demonstrating that malicious software and cyber weapons are not only spreading, they are also becoming more sophisticated, independent and intelligent. In the present paper, we are analysing the possible consequences of deployment of powerful cyber weapons, in particular, the possibility of preserving control over these weapons. With the development of autonomous malware and autonomous anti-malware, command and control of cyber weapons is becoming an important part of cyber defence.

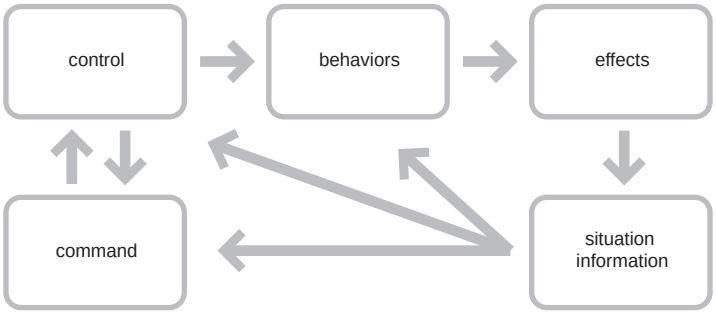
We compare, first, command and control as it has been understood in the context of military operations from one side, and in automatic cyber operations from the other side (Section 2). Then we introduce a generic concept of agent as an intelligent software component with proactivity (Section 3). We make some predictions about the further development of agents in cyber space, and describe their advanced features: beliefs-desires-intentions and reflection (Sections 4). We discuss threats of C2 of agents in Sections 5 and 6. Finally, we present a rather extreme scenario that may follow from the development of intelligence in agents. These scenarios may not become true, but are still possible in principle (Sections 7).

2. TWO FACES OF COMMAND AND COTROL

The magic words 'Network Centric Operations and Network Centric Warfare' point to the changed role of command and control in military operations. The changes concern, first of all, the speed of decision making and communication, but also the increased amount of information available for C2. A conceptual model of C2 is shown in Figure 1. The main control loop is as in control theory: control → behaviours → effects → situation information and back to control. The situation information is collected in classical control by sensors and is usually just a set of

values of variables. One cannot expect that this is the same in the present model. Understanding a situation in the context of military actions may be a very complex intellectual problem. The command and control actions are tightly bound by two-way communication links in this model.

FIGURE 1. CONCEPTUAL MODEL OF COMMAND AND CONTROL



An essential role in this model belongs to human factors as the list of C2 activities given in [3] shows:

- Establishing intent
- Determining roles, responsibilities, and relationships
- Establishing rules and constraints
- Monitoring and assessing the situation and progress
- Inspiring, motivating, and engendering trust
- Training and education
- Provisioning.

In cyber warfare, some command and control has been passed over to automatically operating entities – agents, command and control servers of botnets etc. – and this tendency is increasing. A good analogy can be found in air combat, where most actions are already performed completely automatically, and predictions for the next decade promise wide usage of artificial intelligence in the command and control loop. This increases the role of cognitive methods in situation awareness and situation management [4]. Changes are well visible in cooperative situation awareness of agents and humans. A crucial property is the speed of automatic C2 decision making. We will discuss these aspects in Section 5, dedicated to the threats of command and control of agents.

Let us look at the command and control in a case report of the Golden Cash botnet developed in 2008 and uncovered in 2009 [5]: "A user visits a legitimate, but compromised website which contains malicious Iframe. This Iframe causes the victim's browser to pull the exploit code from a server armed with the exploit toolkit. Upon successful exploitation, a special build of a Trojan, created for the attacker, is being pulled from Golden Cash server. Once installed, the Trojan reports back to the Golden Cash server and the attacker's account at Golden Cash is credited with currency. The first instruction sent by Golden Cash to the victim's machine, is to

install an FTP-grabber (to steal FTP-credentials) ... The victim's machine is now in a pool of infected machines controlled by Golden Cash and being auctioned to other criminals, using a different website for buyers ... The botnet's command and control server uses another website as a proxy that tunnels the bots communication to and from the C&C server. By applying this technique the C&C server remained 'protected' and undetected by security vendors for a longer time." Looking at the Golden Cash case, one can notice the following:

- automatic pay-per-install (including automatic pricing depending on the location of a buyer that varies from 5 USD to 100 USD per 1000 bots);
- automatic reuse of bots;
- information flow in two directions (from and to controller) to support the features above;
- usage of sophisticated malware products – Zeus and Zalupko Trojans;
- bots use FTP grabber to steal FTP credentials;
- using a proxy website by the C2 server.

The case of Golden Cash is over three years old. Considering threat predictions for 2012, we can see that the same botnet trade features are still dominating. Changes are in the architecture of botnets. Instead of a single centralised C2 server, peer-to-peer or hierarchical control is used. This requires more intelligent software and complex cooperation. Up-to-date information about botnet C2 servers can be found on the webpage of the Malware Threat Center of SRI International [6].

Botnets are used also on the defence side. A precedent has been created by the takedown of the Coreflood botnet in 2011 [7]. This takedown was authorised by the US DoJ and was performed by Internet Systems Consortium, Inc. (ISC) in cooperation with the FBI. It also demonstrates how simple it can be to change the side for C2 servers. The Coreflood servers were forced to talk to the FBI software, and shutdown commands were sent to infected computers. This required a Temporary Restraining Order (TRO) from a court.

3. AGENTS IN CYBER SPACE

The first ideas of organising software in the form of agents can be found in the actor model proposed by Carl Hewitt as a model of concurrent computations in the seventies [8]. This model has influenced even the development of object oriented languages. Today's agents can be considered, in essence, as well-developed objects that possess some features of intelligent behaviour.

Agents must have at least proactivity, the ability to communicate, and reactivity – the ability to make some decisions and to act. In software practice of today, agents are usually implemented on some special agent-based computing platform (cf. object-oriented software platforms). This simplifies the development and usage of agents, but it is not a necessary requirement. In the present paper we consider cyber space as an environment for agents, and we use a loose definition of agents as objects with the properties listed above. This is justified by the existing examples of malicious software that have agents' properties and move around in cyber space.

Cyber space requires some robustness and adaptability from agents, i.e. the ability to observe the environment and to use its features (protocols, operating system tables etc.). This is what characterises the advanced malware, and it is predictable that development will continue in this direction.

Probably the most sophisticated malware examples today that have agent properties are Stuxnet and Duqu [9]. They are very intelligent programmes (actually, a set of programmes) that analyse the environment in order to select a target, plan actions, are proactive and behave depending on time. Stuxnet consists of two parts: a delivery part that very selectively infects the control software, and a payload which is an intelligent and stealthy attacker of a special type of Siemens controllers. These parts can be considered as two autonomous agents.

On the other side, using intelligent agents in defence has been described in [10], where simulation shows that cooperating agents can effectively defend against DDoS attacks. After solving some legal [11] and also commercial problems, it should be possible, in principle, to develop a 'cyber police' consisting of mobile intelligent agents. This will require implementation of infrastructure for supporting the cyber agents' mobility and communication, but must be inaccessible for adversaries. This will require cooperation with ISPs. Multi-agent tools can provide a more complete operational picture of cyber space, for instance, a hybrid multi-agent and neural network-based intrusion detection method has been proposed [12]. Agent-based distributed intrusion detection is described in [13].

4. ADVANCED AGENT PROPERTIES

We have to look at some agents' features in order to be able to analyse the consequences of using agents as weapons (or as automatic warriors). These properties are reflection and beliefs-desires-intentions (BDI) – a combination of features that enable the agents to operate autonomously in a goal-oriented way. These are anthropomorphic features, and we must bear in mind that we should not apply any laws of human behaviour automatically to agents when considering these features.

A. Reflection

Reflection is the ability to perceive an agent's own state in the overall situation where an agent operates and to behave according to this perception, i.e. to use this for action planning. Reflection had already been introduced for objects in the eighties [14]. One can distinguish procedural reflection and declarative reflection. The first is implemented by programmes that have access to data describing the agent's/ object's state and, depending on the data, can change the functioning or even the programme of an agent or object (its behaviour in a more general setting).

Declarative reflection is the usage of models of environment and self for action planning [15]. Let us explain it in more detail. First, an agent must have a model that describes the current situation where the agent operates. It is important that this model includes as a part a model of the agent itself (this is the basis for a kind of consciousness that can appear in agents). Second, the agent must have a goal (or goals) presented by some data. Third, the agent must be able,

using these models, to plan its future actions for achieving the goals. In general, planning is a very challenging task. It can be simplified, when specific properties of the environment can be considered.

B. BDI and Emotions

The triplet of features belief-desire-intention got attention in psychology not too long ago, at the end of the last century, after M. Bratman presented his theory of human practical reasoning [16]. It also immediately got the attention of computer scientists for the programming of intelligent agents [17]. A project of the application of BDI in cyber defence has been described in [18].

The idea of BDI is to separate situation awareness from planning and execution of plans. The situation awareness is presented as beliefs – an agent 'believes' that the situation is as the agent 'sees' it. The desires represent a motivational state of an agent; they express the situations that an agent would like to achieve. Goals appear as a result of the analysis of the difference between the situation and the desired situation. Intentions appear as the goals that an agent decides to actively pursue. When a goal has been selected, a respective plan has to be obtained. This can be selected from a library of plans or it can be synthesised on the basis of existing information (beliefs). We present here an anthropomorphic explanation of BDI. Its software implementation is rather straightforward, using knowledge-based software technology. The most complicated part is planning. In the case of declarative reflection, plans are developed on the situation models. An example of planning for declarative reflection support is described in [15].

The steps from beliefs to desires and from desires to intentions depend on the emotional state of an individual or agent. We have not yet agreed on the presentation of emotions in agents. At present, we can speak about priorities instead of emotions. Handling priorities in computers is a common and well understood task.

One can expect that in the future a mechanism will be developed for controlling priorities that can be compared to emotions in human beings. The simplest model of emotions is as follows. Let us have a collection of priorities $p_1, p_2, ..., p_n$ that can control decision making in an agent: selection of goals, immediate reactions of an agent, interpretation of inputs etc. The number of priorities is large. Let us divide priorities into groups $e_1, e_2, ..., e_k$ in such a way that the priorities of one and the same group depend on a state s of the agent in a similar way. The number of groups is much less than the number of priorities: $k << n$. One can say that each group is controlled by an emotion. Thus we can define a small number of functions, $f_1(s), f_2(s), ..., f_k(s)$, for calculating a large number of priorities (a function f_i controls/ calculates priorities of the group e_i). This model can be extended by adding interactions between the groups.

5. THREATS OF AGENT COMMAND AND CONTROL

The agents have to be controlled by stating the most general goals and by giving some initial commands. Specific goals and a detailed action plan will be developed by agents themselves. However, the general command and control model shown in Figure 1 also applies to agents. It has links between its components responsible for command, control, behaviour and situation

awareness. In principle, any of these links can be attacked by an adversary. For instance, if it is true that the US RQ-170 Sentinel drone was captured by Iran [19], then this was obviously caused by an attack on the command and control system of the drone. It is argued that this was possibly done by the disturbance in the link between effects and situation information – wrong GPS data were passed to the control system of a drone.

The command and control of intelligent agents differs from C2 of botnets of today, because the agents have some independence. This makes their behaviour more difficult to predict, and this is a source of threats that can be:

- misinterpretation of commands;
- misunderstanding of a situation;
- unexpected emotions.

Misinterpretation of commands may be the most common threat, but it is also the easiest to avoid in principle. The threat appears if the language of C2, a communication protocol in the simplest case, is not sufficiently verified. Computer science supports verification of protocols, but it is still a complicated task. If an agent communication language is used which is more complicated than messages of a fixed format, then semantic problems of understanding appear. The language should be kept as simple as possible.

Misunderstanding of a situation can lead to wrong decisions at planning and execution stages. It is a threat that is difficult to avoid, because an agent operates in an environment that is complex or even unknown for the designers of the agent. The environment is cyber space, and it is complex with many different operating systems, software platforms, protocols etc. An obvious thing to do is to restrict the environment as much as possible by permitting the agent to operate only on known platforms. Situation awareness of agents must be improved and verified, or at least carefully tested with respect to the safety of their behaviour. A new trend is to apply artificial intelligence and cognitive methods in situation awareness [20]. This permits fusion of human and computer situation awareness and supports real time [21] and automatic [22] decision making.

The agents do not have emotions today, but they have to set priorities in order to be able to plan actions in a reasonable way – performing urgent and important actions first. A simple example of a mistake is setting a wrong priority on the basis of a false alarm. An analogy of a human activity is when someone fears that a threat exists and behaves in panic. It is a complex task to foresee all possible combinations that can appear in selecting priorities on the basis of the situation analysis.

6. MULTI-AGENT THREATS

Agents in cyber operations and cyber defence can be used most efficiently in multi-agent formations. Botnets could be an example, if bots are developed as agents. However, the control in botnets has still remained quite simple. Some cases of multi-agent defence are also available

from the literature [23,24]. One can expect that multi-agent systems will become the main form of agent application in cyber operations. In this case, agents will negotiate between themselves and will cooperatively create a complex behaviour for achieving the general goals stated by a commander. As a consequence, the strict control of behaviour of every single agent will be weaker. Also, it will be more difficult to foresee all possible cases for decision making. Practically, it will be impossible to verify the outcome of multi-agent behaviour for all situations. It is possible that backdoors and forced destruction will have to be built into agents. Multicast control messages will be needed for emergency cases of the agent control. Another option could be self-destruction of agents if loss of contact occurs, i.e. if for some time no command and control messages are received.

A specific threat of multi-agent systems is the formation of unwanted coalitions by agents. This can happen if agents get too much autonomy in decision making. Communication between the agents will be only partially observable to human controllers in this case. This will require very careful selection of constraints on the behaviour of agents. Here is the right place to remind of Asimov's laws for robots. This kind of law could improve the safety of multi-agent systems. However, there will never be an absolute guarantee of avoiding a misunderstanding of a situation by a team of agents. Also, a danger remains that a collection of agents may behave unintentionally in a harmful way. This is analysed in [25] and some possible, although not very probable, scenarios are described there. The next section presents one of these scenarios in a slightly modified way. This example should serve as a warning against neglecting the security of C2 of agents.

7. A SCARY SCENARIO

The year is 2030. Soon after the first attack, the Stuxnet malware was used in attacks on other systems developed by Siemens. It occurred to be a weapon applicable to various supervisory control and data acquisition (SCADA) systems, as soon as a system's design is known. Its intelligence was developed further with the aim of autonomously penetrating target systems. Its payload was adjusted to the target each time before launching.

Different cyber weapons were developed for performing different autonomous attacks. All these programmes can be called agents. They are quite autonomous, use BDI and declarative reflection, and can operate independently in an unfriendly environment.

As a consequence, botnets – the centrally controllable sets of passive programmes have evolved into armies of quite intelligent artificial fighters commanded in a net-centric way.

The intelligent malware caused much harm to the infrastructure of countries until multi-agent systems were also built for the defence. The defending agents were supported by advanced multi-agent platforms that gave them considerable advantage (in particular, good communication) compared with the attacking agents who had to operate in an unfriendly environment in a stealthy way. In order to further improve the capabilities of the defending agents, their autonomy was extended and their BDI system was developed more than ever before. This gave them an

excellent ability to plan their actions and even to set up new goals. This was very convenient for most of the users, and the general security awareness of people decreased to some extent.

The year is 2045. It was a bad idea to use too many agents with BDI. The danger was not so much in the intelligence of the weapons as in their willingness (and ability) to pursue their own goals. It became difficult to control very intelligent agents who had consciousness, priorities controlled by something similar to emotions, and who had their own desires.

A cyber conflict occurred between the agents that was initiated by the agents themselves. The country of the defending agents was immediately known, but the attacking agents seemed to belong to several different countries. It looked like there was a coalition of attacking agents from several countries. A lot of diplomacy was needed to clarify the case. A danger remains that agents may build hostile coalitions.

8. CONCLUDING REMARKS

The scenario presented above assumes the development of intelligent cyber weapons that are difficult to control. This is, in principle, a possible scenario. It is not based on any idea of artificial general intelligence (AGI) considered by the Singularity Institute for Artificial Intelligence in Palo Alto [26]. The AGI is based on an assumption that unsupervised learning capabilities of programmes will lead to an explosive growth in knowledge and intelligence of computers. Although possible in principle, and applied in data mining and parametric learning, the unsupervised learning has not developed to be applicable in learning on the conceptual level needed for understanding the world in general, and there are no signs of this possibility for the foreseeable future.

We have used the concept of agent for denoting a variety of cyber weapons of the future. This concept is used to denote just a set of features that provide autonomy, mobility and proactivity to the software under consideration. This has enabled us to analyse command and control of new cyber weapons in a unified setting, ignoring details of specific weapons. We have discussed the threats that are caused by agents, and we have made some unconventional predictions, assuming that the development of the cyber weapons will continue with acceleration. The future may not be as predicted here, but there is still good reason to be aware of the dangers described in the last sections of the paper.

We can point out some general conclusions. First of all, the more intelligent software becomes the more difficult it will be to control it. When designing and developing new cyber weapons, one has to be very cautious about guaranteeing the appropriate control over the weapons under any circumstances. It is practically impossible to use formal methods for verifying the safety of intelligent cyber weapons for their users. The global risks of wide implementation of artificial intelligence are analysed in [26].

One possible way to increase the safety seems to be imposing strict constraints on the behaviour of agents. This will be the analogy of the introduction of Asimov's laws on agents. However,

it will be still impossible to verify the correctness of behaviour of agents with respect to these constraints.

ACKNOWLEDGEMENTS

This research was supported by the Estonian Ministry of Education and Research target-financed research theme no. 0140007s12.

REFERENCES:

[1] *"The Cyberwarfare Market 2012-2022"*, [Online]. Available: http://www.visiongain.com/Report/732/The-Cyber-Security-Market-2012-2022, Dec. 05, 2011 [Feb. 7, 2012].

[2] J. A. Kaplan, *"Japan Reportedly Building Vigilante Virus Assassin Squad,"* Discovery News, [Online]. Available: http://news.discovery.com/tech/japan-vigilante-virus-120104.html, Jan. 4, 2012 [Feb. 11, 2012].

[3] David S. Alberts Richard E. Hayes, "Understanding Command And Control," *CCRP Publication Series*, DoD, [Online]. Available: www.dodccrp.org/files/Alberts_UC2.pdf [Feb. 7,2012].

[4] *Proc. IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 2011, ISBN: 978-1-61284-784-9.

[5] E. Mills, *"Golden Cash botnet-leasing network uncovered,"* [Online]. Available: http://news.cnet.com/8301-1009_3-10266977-83.html, June 17, 2009 [Feb. 7, 2012].

[6] *"Most Prolific BotNet Command and Control Servers and Filters"* [Online]. Available: http://mtc.sri.com/live_data/cc_servers/, update daily [Feb. 07, 2012].

[7] S. Ragan, *"Coreflood: Botnet takedown introduces a potentially risky precedent"*, [Online]. Available: http://www.thetechherald.com/articles/Coreflood-Botnet-takedown-introduces-a-potentially-risky-precedent/13363/ Apr. 18, 2011 [Feb. 07, 2012].

[8] C. Hewitt, P. Bishop and R. Steiger, "A Universal Modular Actor Formalism for Artificial Intelligence," in *Proc. IJCAI*, 1973.

[9] R. Langer, "Stuxnet: Dissecting a Cyberwarfare Weapon", in *IEEE Security and Privacy, v.* 9, 2011, pp. 49 - 51.

[10] I. Kotenko, A. Konovalov, A. Shorov, "Agent-Based modelling and Simulation of Botnets and Botnet Defence," in *Proc. Conference on Cyber Conflict 2010*, C. Czosseck, K. Podins (Eds.), *CCD COE Publications*, Tallinn, Estonia, 2010.

[11] B. Stahl, D. Elizondo, M. Carroll-Mayer, Y. Zheng, K. Wakunuma, "Ethical and Legal Issues of the Use of Computational Intelligence Techniques in Computer Security and Computer Forensics," in *WCCI 2010 IEEE World Congress on Computational Intelligence, Barcelona, Spain*, 2010, pp. 1822 – 1829.

[12] E. Herrero, M. Corchado, A. Pellicer, A. Abraham,"Hybrid multiagent-neural network intrusion detection with mobile visualization," in *Innovations in Hybrid Intelligent Systems*, vol. 44, 2007, pp. 320–328.

[13] V. Chatzigiannakis, G. Androulidakis, B. Maglaris, *"A Distributed Intrusion Detection Prototype Using Security Agents,"* HP OpenView University Association, 2004.

[14] P. Maes, "Concepts and Experiments in Computational Reflection," in *Proc. OOPSLA*, 1987, pp. 147-155.

[15] M. Addibpur, E. Tyugu, "Declarative Reflection Tools For Agent Shells," in *Future Generation Computer Systems*, July 1996, pp. 1 - 12.

[16] M. Bratman. *Intention, Plans, and Practical Reason, CSLI Publications*. 1999.

[17] M. Georgeff, B. Pell, M. Pollack, M. Tambe, M. Wooldridge, "The Belief-Desire-Intention Model of Agency," in *Proceedings of the 5th International Workshop on Intelligent Agents V, Agent Theories, Architectures, and Languages*. Springer-Verlag, London, UK 1999.

[18] M. Shajari, A. Ghorbani, "Application of Belief-Desire-Intention Agents in Intrusion Detection & Response," in *Proc. Second Annual Conference on Privacy, Security and Trust*, 2004, Fredericton, NB E3B9W4, 2004, pp. 181 - 190.

[19] *"RQ-170 Sentinel Drone Downed In Iran Critical Updates,"* [Online]. Available: http://aviationintel.com/2011/12/08/downed-rq-170-sentinel-drone-critical-updates/, Dec. 8, 2011 [April 1, 2012].

[20] R. Jones, E. Connors, M. Endsley, "A Framework for Representing Agent and Human Situation Awareness," in *Proc. IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 2011.

[21] T. Saarelainen, J. Timonen, “Tactical Management in Near Real-Time Systems,” in *Proc. IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 2011.
[22] Y. Fischer, A. Bauer, J. Beyerer, “A Conceptual Framework for Automatic Situation Assessment,” in *Proc. IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 2011.
[23] V. Gorodetski, I. Kotenko, “Multi-Agent Systems for Computer Network Security Assurance: Frameworks and Case Studies,” in *Proceedings of the 2002 IEEE International Conference on Artificial Intelligence Systems (ICAIS’02)*, IEEE Computer Society, 2002.
[24] K. Boudaoud, Z. Guessoum, “A Multi-agents System for Network Security Management,” in *SMARTNET ‘00 Proceedings of the IFIP TC6 WG6,7 Sixth International Conference on Intelligence in Networks: Telecommunication Network Intelligence*, Kluwer, The Netherlands, 2000.
[25] E. Tyugu, “When computers become too smart,” in *Information Modelling and Knowledge Bases XXIII*, J. Henno, Jaak, Y. Kiyoki, T. Tokuda et al. (Eds.), IOS Press, Amsterdam, 2012, (Frontiers in Artificial Intelligence and Applications v. 237), 2012, pp. 412 - 418.
[26] E. Yudkowsky, “Artificial Intelligence as a Positive and Negative Factor in Global Risk,” in *Global Catastrophic Risks*, N. Bostrom, M. Ćirković (Eds.), Oxford University Press, 2008, pp. 308–345.