U.S. DEPARTMENT OF ENERGY

PNNL-18995

Towards a Research Agenda for Cyber Friendly Fire

Frank L. Greitzer
Samuel L. Clements
Thomas E. Carroll
J.D. Fluckiger

Prepared for the Air Force Research Laboratory Human Effectiveness Directorate

November 9, 2009

Pacific Northwest
NATIONAL LABORATORY

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights**. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America
Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161
ph: (800) 553-6847
fax: (703) 605-6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/ordering.htm

Contents

Abstract

Historical assessments of combat fratricide reveal principal contributing factors in the effects of stress, degradation of skills due to continuous operations or sleep deprivation, poor situation awareness, and lack of training and discipline in offensive/defense response selection. While these problems are typically addressed in R&D focusing on traditional ground-based combat, there is also an emerging need for improving situation awareness and decision making on defensive/offensive response options in the cyber defense arena, where a mistaken response to an actual or perceived cyber attack could lead to destruction or compromise of friendly cyber assets. The purpose of this report is to examine cognitive factors that may affect cyber situation awareness and describe possible research needs to reduce the likelihood and effects of "friendly cyber fire" on cyber defenses, information infrastructures, and data. The approach is to examine concepts and methods that have been described in research applied to the more traditional problem of mitigating the occurrence of combat identification and fratricide. Application domains of interest include cyber security defense against external or internal (insider) threats.

Introduction

Historical assessments of combat fratricide reveal principal contributing factors in the effects of stress, degradation of skills due to continuous operations or sleep deprivation, poor situation awareness (SA), and lack of training and discipline in offensive/defense response selection. While these problems are typically addressed in research and development focusing on traditional ground-based combat, there is also an emerging need for improving SA and decision making on defensive/offensive response options in the cyber arena, where a mistaken reaction to an actual or perceived cyber attack could lead to destruction, degradation or compromise of friendly cyber assets. The purpose of this paper is to discuss research issues underlying cyber SA—particularly cyber friendly fire—toward the development of a research agenda to control and reduce the likelihood or impact of cyber "fratricide." We believe that key areas of study should include the examination of cyber security decision making, available tools and technologies, and the set of tactics/policies/procedures that comprise the cyber security environment. An analysis of possible threats to proper use of these resources is expected to reveal flaws in human machine systems that adversely affect SA as well as research requirements that address gaps/needs in decision aids, technologies, and processes.

Application domains of interest may include cyber security defense against external or internal (insider) threats. Because cyber defense responses often must be applied within seconds, there is a need to consider factors involved in establishing and validating automated responses based on defined a priori courses of action (COAs). Such automated responses would be expected to mitigate operator-under-pressure errors that are associated with more traditional combat identification contexts, but may increase the risk of cyber fratricide. Indeed, new issues arise that center around the degree of autonomy and the associated risks and consequences of enacting prescribed COAs without human intervention.

Definition and Rationale

There is no generally accepted definition of cyber fratricide or cyber friendly fire. There is an understandable desire to produce a definition of cyber friendly fire/fratricide that is consistent with the accepted definition of combat fratricide, as defined by the U.S. Army's Training and Doctrine Command (TRADOC) Fratricide Action Plan: Fratricide is "the employment of friendly weapons and munitions with the intent to kill the enemy or destroy his equipment or facilities, which results in unforeseen and unintentional death or injury to friendly personnel" (US Department of the Army, 1993, p. 1).

After consideration of several perspectives on the problem[1], we define cyber friendly fire as

> "Intentional, offensive or defensive cyber/electronic actions intended to protect cyber systems against enemy forces or to attack enemy cyber systems, which result in inhibiting, damaging or destroying friendly infrastructure or operations."

Important aspects of this definition are that it includes both defensive and offensive actions; and it specifies intentional actions by friendly forces. Thus, for example, insider activity that harms friendly assets is excluded from the definition. There is some controversy about whether or not to include defensive actions in the definition. Traditional warfare only considers friendly fire as injuries or deaths of friendly forces sustained from actions that were intended for an enemy. Thus an alternative definition for cyber friendly fire would only include actions that are intended to harm the enemy but result in injury to friendly assets. While this is true in the case of combat friendly fire, we believe that for the cyber warfare context, active defenses also can be implicated in cyber friendly fire. Two analogies will help illustrate the point:

(a) Suppose that a new type of coating is invented that makes airplanes invisible to radar, and this defensive measure is applied to paint military aircraft. Now suppose

[1] Much discussion of alternative definitions took place at a seminal workshop devoted to the issue, hosted by the Air Force Research Laboratory (AFRL) and organized by AFRL and PNNL: *Cyber Friendly-Fire Avoidance Workshop,* Colorado Springs, CO, Feb 3-4, 2009.

that the application of this coating disrupts the aerodynamics of some models of planes, causing crashes and loss of life. Would this be considered fratricide? Clearly, this would not be classified with cases of friendly fire. It would be considered an accident that resulted from an engineering/design flaw. This argument, in particular, has been made to promote a definition of cyber fratricide that includes offensive but excludes defensive actions.[2]

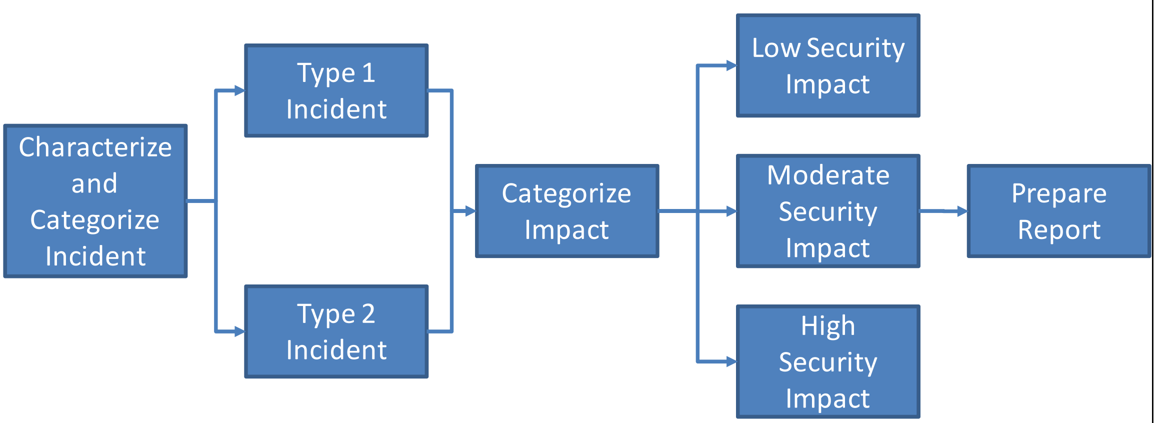
(b) Now suppose a number of airplanes are engaged in combat and a friendly aircraft deploys flares to deflect an enemy attack. These flares are then ingested into another friendly aircraft's engine and disable that airplane. This case, we argue, should be considered friendly fire. Even though the flares were not intended to harm the enemy they still caused damage to friendly forces. Following this line of thinking, we argue that actions taken to *actively defend* against cyber threats and attacks, that have unintended consequences to inhibit, disrupt or compromise friendly cyber systems, should be classified as instances of cyber friendly fire.

Cyber Security Incident Reporting Process

Figure 1 illustrates the process that has been defined by the Office of the Chief Information Officer, U.S. Department of Energy. [3] Once a cyber security event occurs, the cyber security officer must report incidents that are significant or unusually persistent and that meet the criteria identified in this process. The incident must be characterized and categorized according to its potential to cause damage to information and information systems, based on criteria of incident type and security category. While it is not explicitly stated in the DOE incident management documentation, we can recognize elements of SA in the classification requirements of Figure 1.

[2] This argument and example provided by Dr. Kamal Jabour, AFRL, at the *Cyber Friendly-Fire Avoidance Workshop*, Feb 3-4, 2009.

[3] http://www.doecirc.energy.gov/incidentreporting.html



Department of Energy Cyber Security Incident Reporting Process

Required Time Frame for Reporting Cyber Security Incidents

Incident Type	Security Category		
	Low	**Moderate**	**High**
Type 1	Within 4 hours	Within 2 hours	Within 1 hour
Type 2	Within 1 week	Within 48 hours	Within 24 hours

Figure 1. Cyber Security Incident Management

Type 1 incidents are successful incidents that potentially create serious breaches of DOE cyber security or have the potential to generate negative media interest. Examples are system compromise/intrusion; loss, theft, or missing computers or national security information; web site defacement; malicious code; denial of service; critical infrastructure protection; unauthorized use; or information compromise. Type 2 incidents are attempted incidents that pose potential long-term threats to cyber security interests or that may degrade overall effectiveness of DOE's cyber security posture. Examples are attempted intrusions or reconnaissance activity (probes, scans, or social engineering).

It can be seen that different types of incidents are assigned different reporting times, ranging from one hour to one week. On the face of it, this does not seem to present a particularly challenging timeline, except when we consider the magnitude of the threat, i.e., the volume of attacks that occur. Figure 2 (from Symantec, 2009) provides a glimpse of the magnitude of the threat. This chart shows the explosive proliferation of customized malicious code and phishing kits as compiled by Symantec. In 2008, Symantec detected more than 1.6 million new malicious code threats that required development of new signatures so they could be found by malware monitors—this represents over 60 percent of the approximately 2.6 million malicious code threats that Symantec has detected in total over time. Clearly, the number of malicious code signatures that are being addressed annually has increased beyond any unaided human capability to respond!

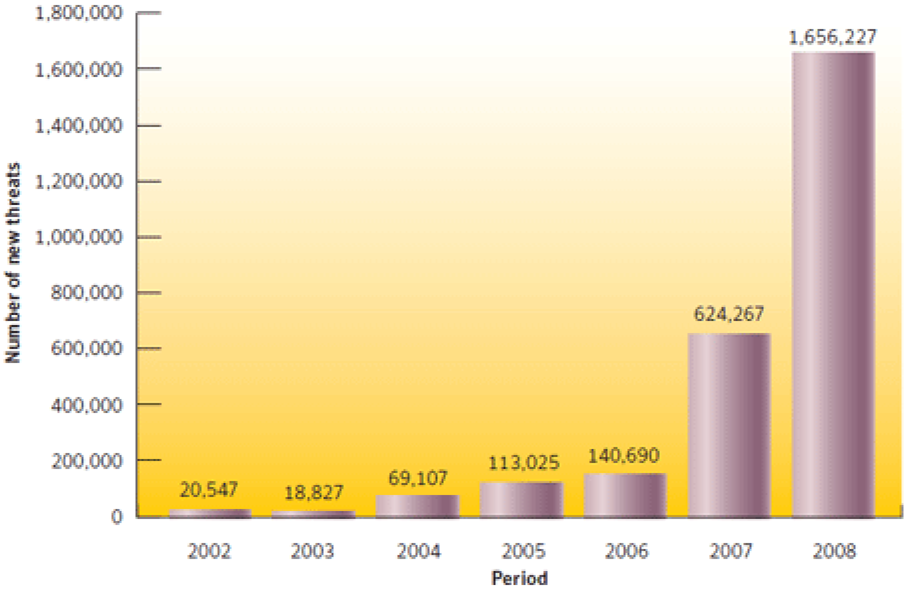


Figure 2. New Malicious Code threats

Source: Symantec (2009)

From a human factors and information processing standpoint, it is important to consider the effects of workload and stress on SA and decision making. Errors, biases, and other information processing deficiencies (as well as organizational factors such as staffing, training, etc.) will impact performance and have the potential to produce performance deficiencies that may lead to cyber friendly fire cases. The next section describes some representative cyber friendly fire cases that have been reported.

Taxonomy

To help delimit the domain of cyber friendly fire, consider the following categories that may or may not be considered to fall in the domain of cyber friendly fire (see also Table 1):

- *Offensive Cyber Action*. We consider any offensive action that causes incidental damage to friendly resources to be cyber friendly fire.
- *Defensive Cyber Action*. We allow that defensive cyber actions that lead to damage or limitations of friendly resources should be considered cyber friendly fire incidents.
- *Insider Threat*. Malicious actions by trusted individuals within an enterprise are examples of insider threat. We do not consider malicious insider threat to be cyber friendly fire, specifically because the intent of the individual is to harm the organization. Further, inadvertent actions by insiders that harm friendly cyber resources should not be considered examples of cyber friendly fire, because there was no intent (or even awareness) by the individual. Because it does not "fit" the analogy to combat fratricide, we do not consider accidental actions by an actor to be cyber friendly fire when the actor is not actively trying to protect resources, attack an enemy via cyber means, or defensively respond to a cyber incident.
- *Cyber Actions Causing Accidental/Inadvertent Damage*. Cyber actions that are unrelated to offensive/defensive tactics that accidentally cause damage to or limit friendly cyber resources are not considered to be examples of cyber friendly fire.

Table 1. Summary Taxonomy of Cyber Actions that Damage/Limit Friendly Resources

Type:	**Offensive Cyber Action**	**Defensive Cyber Action**	**Insider Threat**	**Accidental Damage**
Friendly Fire?	Yes	Yes	No	No
Example:	Case Study #1	Case Study #2 Case Study #3 Case Study #4 Case Study #5 Case Study #6	N/A	• Cellphone interferes with wireless network • Equipment malfunction

Below we discuss some illustrative cases.

Cyber Friendly Fire Cases

Here we describe some representative case studies that elucidate critical issues and needs to be addressed involving cyber friendly fire. We identify relevant issues (such as automation considerations, SA/decision functions, etc.) that apply to these cases, recognizing that the issues are not necessarily independent, and that examples described for one type of issue or area might well be considered to apply in other areas.

Offensive Cyber Action (Cyber Warfare)

Discussion of possible offensive cyber actions is beyond the scope of this report. Hypothetically, we can acknowledge the possibility of such events, methods, and tools. An example is the possible development of a cyber weapon for US Military infantrymen that might inadvertently damage friendly cyber systems.

Another example is an offensive action (such as a Distributed Denial of Service attack, or DDoS) against an external entity (e.g., DNS provider) that caused a widespread outage. The Pentagon is reportedly developing a cyber warfare device that can be used by frontline infantrymen[4]. The tool purportedly will have the capability to tap into satellite communications, any wireless network, voice over Internet (VoIP), and proprietary SCADA networks. Currently the knowledge of how to attack and compromise these systems is known by relatively few people with deep and intricate knowledge. These elite few tend to understand the consequences of particular actions and how to limit damage. Providing this powerful tool to so many individuals without such comprehensive domain knowledge will inevitably result in misuse, whether intentional or accidental, that will degrade or even compromise friendly systems.

In this case it seems likely that many of the causes linked to combat fratricide; namely stress and fatigue, (Greitzer and Andrews, 2008) will play a part in the

[4] http://www.aviationweek.com/aw/generic/story_channel.jsp?channel=defense&id=news/CYBER052109.xml

deployment of this weapon. Since working with computers involves fast and accurate keyboard work and even faster thought processes, stress and fatigue will also play a significant role in these scenarios.

Illustrative Case Study #1: A current example of cyber warfare gone wrong was published by the Chinese Xinhua News Agency[5]. Reportedly a distributed denial of service attack (DDoS) against DNSPod, a Chinese DNS provider and domain registrar, was conducted by feuding underground online gaming service providers. The attack was an attempt to force users away from gaming servers linked to by DNSPod toward rivaling servers. Unfortunately, DNSPod was also the DNS server for Baofeng, a very popular video-streaming service. When requests for video were denied they cascaded upstream to higher-level DNS servers which did not know how to respond effectively creating a multiplier effect on the DDoS. An attack that was apparently intended to knock-out a few servers affected millions of Internet users in over 20 Chinese provinces.

Often both friends and foes are using some of the same network services. Attacking them will inevitably affect both parties. How is this mitigated? When is it appropriate? How is this determined?

Defensive Cyber Action

Defensive cyber actions that damage or limit friendly assets may derive from automated responses that have damaging effects, purposeful actions with inadvertent effects, or overreactions. A basic, underlying factor for most of these issues is a lack of, or insufficient SA.

Automated Response

The speed at which cyber threats change make manual mitigations ineffective, but what level of autonomy is adequate and acceptable? A clear tradeoff is apparent: By

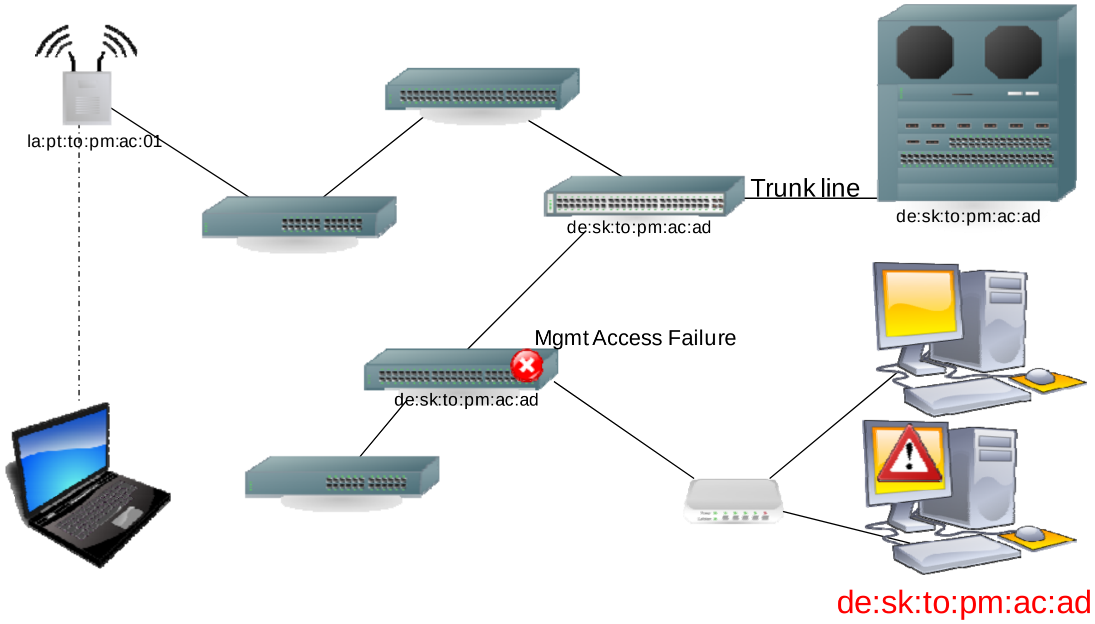
[5] http://www.darkreading.com/story/showArticle.jhtml?articleID=217701926

removing the human from the loop, it is likely that more false positives will occur; but requiring a human to respond to all of today's threats requires massive manpower.

Another issue may be the actions taken automatically may be mis-directed and cause damage to a third party who is completely innocent of any wrongdoing and in fact had nothing to do with the damage inflicted as often it is difficult or impossible to determine the actual bad party. This could even result in a legal suit against you for damage, loss of profit, loss of business, damage to the reputation of the aggrieved party.

Illustrative Case Study #2**:** The operations team at Pacific Northwest National Laboratory constantly monitors network traffic for potentially malicious/unauthorized activity, and if such activity is detected, steps are taken to block or remove it. As the network has grown and the number of incidents has increased, so has the need for an automated response to deal with the increased effort to maintain the network. The operations team began evaluating an automation tool that would automatically disable network switch ports when malicious or unauthorized traffic was detected coming from them. A system performance evaluation determined that the system would function as expected the majority of the time, but there were instances where the management server on the switch would fail. In this case the tool would cascade upstream to the next switch port detecting the traffic and block the offending port on that switch. In effect, this action denies service not only to the offending device but to all devices connected to the original switch. In one instance it would have disabled the network for an entire building. If the traffic were sufficiently malignant, this may be an appropriate response. However, in most cases this denial of service would constitute "overkill" that causes more damage than the malignant traffic alone could produce.

A research challenge for design and development of automated defensive response systems is determining the most appropriate point in the hierarchy to "aim" the response. If a high level IP address is disabled because of inappropriate content at a site located lower in the IP address hierarchy, other unrelated sites that are built from that IP address will be blocked unnecessarily, which may result in financial or other loss.



Case Study #2. Large, complex network architectures challenge the security analyst's ability to maintain an appropriate mental model of the situation—leads to unintended adverse consequences of defensive actions.

In addition to research challenges relating to improving performance and reliability of automated systems, the increasing use of automation brings issues of trust and complacency. Trust in automation should reflect its reliability: as reliability decreases, our trust should decrease (Muir (1987). If trust is not directly related to reliability, it is referred to as *mistrust*. Human trust in automation can be low (*distrust*) or too high (*overtrust*) (Parasuraman & Riley, 1997). Distrust is a type of mistrust where the person fails to trust automation as much as appropriate. It may result from a failure to understand the nature of the automated algorithms that produce the output. When distrust leads an operator to reject good assistance that automation can offer, the consequences can be harmful in addition to merely increasing inefficiency. For example, distrust of faulty automated warning systems can lead to the real danger of ignoring legitimate alarms (Sorkin, 1988).

In contrast to distrust, overtrust of automation (also called complacency) occurs when people trust automation more than is warranted. This can have severe negative consequences if the automation is less than fully reliable (Parasurman et al., 1993; Parasuraman & Riley, 1997). Complacency leads to the failure to monitor adequately, which leads to problems in the infrequent circumstances when automation fails and human intervention is necessary. A complacent operator will likely be slower to detect a real failure. In addition, complacency, which promotes less active participation, tends to produce operators who have a lower level of SA (Sarter & Woods, 2000).

How trustworthy are the tools that aid cyber defender/attackers? How much trust is placed in the tools? Is it an appropriate level where the operator knows the limitations and constraints of the tools or is the trust blind? What can be done to create an appropriate level of trust? How might that level be determined?

The nuclear accident at Three Mile Island shows an inappropriate trust in the tools.

Illustrative Case Study #3**:** The partial core meltdown of Unit 2 of the Three Mile Island Nuclear Generating Station in 1979 is considered the most significant nuclear accident in the history of American commercial nuclear power generation[6]. One of the root causes attributed to this accident was a pilot-operated relief valve (PORV) malfunction. The PORV would open when electrical power was applied to a solenoid. There was an indicator light in the control room that lit up when electrical power was applied to the solenoid. This light was intended to indicate when the PORV was open. In the case of Three Mile Island, the core began to overheat and pressure began to build. The system, as designed, opened the valve until the pressure was within constraints, and then the power was cut; this turned off the indicator light, but the PORV malfunctioned and did not close. The operators in the control room, seeing that the indicator light had extinguished, trusted the valve had closed. There were many other indicators that the PORV had not shut properly, but the operators were blinded by their trust in the tools. The failed PORV was not discovered until a new shift relieved the operators and began to look at the other indicators. This accident essentially ended new construction of nuclear

[6] NRC: Fact Sheet on the Three Mile Island Accident http://www.nrc.gov/reading-rm/doc-collections/fact-sheets/3mile-isle.html

powered electrical generation in the United States—no new plants have broken ground since, and many of those under construction were closed down and never finished. In this case it is obvious that the system design was flawed. The indicator light should have been engineered such that it showed the actual state of the PORV, not whether or not there was electricity applied to the PORV solenoid.

How often are cyber analysts misled by the tools they rely on? An intelligent adversary could craft an attack that causes friendly operators to act or react in a way that helps the adversary achieve his goal because of the trust in the tools.

Purposeful Actions with Inadvertent Effects

Responding to cyber attacks by taking defensive actions to limit their effects always carries the risk that the defensive action can lead to more serious consequences than the original attack. It is therefore crucial for decision makers to analyze risks associated with alternative response options to decide on a response that causes the least damage to friendly assets and operations. The timeframe for making one of these decisions is often very short as is the time to take the action determined to be appropriate.

The response necessary to a given event may differ depending upon the value and priority of the affected system. A response to a given threat can have drastically different consequences than the same response to a different threat. Consider an example where one response was used on four different computers:

***Illustrative Case Study #4*:** An attacker has used a zero day exploit to gain access to the victim's computer. In response, the machine is powered down. Four different victims' machines are: a federal contractor's payroll processing machine; a life support machine at a hospital; a federal travel scheduling system; and a control center workstation at an electrical power utility.

For the payroll processing system, the most obvious consequence is the inability to process payroll. This is definitely a high concern and will inevitably damage the credibility of the contractor. A lesser known consequence is that the potential loss of personally identifiable information (PII) must be reported within 45 minutes[7] of the discovery of the compromise to the appropriate agency or face fines, sanctions, and possibly the right to work. For the hospital life support system, a life threatening situation is created when the affected machine is turned off, terminating life support functions. The case of the travel scheduling system is similar to an event that actually occurred,[8] affecting the travel plans and payment for thousands of employees who were faced with the option of not traveling or paying for travel costs themselves. In that case, further analysis revealed that the system did not have any sensitive data and should not have been disconnected. In the fourth example involving an electric power control system, we note that this can lead to sanctions resulting from the North American Electric Reliability Corporation (NERC) Critical Infrastructure Protection (CIP) standards. Violations of standards can result in significant fines of $1 million per day per infraction that could result when the system is shut down.

Computer networks and systems have become extremely complex. Understanding the downstream effects of actions is essential to limit cyber friendly fire. Often, something as routine as applying a software update will have unintended consequences, as demonstrated in the following case study.

Illustrative Case Study #5**:** On March 7, 2008 the Hatch nuclear power plant was forced into an emergency shutdown for 48 hours after a software update was applied to a monitoring system. The Nuclear Regulatory Commission (NRC) event notification report number 44046[9] records that Unit 2 scrammed because of low water levels as a result of condensate feed-water. Further investigation[10] revealed that a software upgrade was applied to a machine located in the business network. The machine was used to monitor chemical and diagnostic data on a primary control system computer. The upgrade was intended to synchronize both systems. After the upgrade the business system rebooted, causing values to reset, which in turn caused the values on the primary control system to reset as well. The reset values triggered an alarm in the reactor's safety system

[7] http://www.doecirc.energy.gov/incidentreporting.html

[8] http://voices.washingtonpost.com/securityfix/2009/02/travel-booking_site_for_federa.html

[9] http://www.nrc.gov/reading-rm/doc-collections/event-status/event/2008/20080310en.html

[10] http://www.washingtonpost.com/wp-dyn/content/article/2008/06/05/AR2008060501958.html

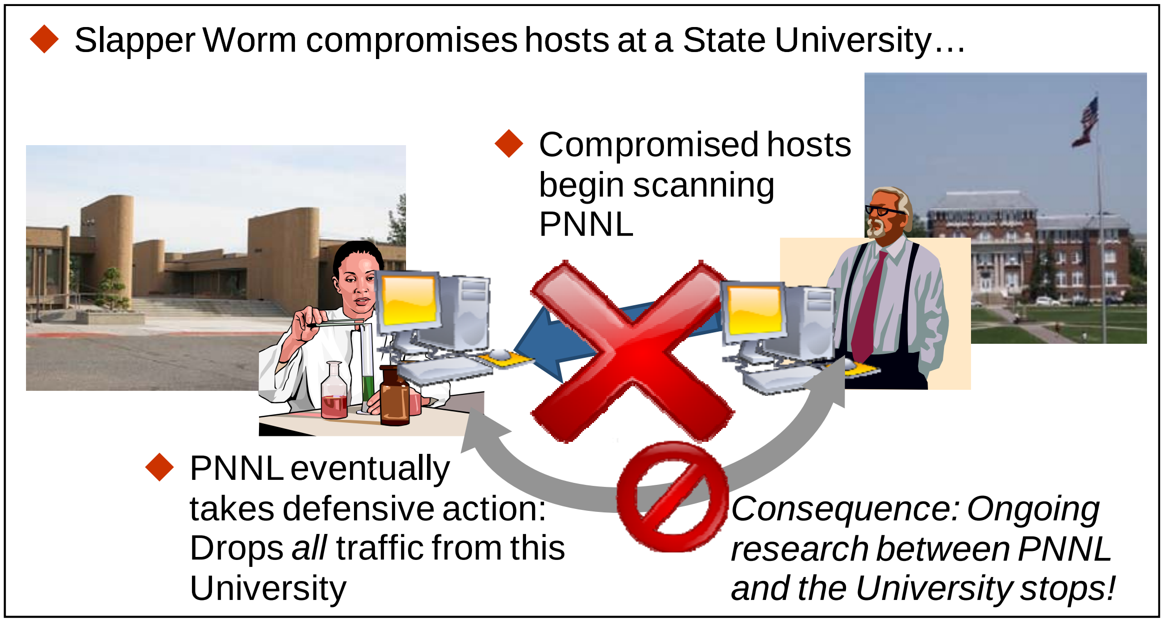
that shut the plant down. The lack of SA created a significant monetary loss to the plant operator.

How are these interdependencies found? What can be done to limit unintended consequences from making changes to these systems?

Overreaction

***Illustrative Case Study #6*:** In another case the Pacific Northwest National Laboratory (PNNL) began recording a massive number of penetration attempts from a state university. PNNL's system administrators contacted the university to find the cause. The university did not readily know the cause and the problem continued to grow. Finally PNNL made the decision to block all traffic from the university until their systems could be cleaned up. It should be noted that this was the only option available to stop the growing threat due to the way the network was configured. The PNNL system administrators later learned that scientists at PNNL had an ongoing research project with the university in question, which was disrupted because of the decision to block all traffic.

In another case like the Slapper Worm, the original infection came as the result of a staff member on travel getting on another organization's network and networking back into PNNL with a VPN. On the user's side of the communications link another system on the network compromised the visiting system. When the staff member got on the PNNL network via VPN, the malware was already on his system and was therefore able to get onto the PNNL main network. Once again, the only available solution was to shut down any communication with the other organization. This impacted numerous projects for almost a week while PNNL cyber security analysts determined that the problem had been addressed on the other organization's network.



- Slapper Worm compromises hosts at a State University…
- Compromised hosts begin scanning PNNL
- PNNL eventually takes defensive action: Drops *all* traffic from this University

Consequence: Ongoing research between PNNL and the University stops!

Case Study #6. Lack of more specific visibility of system status and locus/origination of compromised units leads to a "gross" response that denies service to everyone.

Tradeoffs always exist when taking any action. When are drastic measures appropriate? How is it possible to determine the appropriateness? How can one quickly determine the break-even point between potential compromise and certain loss of productivity? What is the best way to determine this?

Research Agenda

At the outset, it is reasonable to ask whether the factors affecting SA and combat identification (as described by Greitzer and Andrews, 2008) will have similar affects on cyber SA and friendly fire—i.e., there is a need for continued research in a cyber security context to address problems related to training, education, limiting fatigue, and improving SA for cyber defense weapons. Research should address what types of techniques for improving training and preparedness, and generally SA, will be most effective.

Greitzer and Andrews (2008) discussed contributing causes of combat fratricide from cognitive/research perspective. Lack of SA or failures in SA is a primary factor in combat fratricide, and underlying causes are often related to stress, expectancy/bias, failure to follow policy, and training deficiencies. These factors are likely to be relevant in cyber friendly-fire incidents as well. In this section, we consider which of these human factors are implicated in cyber friendly-fire, and whether or not other factors should also be considered.

A report produced by the U.S. Army's CALL center cited primary causes of fratricide (US Department of the Army, 1992) as poor SA, combat identification failures, and weapons errors; with contributing factors including anxiety, confusion, bad weather, inadequate preparation, and leader fatigue. Wilson, Salas, Priest, and Andrews (2007) examined human factors literature for underlying human factors causes of friendly fire incidents. As argued by Wilson et al., to accomplish tasks on the battlefield requires cognitive processes, performed as a collective effort that requires shared cognition. Using a human-centered approach, they concluded that in the absence of adequate shared cognition, warfighters can have problems interpreting cues, making decisions, and taking correct action. They concluded that when shared cognition "fails," the incidence of fratricide increases. They derived a taxonomy of behavioral markers that may help military leaders reduce the consequences of fratricide in war and they identified factors (based on the individual, task, organization, technology, and environment) that influence shared cognition. Addressing CID and fratricide requires mitigation strategies to reduce human errors and better prepare war fighters for factors that undermine SA. These factors—centering on stress, emotion, cognitive biases, and training—apply to individual and group decision makers, i.e., human combat personnel. It is also relevant to consider the effects of technology (automation) on combat identification errors, and we suggest that automation is a particularly important factor in cyber friendly fire.

The foregoing description of some case studies suggests that, in addition to the factors influencing performance in more traditional combat fratricide contexts[11] such as SA, Stress, and the effects of training, specific research needs should be addressed in areas that are specifically tied to cyber SA, cyber defense, and cyber warfare—for example, issues of automation and attribution/identification. Table 2 summarizes factors that we deem to be worthy of consideration in establishing a research agenda for cyber friendly fire. These factors are discussed briefly below.

[11] See, in particular, Greitzer & Andrews (2008) for a more detailed discussion in this context.

Table 2. Factors Underlying Cyber Friendly Fire

Traditional Factors:

- Situation Awareness
- Training Issues
- Stress Effects

Factors Particularly Associated with Cyber Domain

- Attack Attribution
- Automation

Situation Awareness

It is difficult to gain a holistic understanding of any modern network larger than a handful of computers. Knowing what resources are needed, where they are needed, what paths of communication are mission critical and which are ancillary, which systems contain sensitive information, who has access to the computer resources versus who has authorized access to the resources and a myriad of other details are essential for a complete understanding. Add to this the dynamic and ever changing nature of most networks, the barrage and variety of attacks that constantly attempt to compromise these networks[12] and the never ending release of vulnerabilities and it quickly becomes apparent that maintaining SA is very difficult. Yet gaining this awareness and understanding is essential to limit cyber friendly fire.

There are numerous technology-based resources available to aid in SA. Penetration testing tools and assessments help identify resources and services available which can be compared to policies to identify holes and violations; yet they do not provide context or analysis of actions taken to remove these "vulnerabilities." Patch management services detect and patch machines but often break applications running on those machines as illustrated in case study #5. Tools like Nagios[13] or Zenoss[14] monitor system availability and alert when something goes wrong but are not predictive in nature.

[12] See Symantec's Internet Security Threat Report Volume XIV: April, 2009 http://www4.symantec.com/Vrt/wl?tu_id=gCGG123913789453640802

[13] http://www.nagios.org/

Another potential research area that will improve SA and response effectiveness is development of methods for risk assessment and prioritization. Since it is impossible to prevent all attacks, it is the case that some attacks will occur and penetrate the defenses. A research topic is the development of tools/methods for planning and prioritizing resources to determine which resources to defend most vigorously. This includes determining which resources are sufficiently vulnerable or compromised in a given attack. In case study #4 the entire travel system for many government agencies was shut down because of an intrusion only to later find out that it did not contain any personally identifiable information. Had there been a proper risk assessment of the system, it could have remained functional.

Another aspect of SA is awareness of relationships among different parts of the enterprise, and their associated roles and responsibilities. For example, fixing a problem in one organization may limit the capability of another to accomplish its mission. To maintain military superiority, the military may spend millions to develop a new cyber offensive capability; but this can be completely undermined by a third party that releases a patch for an unrelated issue; or by a vendor that discontinues support for a software application or operating system. Even the act of *revealing* a security vulnerability—which might be done to defend against an adversary—could be considered cyber friendly fire to the extent that it would limit the capabilities of friendly cyber operations to protect assets. In addition, research is needed to advance SA by developing tools that can predict how actions taken on a network will affect other resources and mission outcomes. An example is a tool that will notify you that installing security patch X will cause services A and B to fail and C to reboot.

Another research topic is how to create hyper-quiet networks in which all entities are known and understood and much of the noise is removed. By removing much of the dynamic nature of these networks, it would be easier to create useful models for simulating and understanding networks.

[14] http://www.zenoss.com/

Training Issues

Computer-based training (also called distance learning, electronic learning, or e-learning) has roots in traditional learning paradigms deriving from largely behaviorist traditions that reflect passive, rather than active, student-centered training philosophies (Greitzer, Kuchar & Huston, 2007). Although computer-based training approaches have traditionally followed a substantially linear process, they are becoming more student-centered as new methods and computer technologies that allow greater flexibility in the design and delivery of instructional material have emerged. The use of simulations and serious gaming approaches enables the application of an active learning paradigm that engages the learner in what may be defined as a mental contest (Zyda, 2005). Since the 1997 publication of the National Research Council report titled "Modeling and Simulation—Linking Entertainment and Defense" (Zyda & Sheehan, 1997), we have seen several serious games emerge for a variety of domains, including America's Army (http://www.americasarmy.com), SimNavy (Capps, McDowell & Zyda, 2001), and emergency preparedness (Turoff et al., 2006).

Serious games involve activities that educate or instruct, thereby imparting knowledge or skill—that is, they provide a good environment for learning. Game-based or simulation-based training immerses the learner in a realistic setting and engages the learner in an intrinsically motivating experience involving competition, goal-based behavior, and adaptive challenges. Games enable engagement in activities otherwise too costly, dangerous, difficult, or impractical to implement in the classroom. However, game-based e-learning programs are also costly to produce, and although they are immersive, they may not offer the degree of realism and the fidelity to the learner's environment to maximize the impact and effectiveness of the experiential learning approach.

Because the typical game-based and simulation-based approaches to training provide "generic" environments in which to develop and practice skills, there is always a risk that such training contexts will not map sufficiently to the operational world. Thus, transfer of learning may not reach the level that is expected or desired. In essence, what is desired—and missing from the current state of the art in simulation or game-based instruction—is a capability akin to the "holodeck" on the *Starship Enterprise*: an environment that is virtually the same as the “real world” and which offers the learner numerous opportunities to experience different scenarios just as they would experience them in their own work environments. Significantly, this includes emotional (stress) responses that help the learner adapt to decision making under stress in addition the sense of realism that facilitates positive transfer of training (see *Stress Effects*, next sub-section).

Training through game-based simulation excels at developing an understanding of the strategic management of computational assets through application of policy and procedures, but it is less effective in delivering tactical operational training. The primary cause is that most computational elements (software and hardware) are so complex and poorly understood that it is effectively impossible to simulate them with sufficient fidelity to provide a realistic training experience. This deficiency can be filled through use of a computational laboratory.

Research at PNNL has produced a virtual simulation environment that has many of the virtues of an effective instructional platform. We have implemented this environment as a network training simulation laboratory in which network administrators interact with computer screens and simulated problems that are identical to those that they face in their actual work environments. This is because the simulated environment is implemented for, and tailored to, the trainee’s environment—configured to be essentially identical to the trainee’s actual network environment. Because cyber security professionals view the world through a computer screen, the keyboard and monitor is a virtual world—no holodeck needed. They use their own tools and techniques but still use the same keyboard and monitor to view and control their world.

A key technology for simulation based training is *virtualization*. Virtualization is a commonly used alternative to physical devices (even in production environments) due to the lower acquisition and operations cost and greater reliability and scalability. The most realistic training environment has the same physical devices and connections as the live business network. Naturally, a completely physical device approach does not scale well and is typically cost prohibitive and impractical for creating large training networks. As the technology evolves and hardware begins to support virtualization natively, there is a growing capability to create virtual networks that appear as real and in many cases are as real as a physical network. This makes virtualization a viable and cost-effective way to build virtual networks for training.

At PNNL our research focuses on creating the ideal training environment for network administrators. Our research specifically focuses on simulating realistic traffic that is indistinguishable from traffic on an existing production network with a dynamic and flexible training environment (Irvine, Ouderkirk, & Greitzer, 2009). The simulation focuses on people or machines performing actions at stochastically scheduled times. Based on this simple approach, a corporate network can be simulated by knowing how many people or machines are connected, what type of actions they typically perform, and when these actions are typically performed. For most simulations, people are assigned to groups with similar characteristics and given tasks to perform based on a stochastic scheduling model. This approach is extremely flexible because it can scale from a simple model to a highly complex one with high granularity. When properly modeled, simulated users can more closely approximate production behaviors than live users in the same laboratory environment.

Once the simulated network is designed and generating traffic, the training can begin. A mix of ‘canned’ simulated and live attacks provides an excellent training environment that comes closer to approximating what actually happens on a live production network and in a live attack. Typically the focus of training is to gain proficiency with a number of network forensic tools; but for the purposes of training and awareness about cyber fratricide, the simulation environment can be used to build SA, familiarity with network vulnerabilities, interdependencies among network assets, and skills in working under increased workload and stress. One can see that in this environment, it is not only possible to support network administrator training for SA and acquisition of expertise in recognizing possible malicious attack patterns, but it is also possible to train administrators and other cyber security personnel on appropriate responses to these threats. The effect of these responses can be shown to the trainee in after-action debriefs as well as in real time—thus facilitating learning about the most effective responses and responses that limit or reduce the friendly network operations—that is, recognition of situations and defensive/offensive actions that may yield deleterious effects or cyber friendly fire. A by-product of this kind of training is the ability to recognize false-positives and deal with them appropriately.

Effects of Stress

Stress has strong effects on every aspect of cognition from attention to memory to judgment and decision making. In general, under stress, attention appears to channel, reducing focus on peripheral information and centralizing focus on main tasks (Kavanagh, 2005). Thus stress can produce a restriction in the range of cues attended to (tunneling). In decision making under stress, individuals may make decisions based on incomplete information (Janis and Mann, 1977)—failing to consider the full range of alternatives available, ignoring long-term consequences, and making decisions based on oversimplifying assumptions. Friedman and Mann (1993) refer to these “failures” as heuristics.[15]

In the context cyber security monitoring and response, the manifestations of stress in tunneling effects and decision biases may be observed in several ways. The high workload demands imposed by high volumes of malicious traffic and frequency of attempted compromises increase stress on network operators, which may increase tunneling and decision biases: On the one hand, the cyber analyst may become numb to the attacks and not respond aggressively enough, ignoring some attacks and allowing them to compromise system resources. On the other hand, the analyst may overreact to an intrusion and impose excessive limits on services and resources (overreactions were illustrated in Case Studies 4 and 6, for example).

One way to reduce stress effects is to provide over learning or extended practice. Tasks that are well-learned tend to be more resistant to the effects of stress than those that are less-well-learned. Extended practice helps to commit the knowledge to long term memory and facilitates retrieval, and it may produce “automaticity” and the proceduralization of tasks that make lower demands on attentional and mental resources

[15] While researchers who argue that perceptual narrowing reduces the quality of individual decisions, Klein (1996) observed that the use of heuristics may allow individuals to respond more quickly to external demands while under stress or when provided only partial information.

(Leavitt, 1979; Smith & Chamberlin, 1992). This yields greater resistance to the negative effects of stress—i.e., tasks are less likely to be forgotten and more easily recalled under stress.

Another training approach to mitigate the effects of stress is to provide training experiences that help the learner cope with stress. Because high stress during learning may degrade an individual's ability to learn—perhaps due to interference or disruption in the encoding and/or maintenance phases of working memory—a gradual increase in stress levels may offer the most effective instructional strategy. A phased approach should be used, with an initial learning phase under minimum stress, followed by gradually increasing exposure to stress levels that are more consistent with real-world conditions. Such "stress inoculation training" attempts to immunize an individual from reacting negatively to stress exposure. The method provides increasingly realistic pre-exposure to stress through training simulation; through successive approximations, the learner builds a sense of positive expectancy and outcome and a greater sense of mastery and confidence. This approach also helps to habituate the individual to anxiety-producing stimuli. A simulation-based, synthetic environment (as described above in the sub-section on Training) offers an effective approach to training on cyber security monitoring and response under stress.

Attack Attribution

In preparing for a response in cyber warfare, it should be noted that the initial source of an attack may not be in direct communication with the aggressor. Botnets allow a single individual to effectively coordinate the actions of several hundred thousand systems that are dispersed around the Internet and the world. The commands in botnets are relayed through a multitier hierarchy, which obscures the identity of the source issuing the orders. Many of the bots are compromised systems owned and operated by friendly or neutral parties who are unaware of their participation. Similarly, network intrusions are often bounced through a chain of compromised intermediaries with each system acting as a stepping stone to the next. Before a counterattack or another similar response is taken, the trail of compromised systems should be followed to uncover the true identity of the source; otherwise, friendly fire incidents will be increasingly likely.

Current practice requires these forensic investigations to be conducted by humans; they are expensive in both time and manpower. Thus, research is needed to provide automated support for attack attribution.

Automation

Increasing attack frequency and rapidly evolving attack vectors have necessitated the automation of defensive responses. The value in defensive automation is in rapidly identifying threats and then taking actions to mitigate them. But the problem remains in the false positive rate in threat detection. Even though automation can potentially provide a lower false positive rate when compared to people, imperfect SA remains problematic. Due to the efficiencies gained with automation, friendly fire incidents are likely and, without appropriate safeguards, the damage significant. An open research question is what types of responses are acceptable to automate. Active defensive responses such as engaging a counterattack seem inappropriate as the purpose of these responses is to inflict damage, and a misguided response can cause friendly fire incidents. On the other hand, passive defensive responses such as activating a push-back mechanism or modifying firewall rules sets seem appropriate. Any damage as a result of these types can be recified internally within the responding organization. Another open question is where and to what degree humans are inserted into the automation decision-making loop. Too much human intervention slows response times and may result in fatigue as well as frustration or dissatisfaction, as might occur with misuse-based (*i.e.*, signature-based) intrusion detection systems. Too little intervention, however, will not provide the appropriate levels of safeguards necessary to mitigate the risks in deploying automated responses. A balance must be struck between speed of response and intervention to guard against friendly fire incidents.

Conclusions

We began this report with a discussion of definitions of cyber friendly fire, and a question as to whether the factors underlying cyber friendly fire or “fratricide” are the same as those that have been identified in combat fratricide. Our discussion reveals that those factors or research issues that underlie combat fratricide also apply well to cyber friendly fire, but we identified issues of attack attribution and level of automation that present particular challenges for research on mitigating cyber friendly fire.

We conclude this report by composing the following list of the research issues that have been identified in the foregoing case studies and associated discussion:

- A cognitive task analysis of cyber security analysis and decision making should be performed to identify possible sources of information overload and to identify sources of bias and stress.
- SA tools are needed that can predict how actions taken on a network will affect other resources and mission outcomes.
- Research is needed to develop visualizations that enhance SA. What can be done to help analysts understand interdependencies among nodes/entities in a computer network so that they will have sufficient SA to avoid adverse effects of unintended consequences?
- Measures and metrics of workload should be developed for cyber security monitoring and analysis
- Measures of the level of trust (trustworthiness) are needed for cyber defense tools. Research is needed to identify how to create an appropriate level of trust.
- Research on risk analysis/risk metrics is needed to support cyber defense tradeoff decisions. Given that attacks are inevitable, how can risk analysis be used to prioritize system resources and inform decision makers about which systems to protect and which to sacrifice? How can one quickly determine the break-even point between potential compromise and certain loss of productivity?
- Simulation and synthetic environments for training must be used to provide cyber analysts realistic scenarios in network environments that match their operational environments.
- Research on tradeoffs in automation is needed to identify the types of decisions and responses that should be automated and the level of control/oversight by human decision makers. What level of autonomy maximizes response time and minimizes the false positive rates?

References

Capps, M., McDowell, P., and Zyda, M. 2001. A Future for Entertainment – Defense Research Collaboration. *IEEE Computer Graphics and Applications, 21(1)*, 37-43.

Friedman, I. A., & Mann, L. (1993). Coping Patterns in Adolescent Decision-Making: An Israeli-Australian Comparison, *Journal of Adolescence, 16*, 187-199.

Greitzer, F. L., & Andrews, D.H. (2008). Training strategies to mitigate expectancy-induced response bias in combat identification: A research agenda. Keynote Address and Invited Paper, *Human Factors in Combat ID Workshop*, May 2008, Mesa, AZ. PNNL-SA-59942. http://cerici.org/workshop/2008Workshops/Greitzer-Andrews%20CID%20Chapter%20FINAL.pdf

Greitzer, F.L., Kuchar, O.A., & Huston, K. (2007) Cognitive science implications for enhancing training effectiveness in a serious gaming context. *ACM J. Edu. Resources in Comput.*, Vol. 7, No. 3, Article 2 (August 2007).

Irvine, L., Ouderkirk, S.J., & Greitzer, F.L. (2009). Network Administrator Training Using a Secure Virtual Laboratory. PNNL Technical Report PNNL-SA-63441. Richland, WA: Pacific Northwest National Laboratory.

Janis, I. L., & Mann, L. (1977). *Decision Making*, New York: The Free Press.

Kavanagh, J. (2005). *Stress and performance: A review of the literature and its applicability to the military*. (RAND TR-192, ADA439046). Santa Monica CA: Rand Corp. [Online]. Available at: http://www.rand.org/pubs/technical_reports/2005/RAND_TR192.pdf [accessed: February 12, 2008].

Klein, G. (1996). The Effects of Acute Stressors on Decision-Making. In J. Driskell, & E. Salas (Eds.), *Stress and human performance* (pp. 49-88). Hillsdale, NJ: Lawrence Erlbaum.

Leavitt, J. (1979). Cognitive demands of skating and stick handling in ice hockey. *Canadian Journal of Applied Sport Sciences, 4*, 46–55.

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies, 27,* 527-549.

Parasurman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced complacency. *International Journal of Aviation Psychology, 3(1),* 1-23.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39(2),* 230-253.

Sarter, N. B., & Woods, D. D. (2000). Teamplay with a powerful and independent agent: A full-mission simulation study. *Human Factors, 42(3),* 390-402.

Smith, M. D., & Chamberlin, C. J. (1992). Effect of adding cognitively demanding tasks on soccer skill performance. *Perceptual and Motor Skills, 75*, 955–961.

Sorkin, R. (1988). Why are people turning off our alarms? *Journal of Acoustical Society of America, 84*, 1107-1108.

Symantec (2009). *Symantec Global Internet Security Threat Report: Trends for 2008. Volume XIV*. Symantec, April 2009. Retrieved from http://www4.symantec.com/Vrt/wl?tu_id=gCGG123913789453640802, November 2009.

Turoff, M., Chumer, M., Hiltz, S.R., Hendela, A., Konopka, J., and Yao, X. (2006). Gaming Emergency Preparedness. In *International Conference on System Sciences*, 38-47.

US Department of the Army. (1992). *Fratricide: Reducing Self-Inflicted Losses*. Center for Army Lessons Learned (CALL) Newsletter No. 92(4). Fort Leavenworth, KS: Center for Army Lessons Learned, U.S. Army Combined Arms Command.

[Online]. Available at: http://www.globalsecurity.org/military/library/report/call/call_92-4_tblcon.htm# [accessed: October 26, 2009].

US Department of the Army. (1993). Military operations: U.S. Army operations concept for combat identification (TRADOC Pam 525-58). Fort Monroe, VA: Training and Doctrine Command.

Wilson, K. A., Salas, E., Priest, H. A., & Andrews, D. (2007). Errors in the Heat of Battle: Taking a Closer Look at Shared Cognition Breakdowns Through Teamwork. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 49 (2)*, 243-256

Zyda, M. 2005. From Visual Simulation to Virtual Reality to Games. *IEEE Computer*, *38(9)*, 25-31.

Zyda M. & Sheehan, J. (eds.) (1997). *Modeling and Simulation – Linking Entertainment and Defense*. Landover, MD: National Academies Press.